Human Molecular Genetics, 1995, Vol. 4, No. 1 59–69

Integration of physical, breakpoint and genetic maps of chromosome 22. Localization of 587 yeast artificial chromosomes with 238 mapped markers

Callum J.Bell*, Marcia L.Budarf, Bart W.Nieuwenhuijsen[1], Barry L.Barnoski, Kenneth H.Buetow[2], Keely Campbell, Angela M.E.Colbert[3], Joelle Collins, Mark Daly[3] Philippe R.Desjardins[1], Todd DeZwaan[1], Barbara Eckman[1], Simon Foote[3,+], Kyle Hart[1], Kevin Hiester[1], Marius J.Van Het Hoog[1], Elizabeth Hopper, Alan Kaufman[3], Heather E.McDermid[4], G.Christian Overton[1], Mary Pat Reeve[3], David B.Searls[1], Lincoln Stein[3], Vinay H.Valmiki[1], Edward Watson, Sloan Williams, Rachel Winston[1], Robert L.Nussbaum[1,§], Eric S.Lander[3], Kenneth H.Fischbeck[1], Beverly S.Emanuel and Thomas J.Hudson[3]

Children's Hospital of Philadelphia, Division of Human Genetics and Molecular Biology, 34th Street and Civic Center Boulevard, Philadelphia, PA 19104, [1]University of Pennsylvania School of Medicine, 415 Curie Boulevard, Philadelphia, PA 19104-6146, [2]Fox Chase Cancer Center, 7701 Burholme Avenue, Philadelphia, PA 19111-2412, [3]Center for Genome Research, Whitehead Institute for Biological Sciences/Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, MA 02142, USA and [4]Department of Biological Sciences, University of Alberta, Edmonton, Alberta T6G 2E9, Canada

Received October 17, 1994; Revised and Accepted November 2, 1994

Detailed physical maps of the human genome are important resources for the identification and isolation of disease genes and for studying the structure and function of the genome. We used data from STS content mapping of YACs and natural and induced chromosomal breakpoints to anchor contigs of overlapping yeast artificial chromosome (YAC) clones spanning extensive regions of human chromosome 22. The STSs were assigned to specific regions (bins) on the chromosome using cell lines from a somatic hybrid mapping panel defining a maximum of 25 intervals. YAC libraries were screened by PCR amplification of hierarchical pools of yeast DNA with 238 markers, and a total of 587 YAC clones were identified. These YACs were assembled into contigs based upon their shared STS content using a simulated annealing algorithm. Fifteen contigs, containing between 2 and 74 STSs were assembled; and ordered along the chromosome based upon the cytogenetic breakpoint, meiotic and PFG maps. Additional singleton YACs were assigned to unique chromosomal bins. These ordered YAC contigs will be useful for identifying disease genes and chromosomal breakpoints by positional cloning and will provide the foundation for higher resolution physical maps for large scale sequencing of the chromosome.

INTRODUCTION

Human chromosome 22 constitutes approximately 1.9% of the haploid autosomal genome (1). Clinical disorders associated with this chromosome include several acquired, tumor-related translocations such as the t(9;22) of chronic myelogenous leukemia and acute lymphocytic leukemia (2,3), the t(8;22) variant translocation of Burkitt's lymphoma (4) and the t(11;22) of Ewing's sarcoma (5,6). Deletions of all or part of chromosome 22 are associated with meningiomas (7,8), acoustic neuromas (9,10), Neurofibromatosis type 2 (NF2) (11,12), and rhabdoid tumors (13,14). Further, chromosome 22 is also involved in the only recurrent non-Robertsonian constitutional chromosomal translocation in humans (15,16). In addition, a number of syndromes are caused by deletions or duplications of portions of 22q11, including DiGeorge syndrome (17–20), velo-cardio-facial syndrome (21), and cat-eye syndrome (22). Chromosome 22 has a high gene density and contains many duplicated sequences and gene families, which makes it an interesting model for mapping studies. The identification of

*To whom correspondence should be addressed

Present addresses: +Walter and Eliza Hall Institute of Medical Research, PO Royal Melbourne Hospital, Parkville 3050, Australia and §Genetic Disease Research NCHGR/NIH, 9000 Rockville Pike, 49/4A72, Bethesda, MD 20892, USA

STRPs (simple tandem repeat polymorphisms) (32), expressed sequence tags (ESTs) (33,34), YAC vector-insert junction fragments (35), inter-*Alu* PCR fragments (36) and randomly sequenced plasmid clones (26,27). The loci at which STSs or probes were generated are shown in Table 1.

Chromosomal bin assignment of markers

Markers were assigned to chromosomal 'bin' locations by Southern blot hybridization or PCR analysis of DNA from cell lines in a 26 member somatic cell hybrid panel . These cell lines define 22 bins shown schematically in Figure 1. Three of these bins are each further subdivided into two sub-bins, making a total of 25 intervals. The majority of the hybrids have been previously described: GM10888 (37); Cl-6-2/EG, Cl-21-5/CV, Cl-9/GM05878 (38); Rad-110a; Rad 37a (39); GM11220 = X/22 33-TG, GM11224C = 1/22AM-6, GM11223C = 1/22AM-27 (40); GM11685 (41); Cl-4/GB, Cl-1-1/TW (42); AJO 9, APR 8.5 (43); 514 AA2 (44); WESP-2A-TG8 = GM11221 (45); RAJ5BE (46); D6S5 (47). There are eight additional members of the hybrid panel (Cl-3/5878; Cl-1/5878; Cl-2/5878; Cl-8/5878; Cl-15-1/PB; Cl-21-2/PB; Cl-2/DIBA; Cl-8-1/AMB6) which will be further described in another manuscript (26). Not all markers were assigned to a unique bin. STSs binned in the Whitehead Institute/MIT Genome Center were tested on a subset of six somatic cell hybrid lines (Fig. 1), whereas those binned in Philadelphia were tested on the complete panel. A small number of STSs could not be uniquely assigned for technical reasons.

Contained within this hybrid mapping panel are the breakpoints which have been designated by the chromosome 22 mapping community as anchor positions in the physical map. The anchor panel was recently updated (48) and now

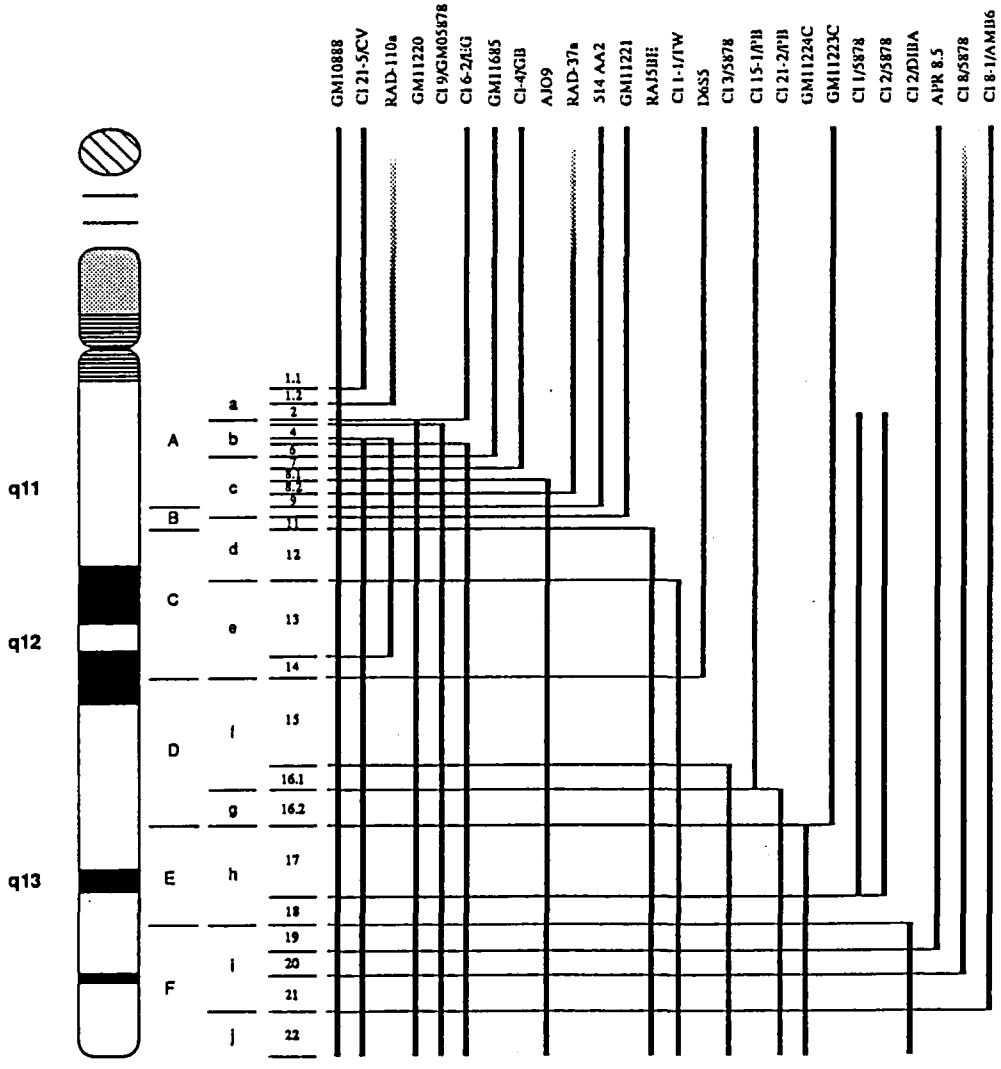


Figure 1. Schematic of the somatic cell hybrid mapping panel used for bin assignment of markers. The heavy vertical black lines represent the segments of chromosome 22 retained in each hybrid. Shaded portions indicate that the extent of the p arm retained in a hybrid is unknown. The names of the hybrids are shown at the top of the figure. The fine horizontal lines indicate the breakpoints that divide the chromosome into 25 intervals, shown numbered from 1.1 through 22 (three 'bins'are further subdivided into two sub-bins). The upper case letters A–F indicate a subset of the 26 member panel defining six intervals that was used for bin assignment at the Whitehead Institute/MIT Genome Center. The full high resolution panel was used at the Children's Hospital of Philadelphia. The lower case letters a–j show the 10 interval panels defined by the 11 hybrid cell lines available from the NIGMS repository (48).

divides chromosome 22q into a total of 10 intervals which represents a subset of this mapping panel. The hybrids defining the anchor points are: GM11220, GM11685, GM11221, GM11222C, GM11224C, GM11223C, D6S5, Cl-15-1/PB (GM13498), Cl-21-2/PB (GM13499), Cl-2/DIBA (GM13501) and Cl-8-1/AMB6 (GM13500). The somatic cell hybrids defining the anchor points of chromosome 22 are available through the NIGMS genetic mutant cell repository, Camden, New Jersey. Since the chromosome 22 reference hybrids represent a subset of the panel utilized for binning markers described in this manuscript, the data presented here can be easily assessed by other groups in order to position markers which they have mapped using the reference panel.

YAC identification

Most YACs in the study were identified in the CEPH/Genethon libraries [original library (29) with an average insert size of 470 kb and mega-YAC library with an average insert size of 0.9 Mbp (30)], by PCR screening of yeast DNAs pooled in two or three dimensions. Additional YACs were isolated from the Washington University YAC library (31), and from a chromosome 22 specific YAC library constructed with DNA from hybrid cell line GM10888 (chromosome 22 in a Chinese hamster background). The chromosome 22 specific YAC library contains approximately 300 YACs with an average insert size of 200 kb, equivalent to 1× coverage of the chromosome. YACs isolated from the Washington University library were kindly provided by collaborators. In addition, limited use was made of a subset of YACs, kindly provided by Ilya Chumakov and Daniel Cohen, identified by hybridization of *Alu*-PCR products of a chromosome 22-only somatic cell hybrid to the CEPH mega-YAC library. YACs from this subset, and from the chromosome 22-specific library were identified by colony hybridization.

Table 1 shows a summary of the YAC screening results. The left-most column shows the bin intervals, numbered 1.1–22. The relative positions of the bins on the chromosome are displayed visually in Figure 1. Loci that were used to identify YACs are shown in boxes in the body of the table; the vertical extent of each box indicates the bin, or range of bins, to which each locus was mapped by referring to the left-most column, and the number of YACs detected by each locus is indicated in parentheses after the locus name. The majority of these results are YACs identified to single microtiter plate addresses, either from unequivocal PCR results in two or three dimensional screens, or from confirmatory PCR tests done on individual YACs. A YAC address consists of three dimensions: plate, row, and column. In initial screening of YAC pools, many of the addresses were incomplete (missing a dimension), or had more than one possible value in a dimension, which occurs when there is more than one positive YAC per block of eight microtiter plates (see Materials and Methods), or from false positive results. Such ambiguous addresses were resolved by several means including fingerprint analysis, comparison with verified YAC addresses of adjacent STSs, or PCR of all possible clones in the degenerate set of addresses. After preliminary contig assembly, most of the clones identified as well as the putative adjacent YACs were individually tested with each STS in the contig.

GGTX, GGTY and GGTZ (Table 1) refer to probes containing sequences homologous to γ-glutamyl transpeptidase 1 (GGT1) (49). These three GGT-like sequences have been shown to be physically linked to the BCR (break point cluster)-like sequences BCRL2 and BCRL4, and to BCR itself, respectively, in 22q11 (50). These BCR-like sequences contain polymorphic *Hin*dIII sites and thus can be distinguished from each other (51), allowing assignment of the YACs detected by the GGT1 STS to be allocated to unique bins. Details of this study will be presented in a separate publication.

Primer sequences for each STS and YAC addresses may be found in the public FTP (file transfer protocol) sites of the Philadelphia (cbil.humgen.upenn.edu/pub/22/) and the Whitehead Institute/MIT (genome.wi.mit.edu /distribution/human_STS_releases/) Genome Centers. World Wide Web access is available through HTTP://www.cis.upenn.edu/~cbil/chr22db/chr22dbhome.html and HTTP://www-genome.wi.mit.edu.

In order to resolve confusion caused by possible cross-contamination among microtiter plate wells we adopted two approaches. The first approach compared the CEPH/Genethon fingerprints, where available, of the putative YAC positives with the fingerprints of other YACs known by STS content to overlap the YAC to be resolved. Shared fingerprint bands

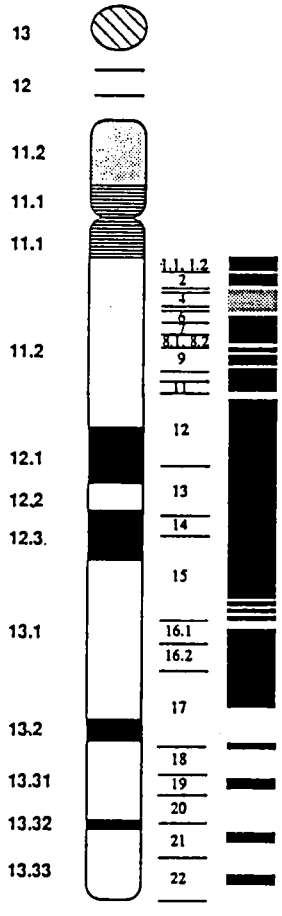


Figure 2. Estimated coverage of the chromosome in contigs. The horizontal lines are the boundaries separating 25 intervals. Contigs are shown as blocks. The stippled block shows the location of a cosmid contig encompassing the DiGeorge critical region (DGCR).

among these YACs identified with a high degree of confidence the true positive YAC address among several neighboring candidates in several cases. The second approach was based on a calculation of the actual distances between wells of two YAC addresses sharing STSs, divided by the number of STS hits in common; when this measure fell below a certain threshold for any pair of addresses, they were consolidated into a single address. This heuristic in all cases corresponded well to human judgments about likely cross-contamination, and was shown to be justified in cases that were checked experimentally. Level 1 data from the CEPH/Genethon genome mapping project were confirmed and included in Table 1.

YAC contig assembly

To date, we have used 238 markers to identify 587 YACs. The YACs and STSs fall into 15 islands, defined as sets of STSs and sets of YACs all of which can be reached from each other by following a path of connectivity alternating between STSs and YACs. Singleton YACs detected by one STS each, numbering 25, are omitted from this total. Although the number of YACs we identified indicates nearly 5× coverage of the chromosome, the depth of coverage is uneven: all somatic cell hybrid bins contain YACs, but the 22q11.23−q12.31 region (bins 12–15; see below) has much deeper coverage than

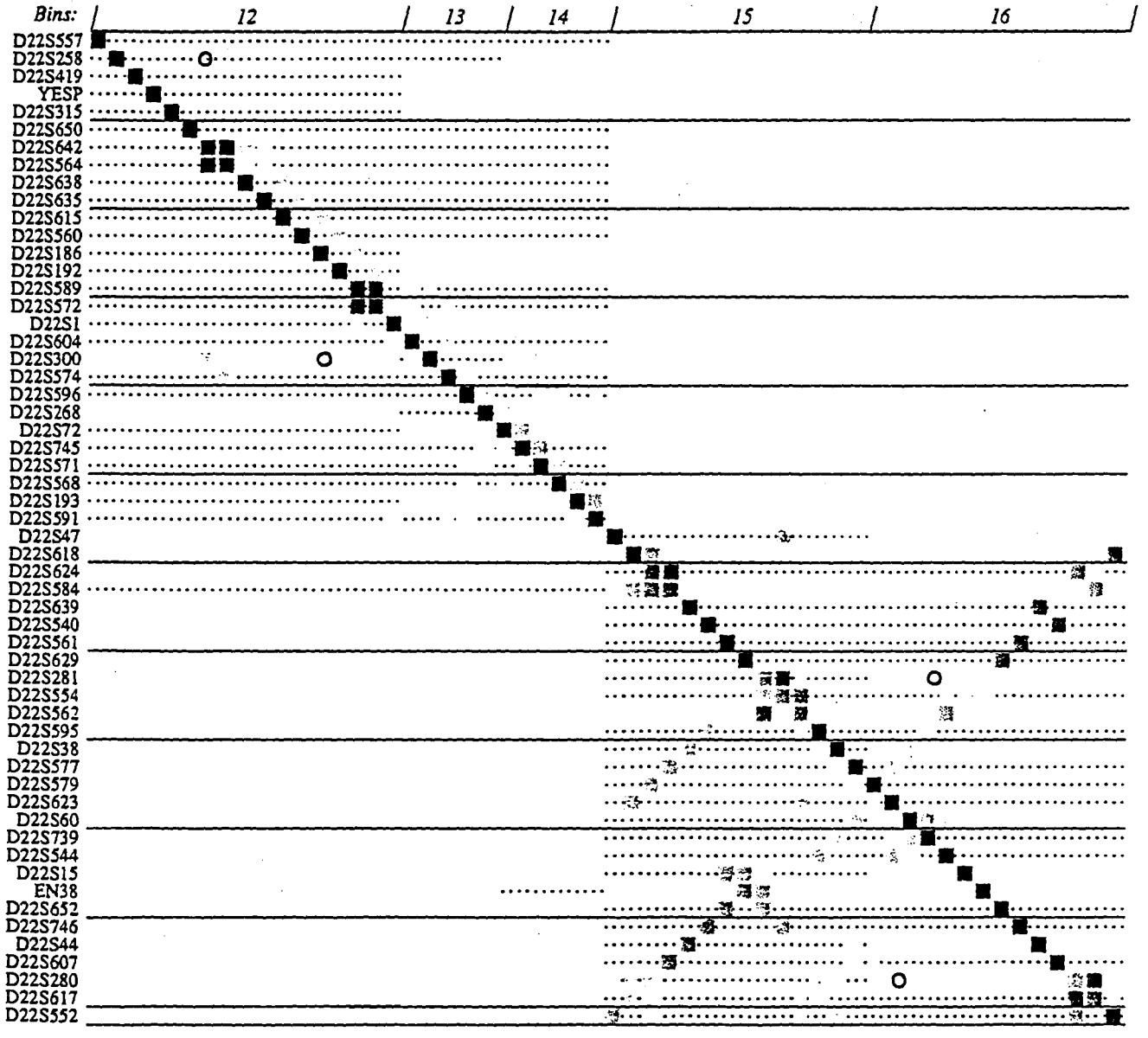


Figure 3. Searls plot of simulated annealing data for the largest contig accumulated from multiple runs of the program. The list of loci down the left of the figure is the 'minimum energy' ordering of markers (see the text for detailed explanation). Gray boxes indicate the position on the horizontal axis at which the indicated STSs occurred during individual runs. Darker boxes indicate that an STS was positioned in the same location in multiple runs. Boxes falling repeatedly on the diagonal indicate high confidence in the minimum energy ordering. Horizontal dotted lines indicate the chromosomal bin location of each STS. The bin intervals are shown at the top of the figure. Circles indicate the consensus positions of markers that are present on the meiotic map.

elsewhere. We had difficulty obtaining unequivocal clone and STS order within the largest of these islands, and a clear clone tiling path, even with deep YAC coverage of the area and many STSs. In the central portion of the chromosome YAC connectivity has been achieved over a distance exceeding 10 Mb, yet an unbroken clone tiling path remains elusive despite extensive testing of YACs versus STSs in that region. This may be due in part to false positive and negative YAC/STS results (although results have been carefully confirmed), internal deletions within YAC clones, and sequences present at more than one location on the chromosome. Given these problems, the objective becomes to find an ordering of STSs that minimizes gaps. In ideal data, there should be an order of STSs, corresponding to a true YAC contig, such that there are no such gaps. However, in our data all postulated orders of STSs in an island result in some number of 'gaps' within YACs in the island, defined as cases where a YAC is negative for some STS but positive for STSs located to both the left and right in the ordering.

For very large islands, finding the STS order with the absolute minimum number of gaps is computationally intractable, but several approaches have been developed to finding approximate solutions. A simulated annealing (52,53) program we developed employs a random search strategy that seeks local energy minima in the space of all possible orderings, where energy is defined in terms of numbers and sizes of gaps (see Materials and Methods). This approach can be expected to yield somewhat different results for multiple runs, both because there may be more than one valid ordering even for ideal data, and because for 'noisy' data the search may find different local energy minima which are near the actual optimum. In practice, the results of multiple runs of simulated annealing are generally similar, although not identical. We refer to these orderings of STSs and YACs as contigs, though it should be emphasized that the larger islands should be viewed as putative contigs at present.

A schematic representation of the coverage of the chromosome in contigs is shown in Figure 2. The chromosome is shown divided into 25 intervals derived from the somatic cell hybrid map of Budarf *et al.* (26). Bin 1 formally includes the

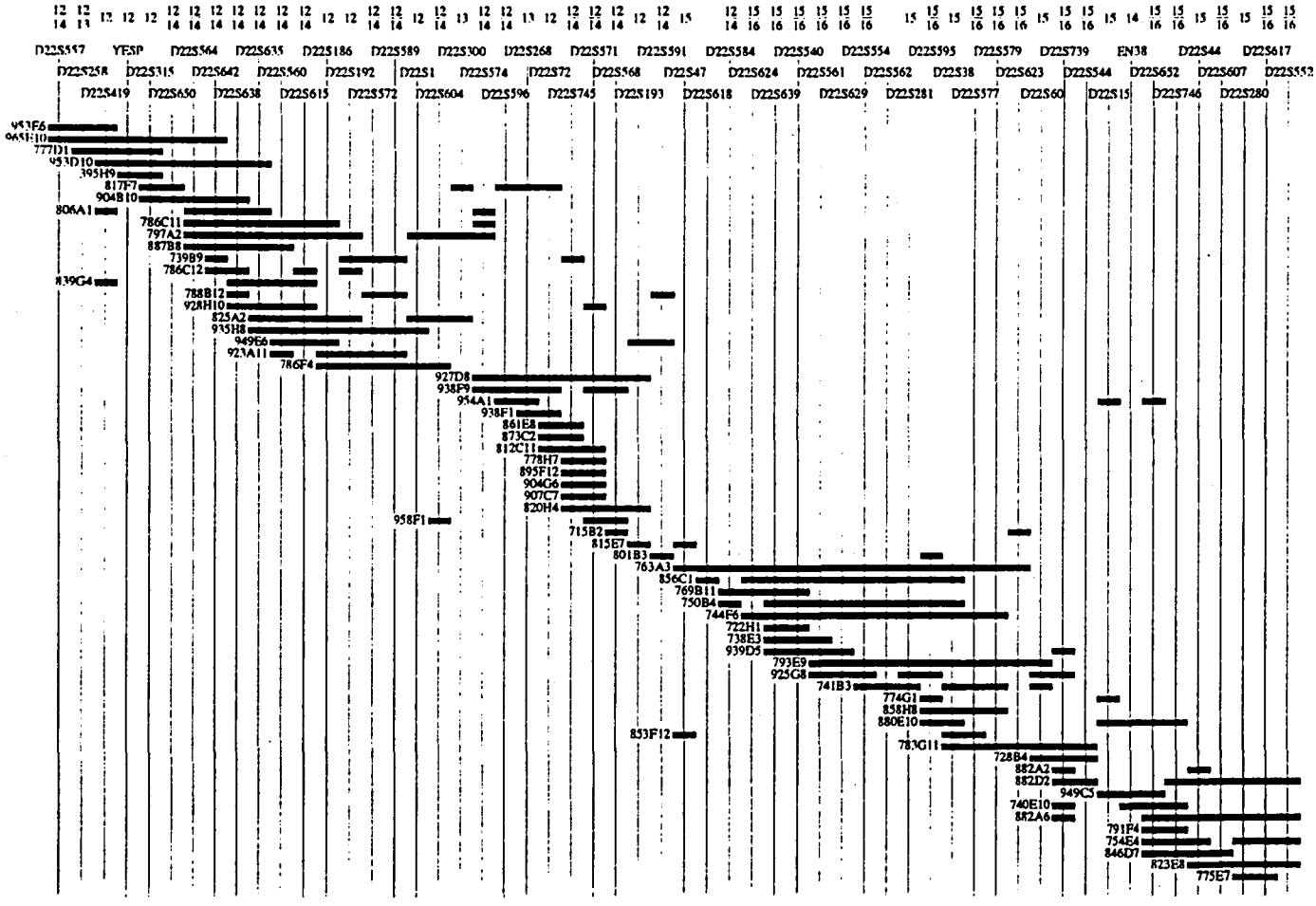


Figure 4. A single solution for the largest contig in the central region of chromosome 22q. The contig was constructed as follows: YACs and STSs were selected by connectivity to D22S1, obeying the double linkage rule. Singletons (YACs detected by one STS only) were then eliminated, as were markers that detected more than 14 YACs. Singletons were eliminated a second time, and the resulting set of markers and YACs were subjected to simulated annealing. Marker order is shown along the top of the figure. Above each marker name is the bin interval that the marker was mapped to, e.g. 15/16 indicates the marker is in bin 15–16. YACs are shown as heavy horizontal black lines.

short arm of the chromosome. The contigs, based on the bin assignment of the STSs that detected the YACs in each, are shown as dark blocks. Since STS content mapping provides only limited information on contig size, the true extent of coverage and the sizes of the gaps separating the contigs are unknown. The stippled block represents a contig of cosmids in a region that proved difficult to clone in YACs. YACs detected by STSs in this part of the chromosome were unstable, and were underrepresented in the libraries screened (M.Budarf, unpublished observations). The cosmid map of this region will be described in a separate publication. Figure 2 makes clear the low coverage of the distal portion of the chromosome. This arises in part from the lower density of markers but is largely due to underrepresentation of the region in the mega-YAC library.

Figure 3 shows simulated annealing results for the largest contig, using a novel method of representing such data to which we have given the name 'Searls plot', after the author of the program. As noted, results of simulated annealing tend toward local minima of the objective function that may differ among runs. The relative merits of these STS orderings and implied YAC contigs cannot be judged with confidence on the basis of the STS data alone. On the other hand, a number of such orderings independently arrived at may be expected to represent a reasonable sampling of the contours of the search space of possible STS orderings. If the predicted orderings do not resemble each other, then little can be said about which is closest to the true optimum, but if they are all similar, one may be more confident in their consensus. Figure 3 shows the degree and nature of the consensus for multiple simulated annealings. The minimum energy ordering among all runs is indicated by the list of STSs running down the left hand side. The gray boxes in the diagram show the positions along the horizontal axis at which the indicated STS occurs in a run, so that the major diagonal denotes complete agreement with the minimum energy run. Other gray boxes indicate positions at which that STS occurred in other runs, and the shading of a box reflects the number of times a particular STS occurred at the same position in a run. If the predictions for an STS tend to cluster at more than one position in multiple runs, one may infer that the evidence is not strong enough to greatly favor one position over another, though it may be possible to narrow the possibilities to a few regions.

As noted above, even with ideal data it may be possible to have more than one ordering, particularly over subregions of the contig. Obviously, a given ordering of STSs may be reversed in its entirety, without changing the apparent fit to the YAC data in isolation, and for that reason each simulated annealing run is reversed, if necessary, to more closely approach the consensus. However, there may also be subregions over which the STSs can be reversed without affecting the energy materially, and in this case the Searls plot will display a characteristic 'X' pattern across the diagonal, representing the alternative orderings. Another characteristic pattern is a displacement of a subregion laterally on the plot, with either a forward or reversed directionality, indicating parts of the contig that display local integrity but which can be moved elsewhere in the larger scheme of things, with little or no penalty. Finally, there are subregions where STSs tend to be in proximity to each other, but where there is little support for ordering them with respect to each other. This may occur, for example, where there are multiple YACs with the same STS hits, but no YACs with only partial overlap to split the STSs and provide order information. These appear as 'clouds' of points at or near the diagonal; it can be seen that with a sufficient sample size such regions would approach a uniform distribution of points within a diffuse 'superblock'. Figure 3 shows a major 'X' indicating that the ordering in the distal half of the contig was inverted in a significant number of the simulated annealing runs. We interpret this to mean that the link between D22S591 and D22S47 should be viewed with caution. We have yet to confirm by other means whether actual continuity of YAC coverage exists in this region.

Figure 3 suggests, with some confidence, a general ordering of STSs in most sections of this region of the chromosome, but in some areas there is significant scatter. Some of this deviation is systematic in nature, as described in the previous paragraph, and some in all likelihood merely reflects regions where the data is error-prone. An external test of the accuracy of this method is provided not only by bin information but by the meiotic and pulsed-field gel maps (54,55) of the region; the orders of the subsets of markers in both of these maps are similar in the converged order arrived at by simulated annealing, which in this case was done without regard to information from any of these other methods. Figure 4 shows a single simulated annealing solution to the largest contig.

DISCUSSION

We used physical, breakpoint, and meiotic maps of human chromosome 22 to localize contigs of overlapping YAC clones that provide extensive coverage of the long arm of the chromosome. The physical map is developing rapidly due to considerable new data obtained by screening YAC libraries with STSs. The contigs, most of which are anchored by landmarks that have been ordered by meiotic or hybrid mapping, provide extensive coverage of the long arm of the chromosome. Although long range continuity of the contigs is not yet complete, the present information is of immediate use to the human genetic mapping community for identifying disease genes and chromosomal breakpoints. The current state of the physical map reported here reflects the fundamental characteristics of the reagents and methods used, as well as the inherent nature of chromosome 22 itself.

STSs that were developed for chromosome 22 are not randomly distributed along the chromosome. The contig spanning interval 22q11.2–q13.1 is the most evolved as the result of the high density of markers in this region and greater than average representation of the region in the YAC libraries. The distribution of markers shows a bias towards the center of the long arm of the chromosome (26,27). This is partly because many STRP markers were used as STSs, and these are known to be concentrated in the 22q12 Giemsa-dark chromosomal band (55). However, it is not known why other randomly chosen STSs generated from flow-sorted material should also be biased in this way. The distal third of the long arm is correspondingly poor in STSs, and appears to be underrepresented in the YAC libraries, and as a consequence, contains only two small contigs and seven singleton YACs. Interestingly, the distal portion of the long arm appears to be resistant to cloning in both plasmid and YAC libraries, and the consequent paucity of mapping information indicates the need for alternat-

ive strategies for covering this region. Currently, we are targetting the region by generating STSs from inter-*Alu* plasmid libraries made from radiation hybrid cell lines that retain only the distal portion of the chromosome. Success in developing new STSs in this way has shown that YACs, not markers, are likely to be limiting for YAC−STS contig mapping, and that complete coverage of this region will probably depend on a different cloning vehicle. Current candidate systems are bacterial artificial chromosomes (BACs) (56), P1 phage clones (57), P1 artificial chromosomes (PACs) (58), and cosmids.

Screening multi-dimensional pools of YACs was the only practical way to test all 25,000 mega-YACs for the presence or absence of a given STS, but created several types of problems. Contamination of adjacent wells during preparation of the pools, absence of amplification in one dimension, or the presence of more than one positive YAC in the same pool were examples of difficulties that are inherent to pooling schemes which can result in false positive, false negative, and ambiguous YAC addresses. Most of the results obtained from the pool screenings have been resolved by a variety of methods, including analysis of YACs seen with adjacent STSs, fingerprint analysis of selected YACs, and ultimately, the verification of the PCR on the individual YAC. To decrease the errors caused by false negatives on STS order, most STSs were screened on adjacent YACs as well.

The CEPH mega-YACs, which have an average insert size of 0.9 Mb (30), provided the best tool for linking STSs and assembling contigs, and were screened with all available markers. By requiring double linkage (59) before declaring contiguity among STSs in the largest contig, large clones were required, and YACs from the other libraries, while contributing to deep coverage in most regions, did not, in general, contribute to contig assembly. However, in some notable cases contig construction was dependent upon the smaller clones, and as the map matures, they will be useful in resolving the order of closely spaced STSs, and as tools for isolating cosmids or other smaller clones as the map moves towards a higher level of resolution required for eventual sequencing.

In addition to the known families of chromosome 22 specific repeats on long arm, such as the BCR, immunoglobulin and GGT gene families, we observed several markers which appear to behave as low copy repeats. In such cases, the PCR assay amplifies two identical or related sequences with products of similar molecular weights. Examples of this were D22S33 and D22S275, which gave several bands of similar size, and detected 15 and 14 YACs respectively. Repetitive STSs created inconsistencies in the data, manifested as large apparent gaps in YAC clones, since contig assembly software tries to assign them single contig locations. In fact, they may be present at two or more locations. Repeats therefore artificially connect YACs at disparate locations. We arbitrarily decided that STSs detecting 14 or more mega-YACs would be declared potentially repetitive and excluded them from contig construction.

The CEPH-Genethon tiling paths (60), provided relatively little additional information because the areas covered by tiling paths coincided with the region where the STS physical map was already well covered. We independently screened the mega-YAC library for the same Genethon genetic markers (61,62), and confirmed the YAC addresses and the level-1 tiling paths present in the November 1993 CEPH-Genethon data release (60). We extracted a few YAC addresses derived by *Alu*-PCR hybridizations in 22q11.2−q13.1 region that were missed during YAC pool screening. Unfortunately, the areas where the STS content map was poor were also not represented in the tiling paths, or present only in higher level paths that could not be confirmed. Fingerprint analysis on the mega-YACs generated by CEPH (60,63), was used to resolve ambiguous addresses derived from screening pools of YACs. This method, successful in one third of ambiguous addresses tested, reduced the number of alternate addresses that need to be verified for YAC determination. We did try to assemble the 22q11.2−q13.1 contig by fingerprint analysis alone using only the fingerprints of YACs that were previously identified to this region. The results had only limited success, yielding small contigs with less than 10 YACs that were already shown to have extensive overlap in STS content.

We chose to represent the data for the large contig in two ways: a single simulated annealing solution, and the Searls plot, derived from multiple runs of simulated annealing. These representations, combined with the YAC−STS results shown in Table 1, provide an objective and useful means of using these data. Previous localization of markers by recombination or breakpoints greatly facilitated the evaluation of the STS content map. The marker order in region 22q11.2−q13.1, spanning more than 11 cM, was broadly consistent with the orders of subsets of markers arrived at by meiotic and pulsed-field gel mapping (54,55). The smaller contigs contain, at most, two genetically ordered markers, which does not allow real comparisons of marker order with the meiotic map. In essence, we have made the assumption that the framework linkage map (55) is correct, and used it to anchor and orient the smaller contigs. The best validation of the smaller contigs came from concordance with the somatic cell hybrid binning results.

It is clear that, due to problems inherent to YACs, the STS-content mapping results from the large contig did not allow us to obtain a fine structure order of the region. This may well be true for many other regions in the genome. The need for additional methodologies to obtain a finer scaffold map of STSs is evident. Radiation hybrids, which allow the study of multiple, larger DNA fragments at a higher redundancy may provide more confidence in generating a high resolution STS order. They will also allow contiguity of the STS map in regions where YAC clones are few or absent.

The contigs reported in this paper will facilitate the study of several disease-related and structural regions of interest on chromosome 22. The YACs that have been localized to bins 1 and 2 (22q11.1−q11.2) form contigs that almost completely cover the cat eye syndrome critical region (CECR). The most proximal of these will allow us to address the problem of defining the physical boundary of the centromere. The completion of a contig containing all of the CECR, facilitated by a pulsed field gel electrophoresis map (64), will permit detailed transcription mapping of the region as a first approach to defining genes that contribute to this syndrome. In the region distal to the CECR, 22q11.2, difficulty was encountered in obtaining stable YACs in bins 3, 4 and 5. YACs identified in this region were frequently smaller than the mean insert sizes of the libraries, indicating that they contained deletions, and several probes failed to detect YACs. These bins represent the DiGeorge syndrome commonly deleted region which is notably unstable in humans giving rise to the deletions seen

atients with DiGeorge syndrome and velo-cardio-facial drome (65). It is interesting to note that these, and other uences mapping to the sites of frequent chromosomal angements in cat eye syndrome are also unstable when ed in yeast. Further characterization of these sequences allow us to investigate the possible causes of instability.

e constitutional t(11;22) translocation breakpoint is the recurrent, non-Robertsonian, constitutional translocation umans (15,16), and defines the boundary between bins 7 8. Contigs spanning this region may help in revealing ctural features on the chromosome that underlie this angement, as well as the identification of genes suspected nvolvement in breast cancer tumorigenesis (66). Identifica- of clones that span the t(11;22) breakpoint has been plicated by the presence of several duplicated regions in 11 which include the GGT and BCRL loci. In addition to e known ancestral duplications, STS screening results ɔest the presence of other low-copy repeat families that e the construction of a contiguous clone map of 22q11 icularly challenging. The largest contig, connecting bins and 15 (22q11.2–q13.1) contains several interesting fea- s that have already been well characterized, including the ng's sarcoma breakpoint (5,6), the NF2 gene (11,12) and candidate meningioma gene β-adaptin (67).

conclusion, the physical map of human chromosome 22 advanced considerably, due to the large scale screening of CEPH mega-YAC library with chromosome 22 specific s, and several regions of interest are now contained within contigs. Current efforts to achieve a complete set of rlapping clones for the long arm of the chromosome are cted at the generation of additional STSs for clone screen- as well as targeted strategies for the distal third of the mosome using *Alu*-PCR hybridization methods.

TERIALS AND METHODS

ing of YAC libraries

e Philadelphia genome center, two dimensional pools of the CEPH/ thon YAC libraries were constructed as described (68). A Biomek 1000 · workstation (Beckman Instruments) was used for yeast DNA isolation ·ling. In brief, yeast clones were grown to saturation in ura- trp- nedium in microtiter plates at 30°C. 50–75 µl of each clone was ıto a 1 ml deep-well plate (Beckman Instruments) in which spheroplast ıtion and lysis were performed as described elsewhere (69). The was extracted twice with Strataclean resin (Stratagene) according ıanufacturers recommendations. The DNA was then precipitated with opanol and the pellet was allowed to dry. After resuspension in TE (10 Tris–HCl pH 8.0, 1 mM EDTA, pH 8.0) and treatment with DNAase-RNAase, the DNA was precipitated with isopropanol and the pellet was and resuspended in water. Limited use was also made of commercially ased DNA pools constructed in three dimensional blocks equivalent to microtiter plates each (Research Genetics, Huntsville, Albama).

R was performed in 20 µl reactions using approximately 20 ng of pooled DNA in standard PCR buffer (1×buffer (Boehringher-Mannheim): 10 Tris–HCl, 1.5 mM Mg^{2+}, 50 mM KCl, pH 8.3) with 20 nM (final entration) primers and 0.5 U *Taq* polymerase (Perkin Elmer Cetus or ringher Mannheim). PCR conditions were: a 5 min denaturation step at followed by 45 cycles of 94°C for 20 s, annealing for 20 s, 72°C for , and a 7 min extension at 72°C. Suitable annealing temperatures were ined for each STS. The majority of the PCR assays were performed J Research PTC-100 thermal cyclers. Products were analyzed by gel rophoresis using 1.5% agarose.

Ss screened at the Whitehead Institute/MIT Center for Genome Research analyzed using a semi-automated system. The STSs were screened on s 709 to 972 of the CEPH mega-YAC library, generously provided by el Cohen. The YAC library was screened by a two-level pooling scheme. At the first level, there are 32 superpools consisting of DNA from the 768 YACs in a block of eight 96 well plates. Corresponding to each block, there are 8 row, 12 column, and 8 plate subpools. STSs positive at the superpool screen were then screened on the corresponding subpools to identify YAC addresses.

PCRs were prepared by a robotic station built by ROSYS and modified by IAS (Intelligent Automation Systems, Inc., Cambridge, MA). PCR was performed in 20 µl volumes containing 10 ng target DNA, 1×PCR Buffer (10 mM Tris–HCl, 50 mM KCl, 1.5 mM Mg^{2+}, and 0.001% gelatin), 4 nmol dNTP, 5 pmol each primer, and 0.5 U *Taq*. PCRs were completed on custom built thermocyclers (locally called waffle irons, by IAS) each having a capacity of 16 192 well plates (Costar, Cambridge MA). PCR conditions were: an initial 4 min denaturation at 94°C followed by 30 cycles of 50 s at 94°C, 1.5 min at 58°C, 1 min at 72°C, and a final extension period of 10 min at 72°C.

STSs were screened by either standard agarose gel stained with ethidium bromide or by high throughput chemiluminescence dot-blot analysis: The PCR products were transferred from the 192 well plates to nylon membranes using a custom built 96 pin pipettor (IAS) and a 6144 reaction capacity dot-blotting apparatus (96×16×4 well density, IAS). Subsequent hybridization and detection of the Hybond N+ membrane (Amersham) membranes was done using the ECL kit (Amersham). Hybridization was done overnight using non-radioactive probes designed from PCR products. STSs known to contain an internal repeat sequence such as CA or AGAT were probed with a molecule containing the repeat structure which had also been labelled with horseradish peroxidase (HRP). All blots were stringently washed with urea, 2×SSC and SDS at 42°C and detected using the standard ECL reagents. Computer images of each autoradiograph were obtained using a CCD camera. The VIEW software (Carl Rosenberg, Whitehead Institute) was used to locate and identify the positive dots, as well as to generate an intensity reading.

Fingerprint resolution of degenerate addresses

The STS screening on YAC pools yielded many degenerate YAC addresses, which occurred as a result of having more than one positive YAC per block of eight microtiter plates, from having one dimension in a two or three dimensional screen consistently fail to amplify, and from false positive results. These degenerate addresses represented a small set of addresses, from 2 to 12, of which usually one or two addresses contained the specific STS. We used fingerprint data to establish overlaps between the set of ambiguous YACs and the set of definite YACs. We applied a simple band-matching test to the CEPH-genethon fingerprint data set and declared pairs of clones with a statistically significant number of matching bands as overlapping. Parameters for declaring overlap were stringent, allowing resolution of only 1/3 of degenerate addresses. However, empirical testing of over 500 fingerprint resolved addresses from random STSs demonstrated that greater than 95% could be confirmed by testing the individual YAC DNAs.

Most YAC addresses obtained by screening the YAC pools, fingerprint analysis, and those derived from adjacent STSs during contig building were verified by testing DNA prepared from individual YACs in the library.

Construction of a chromosome 22 specific YAC library

DNA from hybrid cell line GM10888 (chromosome 22 in a Chinese hamster background) was used to create a chromosome 22 specific YAC library essentially as described (70). In brief, high molecular weight DNA from this cell line was partially digested with *Eco*RI and after ligation to pYAC4 was size selected on a 1% FMC Seaplaque GTG low melting agarose gel in a CHEF-DRII apparatus (BioRad). YACs containing human chromosome 22 DNA were identified by colony hybridization using total human DNA or human C_0t1 DNA as probes.

Contig assembly

Contig assembly was performed using a new software package written for use on SPARCstation Unix workstations (Sun Microsystems, Mountain View CA) in a combination of 'C', the logic programming language Prolog (SICStus Prolog, Swedish Institute of Computer Science, PO Box 1263, S-164 28 KISTA, Sweden), and the graphical user interface language Tcl/Tk (71). The algorithm is based on the technique of simulated annealing, used by a number of others for contig assembly (52,72); our implementation in particular is similar in broad outline to one developed by CEPH for this purpose (53). Briefly, in this technique a search space of probe (STS) order permutations, which would be intractable to explore exhaustively, is randomly reordered by selecting from a set of operations such as movement of single probes, swapping of probes, moving of clusters, and inversion of clusters. Any ordering is assigned a notional 'energy' that reflects its fit to the YAC–STS

data: our energy function involves examining the number and size of apparent gaps required in YACs to account for an ordering of STSs, i.e. positions where an expected STS hit is not observed, as well as arbitrary other objectives reflecting additional sources of information about probe order. The objective is to minimize this energy by accepting moves that reduce the overall energy. In order to avoid being trapped in a local energy minimum, the process takes place in the context of an abstract 'temperature'; a good energy minimum is sought by gradually 'cooling' the random search, so that the entire search space is accessible and poor local minima can be escaped, yet there is a gradual convergence (though it cannot be guaranteed that any one solution is optimal). The graphical user interface was designed for maximum interaction with the user, who has the option of reordering probes manually by any of the operations described above, or of asking the program to do so via simulated annealing, for the entire working probe set or any subregion. Islands of connected probe sets can be accumulated in a controlled fashion and with varying stringency as to degree of connectedness. These sets may then be winnowed based on a variety of heuristics to eliminate non-informative or doubtful probes, clones, or points. For example, adjacent or nearby wells with similar reactivities, likely to be due to cross-contamination, may be automatically combined, or YACs that appear to span non-continuous bins may be removed, etc. The contig assembly software may be obtained by sending a request by email to dsearls@cbil.humgen.upenn.edu.

ACKNOWLEDGEMENTS

The work undertaken in the Human Genome Center for Chromosome 22 in Philadelphia was supported by grant numbers P50-HG00425 (NCHGR) and CA39926 (NCI) from the NIH. Studies in the Whitehead Institute/MIT Center for Genome Research were supported by National Institute of Health Center for Genome Research Grant P50-HG00098. We wish to thank Eric Green and Glen Evans for screening for Washington University YACs, Eckart Meese and Marco Giovannini for providing STSs prior to publication, Daniel Cohen, Ilya Chumakov and Jean Weissenbach for the CEPH YAC libraries and the *Alu*-PCR generated chromosome 22 subset, and Willem Van Loon for biomek routines. Thomas Hudson is a recipient of a Clinician-Scientist Award from the Medical Research Council of Canada.

REFERENCES

1. Morton,N.E. (1991) Parameters of the human genome. *Proc. Natl Acad. Sci. USA* **88**, 7474–7476.
2. Nowell,P.C. and Hungerford,D.A. (1960) A minute chromosome in human chronic granulocytic leukemia. *Science* **132**, 1497–1499.
3. Rowley,J.D. (1973) A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature* **243**, 290–293.
4. Berger,R., Bernheim,A., Weh,H.J., Flandrin,G., Daniel,M.T., Brouet,J.C. and Colbert,N. (1979) A new translocation in Burkitt's tumor cells. *Hum. Genet.* **53**, 111–112.
5. Aurias,A., Rimbaut,C., Buffe,D., Dubousset,J. and Mazabraud,A. (1983) Chromosomal translocations in Ewing's sarcoma. *N. Engl. J. Med.* **309**, 496–497.
6. Turc-Carel,C., Philip,I., Berger,M.P., Philip,T. and Lenoir,G.M. (1983) Chromosomal translocations in Ewing's sarcoma. *N. Engl. J. Med.* **309**, 497–498.
7. Zang,K.D. (1982) Cytological and cytogenetical studies on human meningioma. *Cancer Genet. Cytogenet.* **6**, 249–274.
8. Dumanski,J.P., Carlbom,E., Collins,V.P. and Nordenskjold,M. (1987) Deletion mapping of a locus on human chromosome 22 involved in the oncogenesis of meningioma. *Proc. Natl Acad. Sci. USA* **84**, 9275–9279.
9. Seizinger,B.R., Martuza,R.L. and Gusella,J.F. (1986) Loss of genes on chromosome 22 in tumorigenesis of human acoustic neuroma. *Nature* **322**, 644–647.
10. Seizinger,B.R., Rouleau,G., Ozelius,L.J., Lane,A.H., ST. George-Hyslop,P., Huson,S., Gusella,J.F. and Martuza,R.L. (1987) Common pathogenetic mechanism for three tumor types in bilateral acoustic neurofibromatosis. *Science* **236**, 317–319.
11. Trofatter,J.A., MacCollin,M.M., Rutter,J.L., Murell,J.R., Duyao,M.P., Parry,D.M., Eldridge,R., Kley,N., Menon,A.G., Pulaski,K., Haase,V.H., Ambrose,C.M., Munroe,D., Bove,C., Haines,J.L., Martuza,R.L., MacDonald,M.E., Seizinger,B.R., Short,M.P., Buckler,A.J. and Gusella,J.F. (1993) A novel moesin-, ezrin-, radixin-like gene is a candidate for the neurofibromatosis 2 tumor suppressor. *Cell* **72**, 791–800.
12. Rouleau,G.A., Merel,P., Lutchman,M., Sanson,M., Zucman,J., Marineau,C., Hoang-Xuan,K., Demczuk,S., Desmaze,C., Plougastel,B., Pulst,S.M., Lenoir,G., Bijlsma,E., Fashold,R., Dumanski,J., de Jong,P., Parry,D., Eldridge,R., Aurias,A., Delattre,O. and Thomas,G. (1993) Alteration in a new gene encoding a putative membrane-organizing protein causes neuro-fibromatosis type 2. *Nature* **363**, 515–521.
13. Biegel,J.A., Rorke,L.B., Packer,R.J. and Emanuel,B.S. (1990) Monosomy 22 in rhabdoid or atypical tumors of the brain. *J. Neurosurg.* **73**, 710–714.
14. Biegel,J.A., Burk,C.D., Parmiter,A.H. and Emanuel,B.S. (1992) Molecular analysis of a partial deletion of 22q in a central nervous system rhabdoid tumor. *Genes Chromosom. Cancer* **5**, 104–108.
15. Zackai,E.H. and Emanuel,B.S. (1980) Site-specific reciprocal translocation, t(11;22) (q23;q11), in several unrelated families with 3:1 meiotic disjunction. *Am. J. Med. Genet.* **7**, 507–521.
16. Fraccaro,M., Lindsten,J., Ford,C.E. and Iselius,L. (1980) The 11q;22q translocation: a European collaborative analysis of 43 cases. *Hum. Genet.* **56**, 21–51.
17. De La Chapelle,A., Herva,R., Koivisto,M. and Aula,P. (1981) A deletion in chromosome 22 can cause DiGeorge syndrome. *Hum. Genet.* **57**, 253–256.
18. Kelley,R.I., Zackai,E.H., Emanuel,B.S., Kistenmacher,M., Greenberg,F. and Punnett,H.H. (1982) The association of the DiGeorge anomalad with partial monosomy of chromosome 22. *J. Pediatr.* **101**, 197–200.
19. Driscoll, D.A., Budarf, M.L., Emanuel, B.S. (1992) A genetic etiology for DiGeorge syndrome: Consistent deletions and microdeletions of 22q11. *Am. J. Hum. Genet.* **50**, 924–933
20. Carey,A.H., Roach,S., Williamson,R., Dumanski,J.P., Nordenskjold,M., Collins,V.P., Rouleau,G., Blin,N., Jalbert,P. and Scambler,P. (1990) Localization of 27 DNA markers to the region of human chromosome 22q11–pter deleted in patients with the DiGeorge syndrome and duplicated in the der22 syndrome. *Genomics* **7**, 299–306.
21. Driscoll, D.A., Spinner, N.B., Budarf, M.L., McDonald-McGinn, D.M., Zackai, E.H., Goldberg, R.B., Shprintzen, R.J., Saal, H.M., Zonana, J., Jones, M.C., Mascarello, J.T., Emanuel, B.S. (1992) Deletions and microdeletions of 22q11.2 in velo-cardio-facial syndrome. *Am. J. Med. Genet.* **44**, 261–268
22. McDermid,H.E., Duncan,A.M.V., Brasch,K.R., Holden,J.J.A., Magenis,E., Sheehy,R., Burn,J., Kardon,N., Noel,B., Schinzel,A., Teshima,I. and White,B.N. (1986) Characterization of the supernumery chromosome in cat eye syndrome. *Science* **232**, 646–648.
23. Olson,M., Hood,L., Cantor,C. and Botstein,D. (1989) A common language for physical mapping of the human genome. *Science* **245**, 1434–1435.
24. Green,E.D. and Olson,M. (1990) Chromosomal region of the cystic fibrosis gene in yeast artificial chromosomes: a model for human genome mapping. *Science* **250**, 94–98.
25. Green,E.D. and Green,P. (1991) Sequence-tagged site (STS) content mapping of human chromosomes: theoretical considerations and early experiences. *PCR Methods Applic.* **1**, 77–90.
26. Budarf, M.L., Eckman, B., Michaud, D., Buetow, K.H., Williams, S., McDermid, H., Goldmuntz, E., Gavigan, S., Meese, E., Biegel, J., Dumanski, J., Bell, C.J. and Emanuel, B.S. (1994) Regional localization of over 300 loci on human chromosome 22 with an extended regional mapping panel. Submitted.
27. Hudson, T.J., Colbert, A.M.E., Reeve, M.P., Bae, J.S., Lee, M.K., Nussbaum, R.L., Budarf, M.L., Emanuel, B.S. and Foote, S. (1994) Isolation and regional mapping of 110 chromosome 22 STSs. *Genomics*, in press.
28. Burke,D.T., Carle,G.F. and Olson,M.V. (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* **236**, 806–812.
29. Albertsen,H.M., Abderrahim,H., Cann,H.M., Dausset,J., Le Paslier,D. and Cohen,D. (1990) Construction and characterization of a yeast artificial chromosome library containing seven haploid human genome equivalents. *Proc. Natl Acad. Sci. USA* **87**, 4256–4260.
30. Chumakov,I., Rigault,P., Guillou,S., Ougen,P., Billaut,A., Guasconi,G., Gervy,P., LeGall,I., Soularue,P., Grinas,L., Bougueleret,L., Bellane-Chantelot,C., Lacroix,B., Barillot,E., Gesnouin,P., Pook,S., Vaysseix,G., Frelat,G., Schmitz,A., Sambucy,J., Bosch,A., Estivill,X., Weissenbach,J., Vignal,A., Reithman,H., Cox,D., Patterson,D., Gardiner,K., Hattori,M., Sakaki,Y., Ichikawa,H., Ohki,M., Le Paslier,D., Heilig,R., Antonarakis,S. and Cohen,D. (1992) Continuum of overlapping clones spanning the entire human chromosome 21q. *Nature* **359**, 380–387.
31. Brownstein,B.H., Silverman,G.A., Little,R.D., Burke,D.T., Korsmeyer,S.J., Schlessinger,D. and Olson,M.V. (1989) Isolation of

single-copy human genes from a library of yeast artificial chromosome clones. *Science* **244**, 1348–1351.
32. Weber,J.L. and May,P.E. (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum.Genet.* **44**, 388–396.
33. Wilcox, A.S., Khan,A.S., Hopkins,J.A. and Sikela,J.M. (1991) Use of 3' untranslated sequences of human cDNAs for rapid chromosome assignment and conversion to STSs: implications for an expression map of the genome. *Nucleic Acids Res.* **19**, 1837–1843.
34. Adams,M.D., Kelley,J.M., Gocayne,J.D., Dubnick,M., Polymeropoulos,M.H., Xiao,H., Merril,C.R., Wu,A., Olde,B. and Moreno,R.F. (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* **252**, 1651–1656.
35. Riley,J., Butler,R., Ogilvie,D.J., Finniear,R., Jenner,D., Anand,R., Smith,J.C. and Markham,A.F. (1990) A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. *Nucleic Acids Res.* **18**, 2887–2890.
36. Nelson,D.L., Ledbetter,S.A., Corbo,L., Victoria,M.F., Ramirez-Solis,R., Webster,T.D., Ledbetter,D.H. and Caskey,C.T. (1989) *Alu* polymerase chain reaction: a method for rapid isolation of human-specific sequences from complex DNA sources. *Proc. Natl Acad. Sci. USA* **86**, 6686–6690.
37. Lichter, P., Ledbetter, S.A., Ledbetter, D.H. and Ward, D.C. (1980) Fluorescence *in situ* hybridization with *Alu* and L1 polymerase chain reaction probes for rapid characterization of human chromosomes in hybrid cell lines. *Proc. Natl Acad. Sci. USA* **87**, 6634–6638.
38. Emanuel, B.S., Driscoll, D., Goldmuntz, E., Baldwin, S., Biegel, J., Zackai, E.H., McDonald-McGinn, D., Sellinger, B., Gorman, N., Williams, S., and Budarf, M.L. (1993) Molecular and phenotypic analysis of the chromosome 22 microdeletion syndromes. In Epstein, C.J. (ed.), *Phenotypic Mapping of Down Syndrome and Other Aneuploid Conditions*. Wiley Liss, New York, NY, 207–224.
39. Frazer, K.A., Boehnke, M., Budarf, M.L., Wolff, R.K., Emanuel, B.S., Myers, R.M. and Cox, D.R. (1992) A radiation hybrid map of the region on human chromosome 22 containing the neurofibromatosis type 2 locus. *Genomics* **14**, 574–584.
40. Geurts van Kessel, A.H.M., Westerveld, A., de Groot, P.G., Meera Khan, P. and Hagemeijer, A. (1980) Regional localization of the genes coding for human ACO2, ARSA, and NAGA on chromosome 22. *Cytogenet. Cell. Genet.* **28**, 169–172.
41. Ledbetter, D.H., Rich, D.C., O'Connell, P., Leppert, M. and Carey, J.C. (1989) Precise localization of NF1 to 17q11.2 by balanced translocation. *Am. J. Hum. Genet.* **44**, 20–24.
42. Budarf, M.L., Sellinger, B., Griffin, C., Emanuel, B.S. (1989) Comparative mapping of the constitutional and tumor associated 11:22 translocations. *Am. J. Hum. Genet.* **45**, 128–139.
43. Delattre, O., Azambuja, C.J., Aurias, A., Zucman, J., Peter, M., Zhang, F., Hors-Cayla, M.C., Rouleau, G., and Thomas, G. (1991) Mapping of human chromosome 22 with a panel of somatic cell hybrids. *Genomics* **9**, 721–727.
44. Erikson, J., Griffin, C., ar-Rushdi, A., Valtieri, M., Hoxie, J., Finan, J., Emanuel, B.S., Rovera, G., Nowell, P.C., Croce, C.M. (1986) Heterogeneity of chromosome 22 breakpoint in Ph-positive acute lymphocytic leukemia. *Proc. Natl Acad. Sci. USA* **83**, 1807–1811.
45. Geurts van Kessel, A.H.M., Tetteroo, P.A.T., von dem Borne, A.E.G.Kr., Hagemeijer, A. and Bootsma, D. (1983) Expression of human myeloid-associated surface antigens in human-mouse myeloid cell hybriuds. *Proc. Natl Acad. Sci. USA* **80**, 3748–3752.
46. Bauer,T.R., McDermid,H.E., Budarf, M.L., Van Keuren, M.L. and Bloomberg, B.B. (1993) Physical location of the human immunoglobulin lambda-like genes 14.1, 16.1 and 16.2. *Immunogenetics* **38**, 387–399.
47. Croce, C.M., Huebner, K., Isobe, M., Fainstein, E., Lifshitz, B., Shtivelman, E., Canaani, E. (1987) Mapping of the four distinct BCR-related loci to chromosome region 22q11: order of BCR loci relative to chronic myelogeneous leukemia and acute lymphoblastic leukemia breakpoints. *Proc. Natl Acad. Sci USA* **84**, 7174–7178.
48. Scambler, P.J. (1994) Report of the Fourth International Workshop on Human Chromosome 22 Mapping. *Cytogenet. Cell Genet.* **67**, 277–319.
49. Figlewicz, D.A., Delattre, O., Guellaen, G., Krizus, A., Thomas, G., Zucman,J., and Rouleau, G.A. (1993) Mapping of human γ-glutamyl transpeptidase genes on chromosome 22 and other autosomes. *Genomics* **17**, 299–305.
50. Heisterkamp,N. and Groffen,J. (1988) Duplication of the bcr and gamma-glutamyl transpeptidase genes. *Nucleic Acids Res.* **16**, 8045–8056.
51. Budarf, M.L., Canaani, E. and Emanuel, B.S. (1988) Linear order of the four BCR-related loci in 22q11. *Genomics* **3**, 168–172.
52. Cuticchia,A.J., Arnold,J. and Timberlake,W.E. (1992) The use of simulated annealing in chromosome reconstruction experiments based on binary scoring. *Genetics* **132**, 591–601.
53. Rigault, P. (1993) In Lim,H.A., Fickett,J., Cantor,C.R. and Robbins,R.J. (eds) *Clone Ordering by Simulated Annealing: Application to the STS-Content Map of Chromosome 21*. Proceedings of the Second International Conference on Bioinformatics, Supercomputing, and Complex Genome Analysis. World Scientific Publishing: 169–183.
54. McDermid,H.E., Budarf,M.L. and Emanuel,B.S. (1993) Long-range restriction map of human chromosome 22q11–22q12 between the lambda immunoglobulin locus and the Ewing sarcoma breakpoint. *Genomics* **18**, 308–318.
55. Buetow,K.H., Duggan,D., Yang,B., Ludwigsen,S., Puck,J., Porter,J., Budarf,M., Spielman,R. and Emanuel,B.S. (1993) A microsatellite-based multipoint index map of human chromosome 22. *Genomics* **18**, 329–339.
56. Shizuya,H., Birren,B., Kim,U.J., Mancino,V., Slepak,T., Tachiiri,Y. and Simon,M. (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. *Proc. Natl Acad. Sci. USA* **89**, 8794–8797.
57. Sternberg,N. (1990) Bacteriophage P1 cloning system for the isolation, amplification, and recovery of DNA fragments as large as 100 kilobase pairs. *Proc. Natl Acad. Sci. USA* **87**, 103–107.
58. Ioannou,P.A., Amemiya,C.T., Garnes,J., Kroisel,P.M., Shizuya,H., Chen,C., Batzer,M.A. and de Jong,P.J. (1994) A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nature Genet.* **6**, 84–89.
59. Arratia,R., Lander,E., Tavare,S. and Waterman,M. (1992) Genomic mapping by anchoring random clones: a mathematical analysis. *Genomics* **11**, 806–827.
60. Cohen,D., Chumakov,I. and Weissenbach,J. (1993) A first-generation physical map of the human genome. *Nature* **366**, 698–701.
61. Weissenbach, J., Gyapay, G., Dib, C., Vignal, A., Morissette, J., Millasseau, P., Vaysseix, G. and Lathrop M. (1992) A second-generation linkage map of the human genome. *Nature* **359**, 777–778.
62. Gyapay, G., Morrisette, J., Vignal, A., Dib, C., Fizames, C., Millaseau, P., Marc, S., Bernardi, G., Lathrop, M. and Weissenbach, J. (1994) The 1993–94 genethon human genetic linkage map. *Nature Genet.* **7**, 246–339.
63. Barillot,E., Lacroix,B. and Cohen,D. (1991) Theoretical analysis of library screening using a N-dimensional pooling strategy. *Nucleic Acids Res.* **19**, 6241–6247.
64. Riazi, M.A., Mears, A.J., Bell, C.J., Budarf, M.L., Emanuel, B.S., Murray, J.C., Patil, S.R., and McDermid, H.E.(1994) Long range mapping and construction of a YAC contig within the cat eye syndrome critical region. *Am. J. Hum. Genet.* **55**, A268.
65. Driscoll, D.A., Salvin, J., Sellinger, B., McGinn-McDonald, D., Zackai, E.H., Emanuel, B.S. (1993) Prevalence of 22q11 microdeletions in DGS and VCFS: implications for genetic counseling and prenatal diagnosis. *J. Med. Genet.* **30**, 813–817.
66. Lindblom, A., Sandelin,K., Iselius,L., Dumanski,J., White,I., Nordenskjold,M. and Larsson, C. (1994) Predisposition for breast cancer in carriers of constitutional translocation 11q;22q. *Am.J. Hum. Genet.* **54**, 871–876.
67. Peyard,M., Fransson,I., Xie, Y.-G., Han, F.-Y., Ruttledge,M.H., Swahn,S., Collins,J.E., Dunham,I., Collins,V.P., and Dumanski,J.P. (1994) Characterization of a new member of the human β-adaptin gene family from chromosome 22q12, a candidate meningioma gene. *Hum. Mol. Genet.*3, 1393–1399.
68. Amemiya,C.T., Alegria-Hartman,M.J., Aslanidis,C., Chen,C., Nikolic,J., Gingrich,J.C. and de Jong,P.J. (1992) A two-dimensional YAC pooling strategy for library screening via STS and *Alu*-PCR methods. *Nucleic Acids Res.* **20**, 2559–2563.
69. Green,E.D and Olson,M.V. (1990) Systematic screening of yeast artificial-chromosome libraries by use of the polymerase chain reaction. *Proc. Natl Acad. Sci. USA* **87**, 1213–1217.
70. Lee,J.T., Murgia,A., Sosnoski,D.M., Olivos,I.M. and Nussbaum,R.L. (1992) Construction and characterization of a yeast artificial chromosome library for Xpter–Xq27.3: a systematic determination of cocloning rate and X-chromosome representation. *Genomics* **12**, 526–533.
71. Ousterhout,J.K. (1994) *Tcl and the Tk Toolkit*. Addison-Wesley, Reading, MA.
72. Mott,R., Grigoriev,A., Maier,E., Hoheisel,J. and Lehrach,H. (1993) Algorithms and software tools for ordering clone libraries: application to the mapping of the genome of Schizosaccharomyces pombe. *Nucleic Acids Res.* **21**, 1965–1974.

Human Molecular Genetics, 1995, Vol. 4, No. 1 59–69

Integration of physical, breakpoint and genetic maps of chromosome 22. Localization of 587 yeast artificial chromosomes with 238 mapped markers

Callum J.Bell*, Marcia L.Budarf, Bart W.Nieuwenhuijsen[1], Barry L.Barnoski, Kenneth H.Buetow[2], Keely Campbell, Angela M.E.Colbert[3], Joelle Collins, Mark Daly[3] Philippe R.Desjardins[1], Todd DeZwaan[1], Barbara Eckman[1], Simon Foote[3,+], Kyle Hart[1], Kevin Hiester[1], Marius J.Van Het Hoog[1], Elizabeth Hopper, Alan Kaufman[3], Heather E.McDermid[4], G.Christian Overton[1], Mary Pat Reeve[3], David B.Searls[1], Lincoln Stein[3], Vinay H.Valmiki[1], Edward Watson, Sloan Williams, Rachel Winston[1], Robert L.Nussbaum[1,§], Eric S.Lander[3], Kenneth H.Fischbeck[1], Beverly S.Emanuel and Thomas J.Hudson[3]

Children's Hospital of Philadelphia, Division of Human Genetics and Molecular Biology, 34th Street and Civic Center Boulevard, Philadelphia, PA 19104, [1]University of Pennsylvania School of Medicine, 415 Curie Boulevard, Philadelphia, PA 19104-6146, [2]Fox Chase Cancer Center, 7701 Burholme Avenue, Philadelphia, PA 19111-2412, [3]Center for Genome Research, Whitehead Institute for Biological Sciences/Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, MA 02142, USA and [4]Department of Biological Sciences, University of Alberta, Edmonton, Alberta T6G 2E9, Canada

Received October 17, 1994; Revised and Accepted November 2, 1994

Detailed physical maps of the human genome are important resources for the identification and isolation of disease genes and for studying the structure and function of the genome. We used data from STS content mapping of YACs and natural and induced chromosomal breakpoints to anchor contigs of overlapping yeast artificial chromosome (YAC) clones spanning extensive regions of human chromosome 22. The STSs were assigned to specific regions (bins) on the chromosome using cell lines from a somatic hybrid mapping panel defining a maximum of 25 intervals. YAC libraries were screened by PCR amplification of hierarchical pools of yeast DNA with 238 markers, and a total of 587 YAC clones were identified. These YACs were assembled into contigs based upon their shared STS content using a simulated annealing algorithm. Fifteen contigs, containing between 2 and 74 STSs were assembled, and ordered along the chromosome based upon the cytogenetic breakpoint, meiotic and PFG maps. Additional singleton YACs were assigned to unique chromosomal bins. These ordered YAC contigs will be useful for identifying disease genes and chromosomal breakpoints by positional cloning and will provide the foundation for higher resolution physical maps for large scale sequencing of the chromosome.

INTRODUCTION

Human chromosome 22 constitutes approximately 1.9% of the haploid autosomal genome (1). Clinical disorders associated with this chromosome include several acquired, tumor-related translocations such as the t(9;22) of chronic myelogenous leukemia and acute lymphocytic leukemia (2,3), the t(8;22) variant translocation of Burkitt's lymphoma (4) and the t(11; 22) of Ewing's sarcoma (5,6). Deletions of all or part of chromosome 22 are associated with meningiomas (7,8), acoustic neuromas (9,10), Neurofibromatosis type 2 (NF2) (11,12), and rhabdoid tumors (13,14). Further, chromosome 22 is also involved in the only recurrent non-Robertsonian constitutional chromosomal translocation in humans (15,16). In addition, a number of syndromes are caused by deletions or duplications of portions of 22q11, including DiGeorge syndrome (17–20), velo-cardio-facial syndrome (21), and cat-eye syndrome (22). Chromosome 22 has a high gene density and contains many duplicated sequences and gene families, which makes it an interesting model for mapping studies. The identification of

*To whom correspondence should be addressed

Present addresses: [+]Walter and Eliza Hall Institute of Medical Research, PO Royal Melbourne Hospital, Parkville 3050, Australia and [§]Genetic Disease Research NCHGR/NIH, 9000 Rockville Pike, 49/4A72, Bethesda, MD 20892, USA

new disease genes will be facilitated by the integration of detailed genetic and physical maps of this chromosome. Moreover, integrated maps can be used to make sequence-ready DNA templates, to facilitate the identification of novel structural elements and to study chromosome structure.

We have used STS-content mapping (23–25) to assemble contigs representing most of the chromosome. 316 STSs and 22 hybridization probes were developed by our centers and outside investigators. Markers suspected of containing repeats, and others giving unsatisfactory results in control experiments were eliminated. The remainder were localized by PCR or Southern hybridization to 'bins', which are defined by breakpoints in a somatic cell hybrid mapping panel (26,27). These markers were used to identify YACs (28) in four libraries: the CEPH/Genethon YAC libraries (29,30), a chromosome 22-only hybrid cell line derived YAC library, and the Washington University YAC library (31). 216 STSs and 22 hybridization probes identified a total of 587 individual YACs which were then assembled into 15 contigs containing between 2 and 196 YACs.

RESULTS

Marker generation

The STSs and hybridization probes used in this study were derived from genes and other sequences in the public domain,

Table 1. Loci used to identify YACs[a]

Bin interval	LOCI AND NUMBER OF YACS DETECTED
1.1	ATP6E(8), D22S9(9), D22S789(1), D22S795(5), D22S24(7), D22S50(3), F8VWPF(3), IGKVP3(3)
1.2	D22S43(3)
2	D22S57(2), D22S137(3), D22S420(5), D22S111(1)
3-5	COMT(8), D22S139(2), D22S259(2)
6	D22S184(9), KI-1547(5)
7	BCRL2(1), GGTX(1)
8.1	CRKL(1), D22S117(1), D22S264(4), D22S306(5), D22S308(7), HCF2(11), MEST39(4)
9	BCRL4(3), D22S303(4), D22S425(2), D22S563(1), D22S655(7), GGTY(3), GNAZ(4), IGL@(2), IGLC2(4),
10	BCR(5)
11	ADORA1(10), D22S156(1), D22S794(3), MMP11(2)
12	BCRL3(3), CRYBB2(6), D22S1(8), D22S33(16), D22D42(20), D22S56(20), D22S72(12),D22S186(8), D22S190(16), D22S192(11), D22S193(4), D22S310(14), D22S315(7), D22S351(15), D22S419(7), D22S431(4), EWSex5(1), GGTZ(3), TOP1P2(5), YESP(7),
13	D22S268(12), D22S300(9), D22S447(3), EWS3'(1), LIF(4), MEST14(1), NEFH(7), OSM(1), TCN2(4)
14	D22S37(6), D22S776(3), EN38(5)
12–14	D22S541(2), D22S557(2), D22S560(9), D22S564(7), D22S566(20), D22S568(6), D22S569(15),D22S570(22), D22S571(12), D22S572(6),D22S574(5), D22S576(19), D22S582(1), D22S584(3), D22S588(1), D22S589(5), D22S591(3), D22S594(14), D2S596(4), D22S604(4),D22S615(8), D22S631(1), D22S635(9), D22S638(11), D22S642(9),D22S650(4), D22S653(3), D22S669(14), D22S745(12)
15	22-5(2), D22S15(7), D22S28(4), D22S29(15), D22S38(13), D22S44(14), D22S47(9), D22S51(4), D22S58(2), D22S60(13), D22S61(3), D22S91(1), D22S102(3), D22S273(1), D11S277(4), D22S278(5), D22S280(5), D22S281(8), D22S283(2), D22S292(2), D22S304(7), D22S412(1), D22S778(1), D22S792(1), D22S793(3), D22S799(1), IL2RB(1), PVALB(2), SGLT1(13), TIMP3(1)
16.1	D22S299(12), D22S302(6), PDGFB(1)
16.2	D22S279(9)
15–16.2	D22S540(9), D22S544(4), D22S552(4), D22S554(8),D22S561(10),D22S577(7), D22S579(6), D22S595(12), D22S607(4), D22S617(6),D22S620(1), D22S623(4), D22S624(5), D22S629(7), D22S630(2), D22S633(7), D22S639(8), D22S644(1), D22S652(9), D22S739(10), D22S746(7)
17	CYP2D8P(6), D22S307(5), G22P1(17), NAGA(6)
18	D22S417(1), DIA1(4),
19	ACR(7)D22S270(2), D22S418(2)
17–19	D22S565(3)
20	BZRP(5), D22S64(3), D22S282(1), D22S297(5)
21	D22S40(5), D22S274(6), D22S294(4), D22S781(1), FIBB(2),
22	ARSA(1), CB10(1), D22S23(1), D22S45(4), D22S55(2), D22S63(7), D22S295(9)
21–22	Z7(8)
20–22	D22S611(2)
Not Binned	Cos7-1(5), D22S272(7), D22S427(2), D22S543(12), D22S553(1), D22S559(13), D22S562(6), D22S618(2), D22S627(1), D22S656(4), GGT1(16)**

D22S556(3), D22S609(1), D22S626(2)

D22S567(14)

D22S301(7), IGLLbb1(12)

D22S258(7), EWSR1(1)

D22S1191(1)

D22S546(1)

[a]The bins defined by the somatic cell hybrid panel are indicated on the left. Loci localized to bins or ranges of bins are shown in boxes. The vertical extent of each box indicates the bin interval to which the loci were mapped. In parentheses are the numbers of YACs detected at each locus.

**GGT1-like sequences are found in bins 7, 9 and 12.

STRPs (simple tandem repeat polymorphisms) (32), expressed sequence tags (ESTs) (33,34), YAC vector-insert junction fragments (35), inter-*Alu* PCR fragments (36) and randomly sequenced plasmid clones (26,27). The loci at which STSs or probes were generated are shown in Table 1.

Chromosomal bin assignment of markers

Markers were assigned to chromosomal 'bin' locations by Southern blot hybridization or PCR analysis of DNA from cell lines in a 26 member somatic cell hybrid panel . These cell lines define 22 bins shown schematically in Figure 1. Three of these bins are each further subdivided into two sub-bins, making a total of 25 intervals. The majority of the hybrids have been previously described: GM10888 (37); Cl-6-2/EG, Cl-21-5/CV, Cl-9/GM05878 (38); Rad-110a; Rad 37a (39); GM11220 = X/22 33-TG, GM11224C = 1/22AM-6, GM11223C = 1/22AM-27 (40); GM11685 (41); Cl-4/GB, Cl-1-1/TW (42); AJO 9, APR 8.5 (43); 514 AA2 (44); WESP-2A-TG8 = GM11221 (45); RAJ5BE (46); D6S5 (47). There are eight additional members of the hybrid panel (Cl-3/5878; Cl-1/5878; Cl-2/5878; Cl-8/5878; Cl-15-1/PB; Cl-21-2/PB; Cl-2/DIBA; Cl-8-1/AMB6) which will be further described in another manuscript (26). Not all markers were assigned to a unique bin. STSs binned in the Whitehead Institute/MIT Genome Center were tested on a subset of six somatic cell hybrid lines (Fig. 1), whereas those binned in Philadelphia were tested on the complete panel. A small number of STSs could not be uniquely assigned for technical reasons.

Contained within this hybrid mapping panel are the breakpoints which have been designated by the chromosome 22 mapping community as anchor positions in the physical map. The anchor panel was recently updated (48) and now

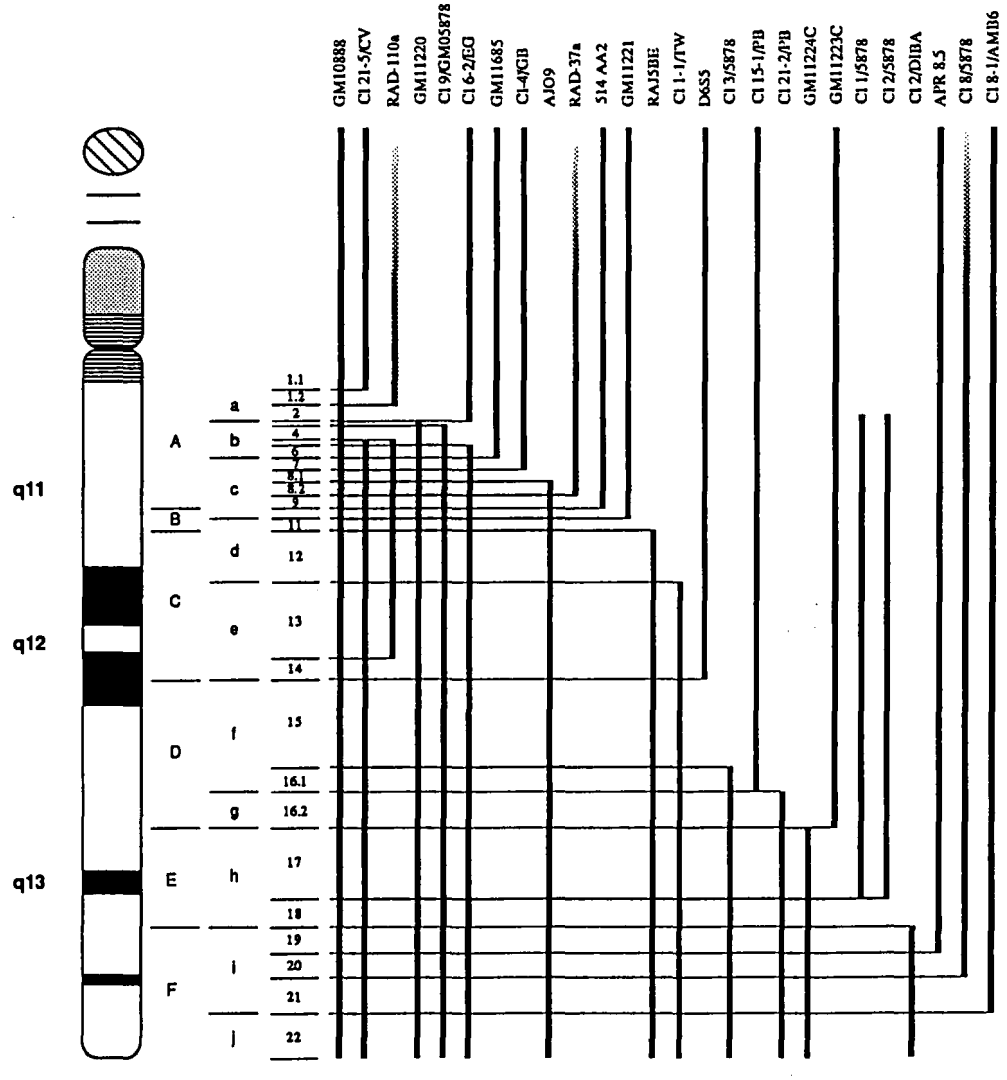


Figure 1. Schematic of the somatic cell hybrid mapping panel used for bin assignment of markers. The heavy vertical black lines represent the segments of chromosome 22 retained in each hybrid. Shaded portions indicate that the extent of the p arm retained in a hybrid is unknown. The names of the hybrids are shown at the top of the figure. The fine horizontal lines indicate the breakpoints that divide the chromosome into 25 intervals, shown numbered from 1.1 through 22 (three 'bins'are further subdivided into two sub-bins). The upper case letters A–F indicate a subset of the 26 member panel defining six intervals that was used for bin assignment at the Whitehead Institute/MIT Genome Center. The full high resolution panel was used at the Children's Hospital of Philadelphia. The lower case letters a–j show the 10 interval panels defined by the 11 hybrid cell lines available from the NIGMS repository (48).

divides chromosome 22q into a total of 10 intervals which represents a subset of this mapping panel. The hybrids defining the anchor points are: GM11220, GM11685, GM11221, GM11222C, GM11224C, GM11223C, D6S5, Cl-15-1/PB (GM13498), Cl-21-2/PB (GM13499), Cl-2/DIBA (GM13501) and Cl-8-1/AMB6 (GM13500). The somatic cell hybrids defining the anchor points of chromosome 22 are available through the NIGMS genetic mutant cell repository, Camden, New Jersey. Since the chromosome 22 reference hybrids represent a subset of the panel utilized for binning markers described in this manuscript, the data presented here can be easily assessed by other groups in order to position markers which they have mapped using the reference panel.

YAC identification

Most YACs in the study were identified in the CEPH/Genethon libraries [original library (29) with an average insert size of 470 kb and mega-YAC library with an average insert size of 0.9 Mbp (30)], by PCR screening of yeast DNAs pooled in two or three dimensions. Additional YACs were isolated from the Washington University YAC library (31), and from a chromosome 22 specific YAC library constructed with DNA from hybrid cell line GM10888 (chromosome 22 in a Chinese hamster background). The chromosome 22 specific YAC library contains approximately 300 YACs with an average insert size of 200 kb, equivalent to 1× coverage of the chromosome. YACs isolated from the Washington University library were kindly provided by collaborators. In addition, limited use was made of a subset of YACs, kindly provided by Ilya Chumakov and Daniel Cohen, identified by hybridization of *Alu*-PCR products of a chromosome 22-only somatic cell hybrid to the CEPH mega-YAC library. YACs from this subset, and from the chromosome 22-specific library were identified by colony hybridization.

Table 1 shows a summary of the YAC screening results. The left-most column shows the bin intervals, numbered 1.1–22. The relative positions of the bins on the chromosome are displayed visually in Figure 1. Loci that were used to identify YACs are shown in boxes in the body of the table; the vertical extent of each box indicates the bin, or range of bins, to which each locus was mapped by referring to the left-most column, and the number of YACs detected by each locus is indicated in parentheses after the locus name. The majority of these results are YACs identified to single microtiter plate addresses, either from unequivocal PCR results in two or three dimensional screens, or from confirmatory PCR tests done on individual YACs. A YAC address consists of three dimensions: plate, row, and column. In initial screening of YAC pools, many of the addresses were incomplete (missing a dimension), or had more than one possible value in a dimension, which occurs when there is more than one positive YAC per block of eight microtiter plates (see Materials and Methods), or from false positive results. Such ambiguous addresses were resolved by several means including fingerprint analysis, comparison with verified YAC addresses of adjacent STSs, or PCR of all possible clones in the degenerate set of addresses. After preliminary contig assembly, most of the clones identified as well as the putative adjacent YACs were individually tested with each STS in the contig.

GGTX, GGTY and GGTZ (Table 1) refer to probes containing sequences homologous to γ-glutamyl transpeptidase 1 (GGT1) (49). These three GGT-like sequences have been shown to be physically linked to the BCR (break point cluster)-like sequences BCRL2 and BCRL4, and to BCR itself, respectively, in 22q11 (50). These BCR-like sequences contain polymorphic *Hin*dIII sites and thus can be distinguished from each other (51), allowing assignment of the YACs detected by the GGT1 STS to be allocated to unique bins. Details of this study will be presented in a separate publication.

Primer sequences for each STS and YAC addresses may be found in the public FTP (file transfer protocol) sites of the Philadelphia (cbil.humgen.upenn.edu/pub/22/) and the Whitehead Institute/MIT (genome.wi.mit.edu /distribution/human_STS_releases/) Genome Centers. World Wide Web access is available through HTTP://www.cis.upenn.edu/~cbil/chr22db/chr22dbhome.html and HTTP://www-genome.wi.mit.edu.

In order to resolve confusion caused by possible cross-contamination among microtiter plate wells we adopted two approaches. The first approach compared the CEPH/Genethon fingerprints, where available, of the putative YAC positives with the fingerprints of other YACs known by STS content to overlap the YAC to be resolved. Shared fingerprint bands

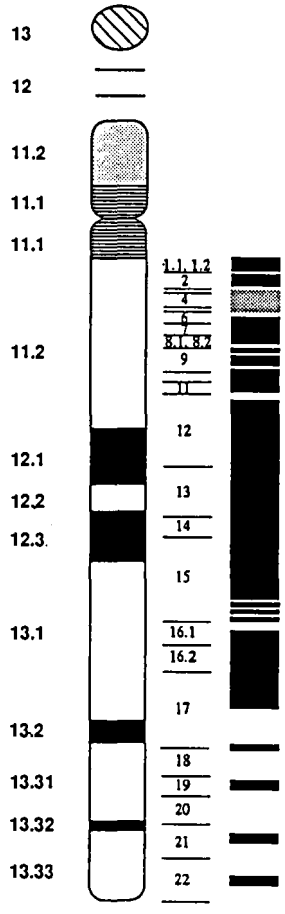


Figure 2. Estimated coverage of the chromosome in contigs. The horizontal lines are the boundaries separating 25 intervals. Contigs are shown as blocks. The stippled block shows the location of a cosmid contig encompassing the DiGeorge critical region (DGCR).

among these YACs identified with a high degree of confidence the true positive YAC address among several neighboring candidates in several cases. The second approach was based on a calculation of the actual distances between wells of two YAC addresses sharing STSs, divided by the number of STS hits in common; when this measure fell below a certain threshold for any pair of addresses, they were consolidated into a single address. This heuristic in all cases corresponded well to human judgments about likely cross-contamination, and was shown to be justified in cases that were checked experimentally. Level 1 data from the CEPH/Genethon genome mapping project were confirmed and included in Table 1.

YAC contig assembly

To date, we have used 238 markers to identify 587 YACs. The YACs and STSs fall into 15 islands, defined as sets of STSs and sets of YACs all of which can be reached from each other by following a path of connectivity alternating between STSs and YACs. Singleton YACs detected by one STS each, numbering 25, are omitted from this total. Although the number of YACs we identified indicates nearly 5× coverage of the chromosome, the depth of coverage is uneven: all somatic cell hybrid bins contain YACs, but the 22q11.23−q12.31 region (bins 12–15; see below) has much deeper coverage than

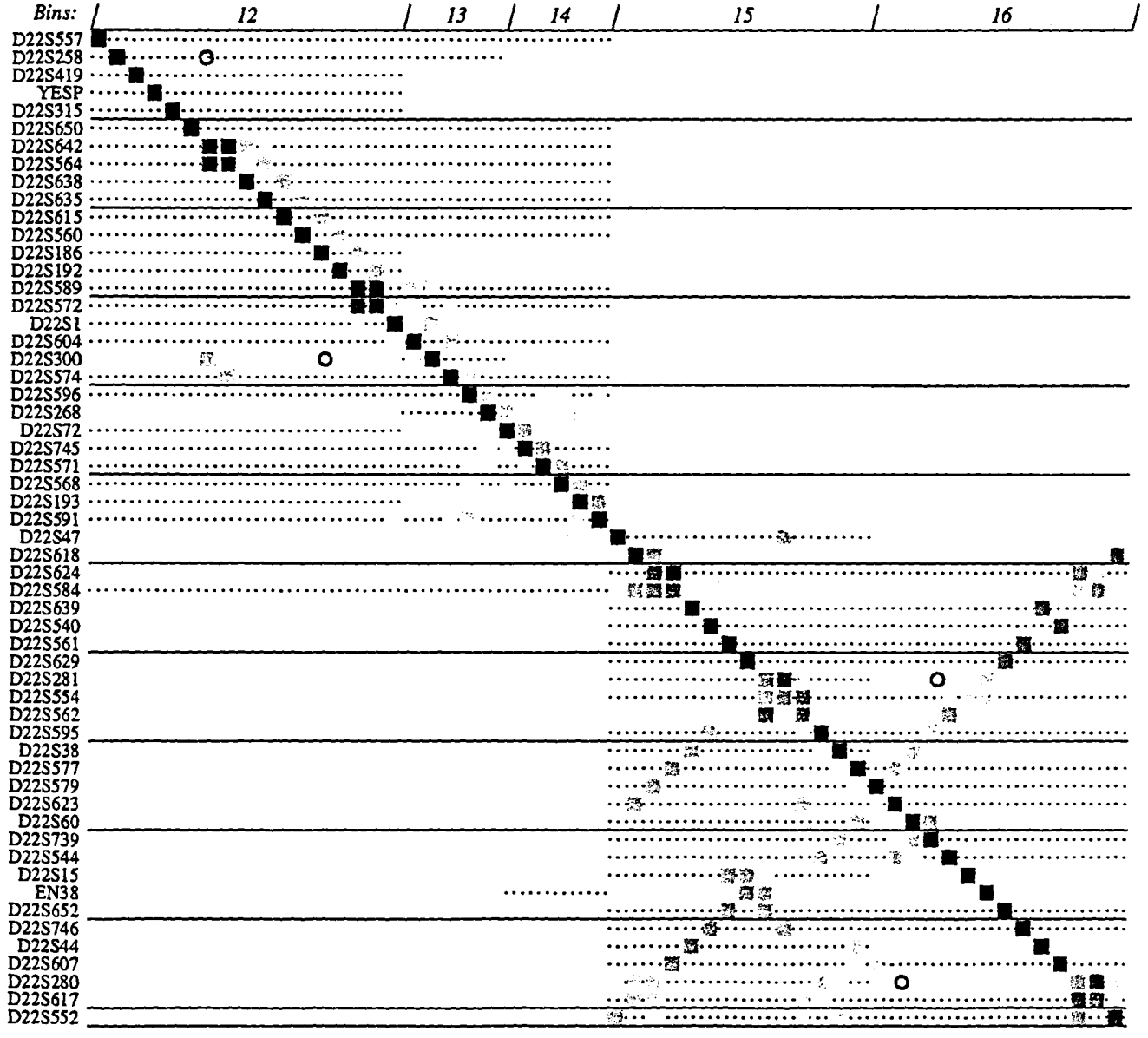


Figure 3. Searls plot of simulated annealing data for the largest contig accumulated from multiple runs of the program. The list of loci down the left of the figure is the 'minimum energy' ordering of markers (see the text for detailed explanation). Gray boxes indicate the position on the horizontal axis at which the indicated STSs occurred during individual runs. Darker boxes indicate that an STS was positioned in the same location in multiple runs. Boxes falling repeatedly on the diagonal indicate high confidence in the minimum energy ordering. Horizontal dotted lines indicate the chromosomal bin location of each STS. The bin intervals are shown at the top of the figure. Circles indicate the consensus positions of markers that are present on the meiotic map.

elsewhere. We had difficulty obtaining unequivocal clone and STS order within the largest of these islands, and a clear clone tiling path, even with deep YAC coverage of the area and many STSs. In the central portion of the chromosome YAC connectivity has been achieved over a distance exceeding 10 Mb, yet an unbroken clone tiling path remains elusive despite extensive testing of YACs versus STSs in that region. This may be due in part to false positive and negative YAC/STS results (although results have been carefully confirmed), internal deletions within YAC clones, and sequences present at more than one location on the chromosome. Given these problems, the objective becomes to find an ordering of STSs that minimizes gaps. In ideal data, there should be an order of STSs, corresponding to a true YAC contig, such that there are no such gaps. However, in our data all postulated orders of STSs in an island result in some number of 'gaps' within YACs in the island, defined as cases where a YAC is negative for some STS but positive for STSs located to both the left and right in the ordering.

For very large islands, finding the STS order with the absolute minimum number of gaps is computationally intractable, but several approaches have been developed to finding approximate solutions. A simulated annealing (52,53) program we developed employs a random search strategy that seeks local energy minima in the space of all possible orderings, where energy is defined in terms of numbers and sizes of gaps (see Materials and Methods). This approach can be expected to yield somewhat different results for multiple runs, both because there may be more than one valid ordering even for ideal data, and because for 'noisy' data the search may find different local energy minima which are near the actual optimum. In practice, the results of multiple runs of simulated annealing are generally similar, although not identical. We refer to these orderings of STSs and YACs as contigs, though it should be emphasized that the larger islands should be viewed as putative contigs at present.

A schematic representation of the coverage of the chromosome in contigs is shown in Figure 2. The chromosome is shown divided into 25 intervals derived from the somatic cell hybrid map of Budarf *et al.* (26). Bin 1 formally includes the

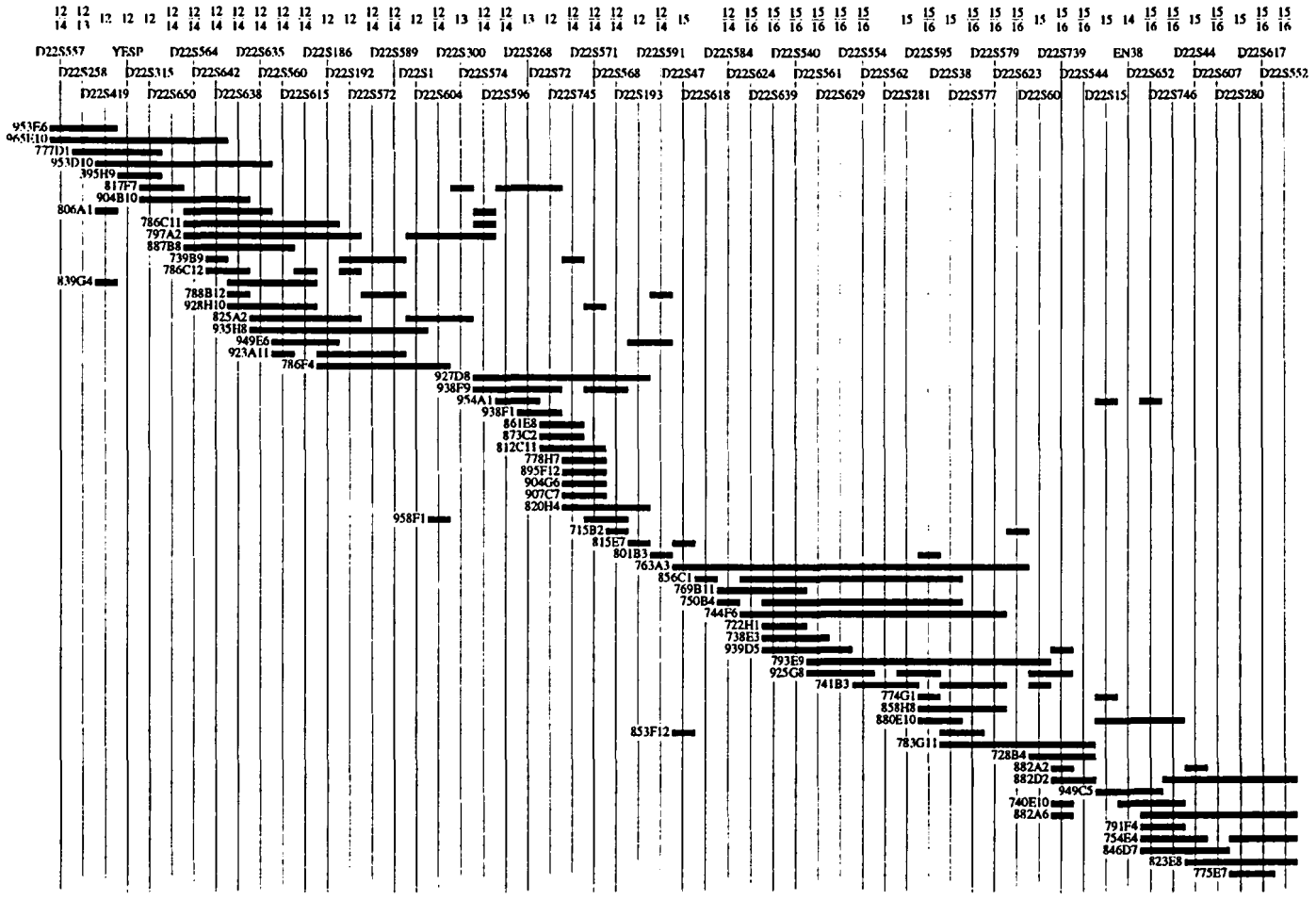


Figure 4. A single solution for the largest contig in the central region of chromosome 22q. The contig was constructed as follows: YACs and STSs were selected by connectivity to D22S1, obeying the double linkage rule. Singletons (YACs detected by one STS only) were then eliminated, as were markers that detected more than 14 YACs. Singletons were eliminated a second time, and the resulting set of markers and YACs were subjected to simulated annealing. Marker order is shown along the top of the figure. Above each marker name is the bin interval that the marker was mapped to, e.g. 15/16 indicates the marker is in bin 15–16. YACs are shown as heavy horizontal black lines.

short arm of the chromosome. The contigs, based on the bin assignment of the STSs that detected the YACs in each, are shown as dark blocks. Since STS content mapping provides only limited information on contig size, the true extent of coverage and the sizes of the gaps separating the contigs are unknown. The stippled block represents a contig of cosmids in a region that proved difficult to clone in YACs. YACs detected by STSs in this part of the chromosome were unstable, and were underrepresented in the libraries screened (M.Budarf, unpublished observations). The cosmid map of this region will be described in a separate publication. Figure 2 makes clear the low coverage of the distal portion of the chromosome. This arises in part from the lower density of markers but is largely due to underrepresentation of the region in the mega-YAC library.

Figure 3 shows simulated annealing results for the largest contig, using a novel method of representing such data to which we have given the name 'Searls plot', after the author of the program. As noted, results of simulated annealing tend toward local minima of the objective function that may differ among runs. The relative merits of these STS orderings and implied YAC contigs cannot be judged with confidence on the basis of the STS data alone. On the other hand, a number of such orderings independently arrived at may be expected to represent a reasonable sampling of the contours of the search space of possible STS orderings. If the predicted orderings do not resemble each other, then little can be said about which is closest to the true optimum, but if they are all similar, one may be more confident in their consensus. Figure 3 shows the degree and nature of the consensus for multiple simulated annealings. The minimum energy ordering among all runs is indicated by the list of STSs running down the left hand side. The gray boxes in the diagram show the positions along the horizontal axis at which the indicated STS occurs in a run, so that the major diagonal denotes complete agreement with the minimum energy run. Other gray boxes indicate positions at which that STS occurred in other runs, and the shading of a box reflects the number of times a particular STS occurred at the same position in a run. If the predictions for an STS tend to cluster at more than one position in multiple runs, one may infer that the evidence is not strong enough to greatly favor one position over another, though it may be possible to narrow the possibilities to a few regions.

As noted above, even with ideal data it may be possible to have more than one ordering, particularly over subregions of the contig. Obviously, a given ordering of STSs may be reversed in its entirety, without changing the apparent fit to the YAC data in isolation, and for that reason each simulated annealing run is reversed, if necessary, to more closely approach the consensus. However, there may also be subregions over which the STSs can be reversed without affecting the energy materially, and in this case the Searls plot will display a characteristic 'X' pattern across the diagonal, representing the alternative orderings. Another characteristic pattern is a displacement of a subregion laterally on the plot, with either a forward or reversed directionality, indicating parts of the contig that display local integrity but which can be moved elsewhere in the larger scheme of things, with little or no penalty. Finally, there are subregions where STSs tend to be in proximity to each other, but where there is little support for ordering them with respect to each other. This may occur, for example, where there are multiple YACs with the same STS hits, but no YACs with only partial overlap to split the STSs and provide order information. These appear as 'clouds' of points at or near the diagonal; it can be seen that with a sufficient sample size such regions would approach a uniform distribution of points within a diffuse 'superblock'. Figure 3 shows a major 'X' indicating that the ordering in the distal half of the contig was inverted in a significant number of the simulated annealing runs. We interpret this to mean that the link between D22S591 and D22S47 should be viewed with caution. We have yet to confirm by other means whether actual continuity of YAC coverage exists in this region.

Figure 3 suggests, with some confidence, a general ordering of STSs in most sections of this region of the chromosome, but in some areas there is significant scatter. Some of this deviation is systematic in nature, as described in the previous paragraph, and some in all likelihood merely reflects regions where the data is error-prone. An external test of the accuracy of this method is provided not only by bin information but by the meiotic and pulsed-field gel maps (54,55) of the region; the orders of the subsets of markers in both of these maps are similar in the converged order arrived at by simulated annealing, which in this case was done without regard to information from any of these other methods. Figure 4 shows a single simulated annealing solution to the largest contig.

DISCUSSION

We used physical, breakpoint, and meiotic maps of human chromosome 22 to localize contigs of overlapping YAC clones that provide extensive coverage of the long arm of the chromosome. The physical map is developing rapidly due to considerable new data obtained by screening YAC libraries with STSs. The contigs, most of which are anchored by landmarks that have been ordered by meiotic or hybrid mapping, provide extensive coverage of the long arm of the chromosome. Although long range continuity of the contigs is not yet complete, the present information is of immediate use to the human genetic mapping community for identifying disease genes and chromosomal breakpoints. The current state of the physical map reported here reflects the fundamental characteristics of the reagents and methods used, as well as the inherent nature of chromosome 22 itself.

STSs that were developed for chromosome 22 are not randomly distributed along the chromosome. The contig spanning interval 22q11.2–q13.1 is the most evolved as the result of the high density of markers in this region and greater than average representation of the region in the YAC libraries. The distribution of markers shows a bias towards the center of the long arm of the chromosome (26,27). This is partly because many STRP markers were used as STSs, and these are known to be concentrated in the 22q12 Giemsa-dark chromosomal band (55). However, it is not known why other randomly chosen STSs generated from flow-sorted material should also be biased in this way. The distal third of the long arm is correspondingly poor in STSs, and appears to be underrepresented in the YAC libraries, and as a consequence, contains only two small contigs and seven singleton YACs. Interestingly, the distal portion of the long arm appears to be resistant to cloning in both plasmid and YAC libraries, and the consequent paucity of mapping information indicates the need for alternat-

ive strategies for covering this region. Currently, we are targetting the region by generating STSs from inter-*Alu* plasmid libraries made from radiation hybrid cell lines that retain only the distal portion of the chromosome. Success in developing new STSs in this way has shown that YACs, not markers, are likely to be limiting for YAC–STS contig mapping, and that complete coverage of this region will probably depend on a different cloning vehicle. Current candidate systems are bacterial artificial chromosomes (BACs) (56), P1 phage clones (57), P1 artificial chromosomes (PACs) (58), and cosmids.

Screening multi-dimensional pools of YACs was the only practical way to test all 25,000 mega-YACs for the presence or absence of a given STS, but created several types of problems. Contamination of adjacent wells during preparation of the pools, absence of amplification in one dimension, or the presence of more than one positive YAC in the same pool were examples of difficulties that are inherent to pooling schemes which can result in false positive, false negative, and ambiguous YAC addresses. Most of the results obtained from the pool screenings have been resolved by a variety of methods, including analysis of YACs seen with adjacent STSs, fingerprint analysis of selected YACs, and ultimately, the verification of the PCR on the individual YAC. To decrease the errors caused by false negatives on STS order, most STSs were screened on adjacent YACs as well.

The CEPH mega-YACs, which have an average insert size of 0.9 Mb (30), provided the best tool for linking STSs and assembling contigs, and were screened with all available markers. By requiring double linkage (59) before declaring contiguity among STSs in the largest contig, large clones were required, and YACs from the other libraries, while contributing to deep coverage in most regions, did not, in general, contribute to contig assembly. However, in some notable cases contig construction was dependent upon the smaller clones, and as the map matures, they will be useful in resolving the order of closely spaced STSs, and as tools for isolating cosmids or other smaller clones as the map moves towards a higher level of resolution required for eventual sequencing.

In addition to the known families of chromosome 22 specific repeats on long arm, such as the BCR, immunoglobulin and GGT gene families, we observed several markers which appear to behave as low copy repeats. In such cases, the PCR assay amplifies two identical or related sequences with products of similar molecular weights. Examples of this were D22S33 and D22S275, which gave several bands of similar size, and detected 15 and 14 YACs respectively. Repetitive STSs created inconsistencies in the data, manifested as large apparent gaps in YAC clones, since contig assembly software tries to assign them single contig locations. In fact, they may be present at two or more locations. Repeats therefore artificially connect YACs at disparate locations. We arbitrarily decided that STSs detecting 14 or more mega-YACs would be declared potentially repetitive and excluded them from contig construction.

The CEPH-Genethon tiling paths (60), provided relatively little additional information because the areas covered by tiling paths coincided with the region where the STS physical map was already well covered. We independently screened the mega-YAC library for the same Genethon genetic markers (61,62), and confirmed the YAC addresses and the level-1 tiling paths present in the November 1993 CEPH-Genethon data release (60). We extracted a few YAC addresses derived by *Alu*-PCR hybridizations in 22q11.2–q13.1 region that were missed during YAC pool screening. Unfortunately, the areas where the STS content map was poor were also not represented in the tiling paths, or present only in higher level paths that could not be confirmed. Fingerprint analysis on the mega-YACs generated by CEPH (60,63), was used to resolve ambiguous addresses derived from screening pools of YACs. This method, successful in one third of ambiguous addresses tested, reduced the number of alternate addresses that need to be verified for YAC determination. We did try to assemble the 22q11.2–q13.1 contig by fingerprint analysis alone using only the fingerprints of YACs that were previously identified to this region. The results had only limited success, yielding small contigs with less than 10 YACs that were already shown to have extensive overlap in STS content.

We chose to represent the data for the large contig in two ways: a single simulated annealing solution, and the Searls plot, derived from multiple runs of simulated annealing. These representations, combined with the YAC–STS results shown in Table 1, provide an objective and useful means of using these data. Previous localization of markers by recombination or breakpoints greatly facilitated the evaluation of the STS content map. The marker order in region 22q11.2–q13.1, spanning more than 11 cM, was broadly consistent with the orders of subsets of markers arrived at by meiotic and pulsed-field gel mapping (54,55). The smaller contigs contain, at most, two genetically ordered markers, which does not allow real comparisons of marker order with the meiotic map. In essence, we have made the assumption that the framework linkage map (55) is correct, and used it to anchor and orient the smaller contigs. The best validation of the smaller contigs came from concordance with the somatic cell hybrid binning results.

It is clear that, due to problems inherent to YACs, the STS-content mapping results from the large contig did not allow us to obtain a fine structure order of the region. This may well be true for many other regions in the genome. The need for additional methodologies to obtain a finer scaffold map of STSs is evident. Radiation hybrids, which allow the study of multiple, larger DNA fragments at a higher redundancy may provide more confidence in generating a high resolution STS order. They will also allow contiguity of the STS map in regions where YAC clones are few or absent.

The contigs reported in this paper will facilitate the study of several disease-related and structural regions of interest on chromosome 22. The YACs that have been localized to bins 1 and 2 (22q11.1–q11.2) form contigs that almost completely cover the cat eye syndrome critical region (CECR). The most proximal of these will allow us to address the problem of defining the physical boundary of the centromere. The completion of a contig containing all of the CECR, facilitated by a pulsed field gel electrophoresis map (64), will permit detailed transcription mapping of the region as a first approach to defining genes that contribute to this syndrome. In the region distal to the CECR, 22q11.2, difficulty was encountered in obtaining stable YACs in bins 3, 4 and 5. YACs identified in this region were frequently smaller than the mean insert sizes of the libraries, indicating that they contained deletions, and several probes failed to detect YACs. These bins represent the DiGeorge syndrome commonly deleted region which is notably unstable in humans giving rise to the deletions seen

atients with DiGeorge syndrome and velo-cardio-facial drome (65). It is interesting to note that these, and other uences mapping to the sites of frequent chromosomal angements in cat eye syndrome are also unstable when ed in yeast. Further characterization of these sequences allow us to investigate the possible causes of instability.

e constitutional t(11;22) translocation breakpoint is the recurrent, non-Robertsonian, constitutional translocation umans (15,16), and defines the boundary between bins 7 8. Contigs spanning this region may help in revealing ctural features on the chromosome that underlie this angement, as well as the identification of genes suspected nvolvement in breast cancer tumorigenesis (66). Identifica- of clones that span the t(11;22) breakpoint has been plicated by the presence of several duplicated regions in 11 which include the GGT and BCRL loci. In addition to e known ancestral duplications, STS screening results gest the presence of other low-copy repeat families that e the construction of a contiguous clone map of 22q11 icularly challenging. The largest contig, connecting bins and 15 (22q11.2–q13.1) contains several interesting fea- s that have already been well characterized, including the ing's sarcoma breakpoint (5,6), the NF2 gene (11,12) and candidate meningioma gene β-adaptin (67).

n conclusion, the physical map of human chromosome 22 advanced considerably, due to the large scale screening of CEPH mega-YAC library with chromosome 22 specific s, and several regions of interest are now contained within C contigs. Current efforts to achieve a complete set of rlapping clones for the long arm of the chromosome are cted at the generation of additional STSs for clone screen- , as well as targeted strategies for the distal third of the mosome using *Alu*-PCR hybridization methods.

TERIALS AND METHODS

ling of YAC libraries

he Philadelphia genome center, two dimensional pools of the CEPH/ ethon YAC libraries were constructed as described (68). A Biomek 1000 ic workstation (Beckman Instruments) was used for yeast DNA isolation oling. In brief, yeast clones were grown to saturation in ura- trp- medium in microtiter plates at 30°C. 50–75 µl of each clone was nto a 1 ml deep-well plate (Beckman Instruments) in which spheroplast ation and lysis were performed as described elsewhere (69). The was extracted twice with Strataclean resin (Stratagene) according anufacturers recommendations. The DNA was then precipitated with ropanol and the pellet was allowed to dry. After resuspension in TE (10 Tris–HCl pH 8.0, 1 mM EDTA, pH 8.0) and treatment with DNAase-RNAase, the DNA was precipitated with isopropanol and the pellet was and resuspended in water. Limited use was also made of commercially hased DNA pools constructed in three dimensional blocks equivalent to t microtiter plates each (Research Genetics, Huntsville, Albama).

CR was performed in 20 µl reactions using approximately 20 ng of pooled t DNA in standard PCR buffer (1×buffer (Boehringher-Mannheim): 10 Tris–HCl, 1.5 mM Mg^{2+}, 50 mM KCl, pH 8.3) with 20 nM (final entration) primers and 0.5 U *Taq* polymerase (Perkin Elmer Cetus or hringher Mannheim). PCR conditions were: a 5 min denaturation step at followed by 45 cycles of 94°C for 20 s, annealing for 20 s, 72°C for , and a 7 min extension at 72°C. Suitable annealing temperatures were rmined for each STS. The majority of the PCR assays were performed J Research PTC-100 thermal cyclers. Products were analyzed by gel trophoresis using 1.5% agarose.

Ss screened at the Whitehead Institute/MIT Center for Genome Research analyzed using a semi-automated system. The STSs were screened on s 709 to 972 of the CEPH mega-YAC library, generously provided by iel Cohen. The YAC library was screened by a two-level pooling scheme. At the first level, there are 32 superpools consisting of DNA from the 768 YACs in a block of eight 96 well plates. Corresponding to each block, there are 8 row, 12 column, and 8 plate subpools. STSs positive at the superpool screen were then screened on the corresponding subpools to identify YAC addresses.

PCRs were prepared by a robotic station built by ROSYS and modified by IAS (Intelligent Automation Systems, Inc., Cambridge, MA). PCR was performed in 20 µl volumes containing 10 ng target DNA, 1×PCR Buffer (10 mM Tris–HCl, 50 mM KCl, 1.5 mM Mg^{2+}, and 0.001% gelatin), 4 nmol dNTP, 5 pmol each primer, and 0.5 U *Taq*. PCRs were completed on custom built thermocyclers (locally called waffle irons, by IAS) each having a capacity of 16 192 well plates (Costar, Cambridge MA). PCR conditions were: an initial 4 min denaturation at 94°C followed by 30 cycles of 50 s at 94°C, 1.5 min at 58°C, 1 min at 72°C, and a final extension period of 10 min at 72°C.

STSs were screened by either standard agarose gel stained with ethidium bromide or by high throughput chemiluminescence dot-blot analysis: The PCR products were transferred from the 192 well plates to nylon membranes using a custom built 96 pin pipettor (IAS) and a 6144 reaction capacity dot-blotting apparatus (96×16×4 well density, IAS). Subsequent hybridization and detection of the Hybond N+ membrane (Amersham) membranes was done using the ECL kit (Amersham). Hybridization was done overnight using non-radioactive probes designed from PCR products. STSs known to contain an internal repeat sequence such as CA or AGAT were probed with a molecule containing the repeat structure which had also been labelled with horseradish peroxidase (HRP). All blots were stringently washed with urea, 2×SSC and SDS at 42°C and detected using the standard ECL reagents. Computer images of each autoradiograph were obtained using a CCD camera. The VIEW software (Carl Rosenberg, Whitehead Institute) was used to locate and identify the positive dots, as well as to generate an intensity reading.

Fingerprint resolution of degenerate addresses

The STS screening on YAC pools yielded many degenerate YAC addresses, which occurred as a result of having more than one positive YAC per block of eight microtiter plates, from having one dimension in a two or three dimensional screen consistently fail to amplify, and from false positive results. These degenerate addresses represented a small set of addresses, from 2 to 12, of which usually one or two addresses contained the specific STS. We used fingerprint data to establish overlaps between the set of ambiguous YACs and the set of definite YACs. We applied a simple band-matching test to the CEPH-genethon fingerprint data set and declared pairs of clones with a statistically significant number of matching bands as overlapping. Parameters for declaring overlap were stringent, allowing resolution of only 1/3 of degenerate addresses. However, empirical testing of over 500 fingerprint resolved addresses from random STSs demonstrated that greater than 95% could be confirmed by testing the individual YAC DNAs.

Most YAC addresses obtained by screening the YAC pools, fingerprint analysis, and those derived from adjacent STSs during contig building were verified by testing DNA prepared from individual YACs in the library.

Construction of a chromosome 22 specific YAC library

DNA from hybrid cell line GM10888 (chromosome 22 in a Chinese hamster background) was used to create a chromosome 22 specific YAC library essentially as described (70). In brief, high molecular weight DNA from this cell line was partially digested with *Eco*RI and after ligation to pYAC4 was size selected on a 1% FMC Seaplaque GTG low melting agarose gel in a CHEF-DRII apparatus (BioRad). YACs containing human chromosome 22 DNA were identified by colony hybridization using total human DNA or human C_0t1 DNA as probes.

Contig assembly

Contig assembly was performed using a new software package written for use on SPARCstation Unix workstations (Sun Microsystems, Mountain View CA) in a combination of 'C', the logic programming language Prolog (SICStus Prolog, Swedish Institute of Computer Science, PO Box 1263, S-164 28 KISTA, Sweden), and the graphical user interface language Tcl/Tk (71). The algorithm is based on the technique of simulated annealing, used by a number of others for contig assembly (52,72); our implementation in particular is similar in broad outline to one developed by CEPH for this purpose (53). Briefly, in this technique a search space of probe (STS) order permutations, which would be intractable to explore exhaustively, is randomly reordered by selecting from a set of operations such as movement of single probes, swapping of probes, moving of clusters, and inversion of clusters. Any ordering is assigned a notional 'energy' that reflects its fit to the YAC–STS

data: our energy function involves examining the number and size of apparent gaps required in YACs to account for an ordering of STSs, i.e. positions where an expected STS hit is not observed, as well as arbitrary other objectives reflecting additional sources of information about probe order. The objective is to minimize this energy by accepting moves that reduce the overall energy. In order to avoid being trapped in a local energy minimum, the process takes place in the context of an abstract 'temperature'; a good energy minimum is sought by gradually 'cooling' the random search, so that the entire search space is accessible and poor local minima can be escaped, yet there is a gradual convergence (though it cannot be guaranteed that any one solution is optimal). The graphical user interface was designed for maximum interaction with the user, who has the option of reordering probes manually by any of the operations described above, or of asking the program to do so via simulated annealing, for the entire working probe set or any subregion. Islands of connected probe sets can be accumulated in a controlled fashion and with varying stringency as to degree of connectedness. These sets may then be winnowed based on a variety of heuristics to eliminate non-informative or doubtful probes, clones, or points. For example, adjacent or nearby wells with similar reactivities, likely to be due to cross-contamination, may be automatically combined, or YACs that appear to span non-continuous bins may be removed, etc. The contig assembly software may be obtained by sending a request by email to dsearls@cbil.humgen.upenn.edu.

ACKNOWLEDGEMENTS

The work undertaken in the Human Genome Center for Chromosome 22 in Philadelphia was supported by grant numbers P50-HG00425 (NCHGR) and CA39926 (NCI) from the NIH. Studies in the Whitehead Institute/MIT Center for Genome Research were supported by National Institute of Health Center for Genome Research Grant P50-HG00098. We wish to thank Eric Green and Glen Evans for screening for Washington University YACs, Eckart Meese and Marco Giovannini for providing STSs prior to publication, Daniel Cohen, Ilya Chumakov and Jean Weissenbach for the CEPH YAC libraries and the *Alu*-PCR generated chromosome 22 subset, and Willem Van Loon for biomek routines. Thomas Hudson is a recipient of a Clinician-Scientist Award from the Medical Research Council of Canada.

REFERENCES

1. Morton,N.E. (1991) Parameters of the human genome. *Proc. Natl Acad. Sci. USA* **88**, 7474–7476.
2. Nowell,P.C. and Hungerford,D.A. (1960) A minute chromosome in human chronic granulocytic leukemia. *Science* **132**, 1497–1499.
3. Rowley,J.D. (1973) A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature* **243**, 290–293.
4. Berger,R., Bernheim,A., Weh,H.J., Flandrin,G., Daniel,M.T., Brouet,J.C. and Colbert,N. (1979) A new translocation in Burkitt's tumor cells. *Hum. Genet.* **53**, 111–112.
5. Aurias,A., Rimbaut,C., Buffe,D., Dubousset,J. and Mazabraud,A. (1983) Chromosomal translocations in Ewing's sarcoma. *N. Engl. J. Med.* **309**, 496–497.
6. Turc-Carel,C., Philip,I., Berger,M.P., Philip,T. and Lenoir,G.M. (1983) Chromosomal translocations in Ewing's sarcoma. *N. Engl. J. Med.* **309**, 497–498.
7. Zang,K.D. (1982) Cytological and cytogenetical studies on human meningioma. *Cancer Genet. Cytogenet.* **6**, 249–274.
8. Dumanski,J.P., Carlbom,E., Collins,V.P. and Nordenskjold,M. (1987) Deletion mapping of a locus on human chromosome 22 involved in the oncogenesis of meningioma. *Proc. Natl Acad. Sci. USA* **84**, 9275–9279.
9. Seizinger,B.R., Martuza,R.L. and Gusella,J.F. (1986) Loss of genes on chromosome 22 in tumorigenesis of human acoustic neuroma. *Nature* **322**, 644–647.
10. Seizinger,B.R., Rouleau,G., Ozelius,L.J., Lane,A.H., ST. George-Hyslop,P., Huson,S., Gusella,J.F. and Martuza,R.L. (1987) Common pathogenetic mechanism for three tumor types in bilateral acoustic neurofibromatosis. *Science* **236**, 317–319.
11. Trofatter,J.A., MacCollin,M.M., Rutter,J.L., Murell,J.R., Duyao,M.P., Parry,D.M., Eldridge,R., Kley,N., Menon,A.G., Pulaski,K., Haase,V.H., Ambrose,C.M., Munroe,D., Bove,C., Haines,J.L., Martuza,R.L., MacDonald,M.E., Seizinger,B.R., Short,M.P., Buckler,A.J. and Gusella,J.F. (1993) A novel moesin-, ezrin-, radixin-like gene is a candidate for the neurofibromatosis 2 tumor suppressor. *Cell* **72**, 791–800.
12. Rouleau,G.A., Merel,P., Lutchman,M., Sanson,M., Zucman,J., Marineau,C., Hoang-Xuan,K., Demczuk,S., Desmaze,C., Plougastel,B., Pulst,S.M., Lenoir,G., Bijlsma,E., Fashold,R., Dumanski,J., de Jong,P., Parry,D., Eldridge,R., Aurias,A., Delattre,O. and Thomas,G. (1993) Alteration in a new gene encoding a putative membrane-organizing protein causes neuro-fibromatosis type 2. *Nature* **363**, 515–521.
13. Biegel,J.A., Rorke,L.B., Packer,R.J. and Emanuel,B.S. (1990) Monosomy 22 in rhabdoid or atypical tumors of the brain. *J. Neurosurg.* **73**, 710–714.
14. Biegel,J.A., Burk,C.D., Parmiter,A.H. and Emanuel,B.S. (1992) Molecular analysis of a partial deletion of 22q in a central nervous system rhabdoid tumor. *Genes Chromosom. Cancer* **5**, 104–108.
15. Zackai,E.H. and Emanuel,B.S. (1980) Site-specific reciprocal translocation, t(11;22) (q23;q11), in several unrelated families with 3:1 meiotic disjunction. *Am. J. Med. Genet.* **7**, 507–521.
16. Fraccaro,M., Lindsten,J., Ford,C.E. and Iselius,L. (1980) The 11q;22q translocation: a European collaborative analysis of 43 cases. *Hum. Genet.* **56**, 21–51.
17. De La Chapelle,A., Herva,R., Koivisto,M. and Aula,P. (1981) A deletion in chromosome 22 can cause DiGeorge syndrome. *Hum. Genet.* **57**, 253–256.
18. Kelley,R.I., Zackai,E.H., Emanuel,B.S., Kistenmacher,M., Greenberg,F. and Punnett,H.H. (1982) The association of the DiGeorge anomalad with partial monosomy of chromosome 22. *J. Pediatr.* **101**, 197–200.
19. Driscoll, D.A., Budarf, M.L., Emanuel, B.S. (1992) A genetic etiology for DiGeorge syndrome: Consistent deletions and microdeletions of 22q11. *Am. J. Hum. Genet.* **50**, 924–933
20. Carey,A.H., Roach,S., Williamson,R., Dumanski,J.P., Nordenskjold,M., Collins,V.P., Rouleau,G., Blin,N., Jalbert,P. and Scambler,P. (1990) Localization of 27 DNA markers to the region of human chromosome 22q11–pter deleted in patients with the DiGeorge syndrome and duplicated in the der22 syndrome. *Genomics* **7**, 299–306.
21. Driscoll, D.A., Spinner, N.B., Budarf, M.L., McDonald-McGinn, D.M., Zackai, E.H., Goldberg, R.B., Shprintzen, R.J., Saal, H.M., Zonana, J., Jones, M.C., Mascarello, J.T., Emanuel, B.S. (1992) Deletions and microdeletions of 22q11.2 in velo-cardio-facial syndrome. *Am. J. Med. Genet.* **44**, 261–268
22. McDermid,H.E., Duncan,A.M.V., Brasch,K.R., Holden,J.J.A., Magenis,E., Sheehy,R., Burn,J., Kardon,N., Noel,B., Schinzel,A., Teshima,I. and White,B.N. (1986) Characterization of the supernumery chromosome in cat eye syndrome. *Science* **232**, 646–648.
23. Olson,M., Hood,L., Cantor,C. and Botstein,D. (1989) A common language for physical mapping of the human genome. *Science* **245**, 1434–1435.
24. Green,E.D. and Olson,M. (1990) Chromosomal region of the cystic fibrosis gene in yeast artificial chromosomes: a model for human genome mapping. *Science* **250**, 94–98.
25. Green,E.D. and Green,P. (1991) Sequence-tagged site (STS) content mapping of human chromosomes: theoretical considerations and early experiences. *PCR Methods Applic.* **1**, 77–90.
26. Budarf, M.L., Eckman, B., Michaud, D., Buetow, K.H., Williams, S., McDermid, H., Goldmuntz, E., Gavigan, S., Meese, E., Biegel, J., Dumanski, J., Bell, C.J. and Emanuel, B.S. (1994) Regional localization of over 300 loci on human chromosome 22 with an extended regional mapping panel. Submitted.
27. Hudson, T.J., Colbert, A.M.E., Reeve, M.P., Bae, J.S., Lee, M.K., Nussbaum, R.L., Budarf, M.L., Emanuel, B.S. and Foote, S. (1994) Isolation and regional mapping of 110 chromosome 22 STSs.*Genomics*, in press.
28. Burke,D.T., Carle,G.F. and Olson,M.V. (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* **236**, 806–812.
29. Albertsen,H.M., Abderrahim,H., Cann,H.M., Dausset,J., Le Paslier,D. and Cohen,D. (1990) Construction and characterization of a yeast artificial chromosome library containing seven haploid human genome equivalents. *Proc. Natl Acad. Sci. USA* **87**, 4256–4260.
30. Chumakov,I., Rigault,P., Guillou,S., Ougen,P., Billaut,A., Guasconi,G., Gervy,P., LeGall,I., Soularue,P., Grinas,L., Bougueleret,L., Bellane-Chantelot,C., Lacroix,B., Barillot,E., Gesnouin,P., Pook,S., Vaysseix,G., Frelat,G., Schmitz,A., Sambucy,J., Bosch,A., Estivill,X., Weissenbach,J., Vignal,A., Reithman,H., Cox,D., Patterson,D., Gardiner,K., Hattori,M., Sakaki,Y., Ichikawa,H., Ohki,M., Le Paslier,D., Heilig,R., Antonarakis,S. and Cohen,D. (1992) Continuum of overlapping clones spanning the entire human chromosome 21q. *Nature* **359**, 380–387.
31. Brownstein,B.H., Silverman,G.A., Little,R.D., Burke,D.T., Korsmeyer,S.J., Schlessinger,D. and Olson,M.V. (1989) Isolation of

single-copy human genes from a library of yeast artificial chromosome clones. *Science* **244**, 1348–1351.
32. Weber,J.L. and May,P.E. (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum.Genet.* **44**, 388–396.
33. Wilcox, A.S., Khan,A.S., Hopkins,J.A. and Sikela,J.M. (1991) Use of 3' untranslated sequences of human cDNAs for rapid chromosome assignment and conversion to STSs: implications for an expression map of the genome. *Nucleic Acids Res.* **19**, 1837–1843.
34. Adams,M.D., Kelley,J.M., Gocayne,J.D., Dubnick,M., Polymeropoulos,M.H., Xiao,H., Merril,C.R., Wu,A., Olde,B. and Moreno,R.F. (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* **252**, 1651–1656.
35. Riley,J., Butler,R., Ogilvie,D.J., Finniear,R., Jenner,D., Anand,R., Smith,J.C. and Markham,A.F. (1990) A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. *Nucleic Acids Res.* **18**, 2887–2890.
36. Nelson,D.L., Ledbetter,S.A., Corbo,L., Victoria,M.F., Ramirez-Solis,R., Webster,T.D., Ledbetter,D.H. and Caskey,C.T. (1989) *Alu* polymerase chain reaction: a method for rapid isolation of human-specific sequences from complex DNA sources. *Proc. Natl Acad. Sci. USA* **86**, 6686–6690.
37. Lichter, P., Ledbetter, S.A., Ledbetter, D.H. and Ward, D.C. (1980) Fluorescence *in situ* hybridization with *Alu* and L1 polymerase chain reaction probes for rapid characterization of human chromosomes in hybrid cell lines. *Proc. Natl Acad. Sci. USA* **87**, 6634–6638.
38. Emanuel, B.S., Driscoll, D., Goldmuntz, E., Baldwin, S., Biegel, J., Zackai, E.H., McDonald-McGinn, D., Sellinger, B., Gorman, N., Williams, S., and Budarf, M.L. (1993) Molecular and phenotypic analysis of the chromosome 22 microdeletion syndromes. In Epstein, C.J. (ed.), *Phenotypic Mapping of Down Syndrome and Other Aneuploid Conditions*. Wiley Liss, New York, NY, 207–224.
39. Frazer, K.A., Boehnke, M., Budarf, M.L., Wolff, R.K., Emanuel, B.S., Myers, R.M. and Cox, D.R. (1992) A radiation hybrid map of the region on human chromosome 22 containing the neurofibromatosis type 2 locus. *Genomics* **14**, 574–584.
40. Geurts van Kessel, A.H.M., Westerveld, A., de Groot, P.G., Meera Khan, P. and Hagemeijer, A. (1980) Regional localization of the genes coding for human ACO2, ARSA, and NAGA on chromosome 22. *Cytogenet. Cell. Genet.* **28**, 169–172.
41. Ledbetter, D.H., Rich, D.C., O'Connell, P., Leppert, M. and Carey, J.C. (1989) Precise localization of NF1 to 17q11.2 by balanced translocation. *Am. J. Hum. Genet.* **44**, 20–24.
42. Budarf, M.L., Sellinger, B., Griffin, C., Emanuel, B.S. (1989) Comparative mapping of the constitutional and tumor associated 11;22 translocations. *Am. J. Hum. Genet.* **45**, 128–139.
43. Delattre, O., Azambuja, C.J., Aurias, A., Zucman, J., Peter, M., Zhang, F., Hors-Cayla, M.C., Rouleau, G., and Thomas, G. (1991) Mapping of human chromosome 22 with a panel of somatic cell hybrids. *Genomics* **9**, 721–727.
44. Erikson, J., Griffin, C., ar-Rushdi, A., Valtieri, M., Hoxie, J., Finan, J., Emanuel, B.S., Rovera, G., Nowell, P.C., Croce, C.M. (1986) Heterogeneity of chromosome 22 breakpoint in Ph-positive acute lymphocytic leukemia. *Proc. Natl Acad. Sci. USA* **83**, 1807–1811.
45. Geurts van Kessel, A.H.M., Tetteroo, P.A.T., von dem Borne, A.E.G.Kr., Hagemeijer, A. and Bootsma, D. (1983) Expression of human myeloid-associated surface antigens in human-mouse myeloid cell hybriuds. *Proc. Natl Acad. Sci. USA* **80**, 3748–3752.
46. Bauer,T.R., McDermid,H.E., Budarf, M.L., Van Keuren, M.L. and Bloomberg, B.B. (1993) Physical location of the human immunoglobulin lambda-like genes 14.1, 16.1 and 16.2. *Immunogenetics* **38**, 387–399.
47. Croce, C.M., Huebner, K., Isobe, M., Fainstein, E., Lifshitz, B., Shtivelman, E., Canaani, E. (1987) Mapping of the four distinct BCR-related loci to chromosome region 22q11: order of BCR loci relative to chronic myelogeneous leukemia and acute lymphoblastic leukemia breakpoints. *Proc. Natl Acad. Sci USA* **84**, 7174–7178.
48. Scambler, P.J. (1994) Report of the Fourth International Workshop on Human Chromosome 22 Mapping. *Cytogenet. Cell Genet.* **67**, 277–319.
49. Figlewicz, D.A., Delattre, O., Guellaen, G., Krizus, A., Thomas, G., Zucman,J., and Rouleau, G.A. (1993) Mapping of human γ-glutamyl transpeptidase genes on chromosome 22 and other autosomes. *Genomics* **17**, 299–305.
50. Heisterkamp,N. and Groffen,J. (1988) Duplication of the bcr and gamma-glutamyl transpeptidase genes. *Nucleic Acids Res.* **16**, 8045–8056.
51. Budarf, M.L., Canaani, E. and Emanuel, B.S. (1988) Linear order of the four BCR-related loci in 22q11. *Genomics* **3**, 168–172.
52. Cuticchia,A.J., Arnold,J. and Timberlake,W.E. (1992) The use of simulated annealing in chromosome reconstruction experiments based on binary scoring. *Genetics* **132**, 591–601.
53. Rigault, P. (1993) In Lim,H.A., Fickett,J., Cantor,C.R. and Robbins,R.J. (eds) *Clone Ordering by Simulated Annealing: Application to the STS-Content Map of Chromosome 21*. Proceedings of the Second International Conference on Bioinformatics, Supercomputing, and Complex Genome Analysis. World Scientific Publishing: 169–183.
54. McDermid,H.E., Budarf,M.L. and Emanuel,B.S. (1993) Long-range restriction map of human chromosome 22q11–22q12 between the lambda immunoglobulin locus and the Ewing sarcoma breakpoint. *Genomics* **18**, 308–318.
55. Buetow,K.H., Duggan,D., Yang,B., Ludwigsen,S., Puck,J., Porter,J., Budarf,M., Spielman,R. and Emanuel,B.S. (1993) A microsatellite-based multipoint index map of human chromosome 22. *Genomics* **18**, 329–339.
56. Shizuya,H., Birren,B., Kim,U.J., Mancino,V., Slepak,T., Tachiiri,Y. and Simon,M. (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. *Proc. Natl Acad. Sci. USA* **89**, 8794–8797.
57. Sternberg,N. (1990) Bacteriophage P1 cloning system for the isolation, amplification, and recovery of DNA fragments as large as 100 kilobase pairs. *Proc. Natl Acad. Sci. USA* **87**, 103–107.
58. Ioannou,P.A., Amemiya,C.T., Garnes,J., Kroisel,P.M., Shizuya,H., Chen,C., Batzer,M.A. and de Jong,P.J. (1994) A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nature Genet.* **6**, 84–89.
59. Arratia,R., Lander,E., Tavare,S. and Waterman,M. (1992) Genomic mapping by anchoring random clones: a mathematical analysis. *Genomics* **11**, 806–827.
60. Cohen,D., Chumakov,I. and Weissenbach,J. (1993) A first-generation physical map of the human genome. *Nature* **366**, 698–701.
61. Weissenbach, J., Gyapay, G., Dib, C., Vignal, A., Morissette, J., Millasseau, P., Vaysseix, G. and Lathrop M. (1992) A second-generation linkage map of the human genome. *Nature* **359**, 777–778.
62. Gyapay, G., Morrisette, J., Vignal, A., Dib, C., Fizames, C., Millaseau, P., Marc, S., Bernardi, G., Lathrop, M. and Weissenbach, J. (1994) The 1993–94 genethon human genetic linkage map. *Nature Genet.* **7**, 246–339.
63. Barillot,E., Lacroix,B. and Cohen,D. (1991) Theoretical analysis of library screening using a N-dimensional pooling strategy. *Nucleic Acids Res.* **19**, 6241–6247.
64. Riazi, M.A., Mears, A.J., Bell, C.J., Budarf, M.L., Emanuel, B.S., Murray, J.C., Patil, S.R., and McDermid, H.E.(1994) Long range mapping and construction of a YAC contig within the cat eye syndrome critical region. *Am. J. Hum. Genet.* **55**, A268.
65. Driscoll, D.A., Salvin, J., Sellinger, B., McGinn-McDonald, D., Zackai, E.H., Emanuel, B.S. (1993) Prevalence of 22q11 microdeletions in DGS and VCFS: implications for genetic counseling and prenatal diagnosis. *J. Med. Genet.* **30**, 813–817.
66. Lindblom, A., Sandelin,K., Iselius,L., Dumanski,J., White,I., Nordenskjold,M. and Larsson, C. (1994) Predisposition for breast cancer in carriers of constitutional translocation 11q;22q. *Am.J. Hum. Genet.* **54**, 871–876.
67. Peyard,M., Fransson,I., Xie, Y.-G., Han, F.-Y., Ruttledge,M.H., Swahn,S., Collins,J.E., Dunham,I., Collins,V.P., and Dumanski,J.P. (1994) Characterization of a new member of the human β-adaptin gene family from chromosome 22q12, a candidate meningioma gene. *Hum. Mol. Genet.***3**, 1393–1399.
68. Amemiya,C.T., Alegria-Hartman,M.J., Aslanidis,C., Chen,C., Nikolic,J ., Gingrich,J.C. and de Jong,P.J. (1992) A two-dimensional YAC pooling strategy for library screening via STS and *Alu*-PCR methods. *Nucleic Acids Res.* **20**, 2559–2563.
69. Green,E.D and Olson,M.V. (1990) Systematic screening of yeast artificial-chromosome libraries by use of the polymerase chain reaction. *Proc. Natl Acad. Sci. USA* **87**, 1213–1217.
70. Lee,J.T., Murgia,A., Sosnoski,D.M., Olivos,I.M. and Nussbaum,R.L. (1992) Construction and characterization of a yeast artificial chromosome library for Xpter–Xq27.3: a systematic determination of cocloning rate and X-chromosome representation. *Genomics* **12**, 526–533.
71. Ousterhout,J.K. (1994) *Tcl and the Tk Toolkit*. Addison-Wesley, Reading, MA.
72. Mott,R., Grigoriev,A., Maier,E., Hoheisel,J. and Lehrach,H. (1993) Algorithms and software tools for ordering clone libraries: application to the mapping of the genome of Schizosaccharomyces pombe. *Nucleic Acids Res.* **21**, 1965–1974.

A YAC contig map of the human genome

Ilya M. Chumakov*, Philippe Rigault*, Isabelle Le Gall*, Christine Bellanné-Chantelot*, Alain Billault*, Sophie Guillou*, Pascal Soularue*, Ghislaine Guasconi*, Eric Poullier*, Isabelle Gros*, Maria Belova*, Jean-Luc Sambucy*, Laurent Susini*, Patricia Gervy*, Fabrice Gilbert*, Sandrine Beauflis*, Hung Bui*, Catherine Massart*, Marie-France De Tand*, Frédérique Dukasz*, Sandrine Lecouiant*, Pierre Ougen*, Virginie Perrot*, Martial Saumier*, Catherine Soravito*, Rita Bahouayila*, Annick Cohen-Akenine*, Emmanuel Barillot†, Stéphane Bertrand†, Jean-Jacques Codani†, Dominique Caterina*, Isabelle Georges†, Bruno Lacroix*, Georges Lucotte*, Mourad Sahbatou*, Christian Schmit*, Muriel Sangouard*, Emmanuel Tubacher*, Colette Dib†, Sabine Fauré†, Cécile Fizames†, Gabor Gyapay†, Philippe Millasseau†, Simon NGuyen†, Delphine Muselet†, Alain Vignal†, Jean Morissette†¶, Joan Menninger‡, Jonathan Lieman‡, Trushna Desai‡, Amy Banks‡, Patricia Bray-Ward‡, David Ward‡, Thomas Hudson§, Sebastian Gerety§, Simon Foote§, Lincoln Stein§, David C. Page§||, Eric S. Lander§||, Jean Weissenbach†, Denis Le Paslier* & Daniel Cohen*

* Fondation Jean Dausset Centre d'Etude du Polymorphisme Humain, 27 rue Juliette Dodu, 75010 Paris, France
† Généthon, 1 rue de l'Internationale, 91000 Evry, France
‡ Department of Genetics, Yale University School of Medicine, New Haven, Connecticut CT06510, USA
§ Whitehead Institute/MIT Center for Genome Research, Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, Massachusetts 02142, USA
|| Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA
¶ Centre de Recherche du Centre Hospitalier, Université Laval, Quebec, G1V 4G2, Canada

A yeast artificial chromosome library containing 33,000 clones with an average insert size of one megabase of human genomic DNA was extensively analysed by several different procedures for detecting overlaps and positional information. We developed an analysis strategy that resulted, after confirmatory tests, in a YAC contig map reliably covering about 75% of the human genome in 225 contigs having an average size of about ten megabases.

PHYSICAL maps of the human genome are essential tools for unravelling the genetic basis of disease[1], localizing the complete inventory of human genes, understanding the principles of genome organization and achieving other objectives of the Human Genome Project. Physical maps consist of ordered, overlapping cloned fragments of genomic DNA covering each chromosome.

Given the large size of the mammalian genomes, physical mapping of the entire human genome requires using clones with extremely large inserts, of the order of 1 megabase (Mb). Yeast artificial chromosomes (YACs)[2] are currently the only cloning system capable of propagating such large DNA fragments. Indeed, YACs have provided the basis for the first two physical maps of entire human chromosomes: 21q (ref. 3) and Y (ref. 4). More generally, YACs have been crucial tools in cloning disease genes based on their chromosomal location[5,6]. Such positional cloning[7] projects begin by genetically mapping a disease gene to a region of a few centiMorgans by tracing its inheritance relative to polymorphic DNA markers, a task made feasible by the recent availability of a complete genetic map[8] containing thousands of highly polymorphic, polymerase chain reaction (PCR)-typeable markers[9,10] known as microsatellites, simple tandem repeat polymorphisms or simple sequence length polymorphisms. One must then analyse the entire chromosomal region between the closest flanking genetic markers to identify the disease gene. YACs are invaluable for the purpose of covering such large regions, although their utility for detailed genomic analysis is somewhat limited by problems of infidelity–notably, a high frequency of chimaeric clones[11] (containing fragments from more than one genomic region)—and instability of some regions. In addition, YAC-based physical maps are important intermediates in producing a 'sequence-ready' physical map consisting of smaller and more stable clones.

Here we report our progress towards making a physical map of the human genome consisting of overlapping YACs anchored to a comprehensive set of genetic markers.

General strategy

To construct a physical map, we analysed a large-insert YAC library providing tenfold coverage of the human genome by three different experimental procedures: (1) sequence-tagged sites (STSs)[12] content mapping, involving PCR-based screening with genetically-mapped microsatellite markers: YACs identified as containing such markers were referred to as 'genetically-anchored YACs'; (2) cross-hybridization, involving hybridizing the library with probes derived from individual YACs; and (3) fingerprinting, involving characterizing each YAC in terms of the pattern of restriction fragments detected by two human repetitive sequence probes.

These three procedures provide different ways of establishing 'links', representing potential overlaps between clones. In the case of STS content mapping and cross-hybridization, the experiment yields a binary result from which links can be immediately deduced. In the case of fingerprinting, links between YACs are inferred statistically[13] when the fingerprint patterns are sufficiently similar, as described below.

It is not possible to construct a physical map based solely on the complete collection of links: most YACs aggregate into a few huge, branched, artefactual contigs. This can be expected because of the high rate of YAC chimaerism (40–50%), intra- or interchromosomal sequence similarities in the human genome, and the possibility of laboratory errors.

To circumvent this problem, we sought to build only short 'paths' between genetically anchored YACs. Paths connecting nearby points are less likely to be affected by false connections within or between the intervening YACs (such a false connection would require two chimaeric clones: one leaving away from the region and another returning to it). We also obtained partial information about the chromosomal origin(s) of many YACs through our cross-hybridization procedure and used this information to choose between paths.

We have previously given a brief description of this general strategy[14] and reported that an automatic computer implementa-

tion appeared to cover most of the human genome, but we did not provide a detailed map. We have since inspected each interval and performed confirmatory tests when necessary and more YAC links have been established. Here we describe the specific methodology of the map construction and discuss the reliability of the procedures. We also present the improved map and evaluate its coverage of the genome.

The CEPH YAC library

The entire CEPH YAC library comprises 98,208 clones representing about 17 genome equivalents. It was derived from a human male lymphoblastoid cell line, Boleth[15], and is arrayed in 1,023 96-well microtitre plates. Inserts consist of *Eco*RI partial digested human genome fragments cloned into the pYAC4 vector[2] and transfected into the host strain AB1380, as previously described[16]. (The sole exception is a set of 237 clones, in plates 2001–3, for which a recombination deficient host Rad52-3a was used[17].)

The first portion of the library, termed Mark I (containing 52,992 clones in plates 1–551) has an average insert size of 431 kilobases (kb). By using different size fractionation conditions, a Mark II library (containing 17,760 clones in plates 552–736

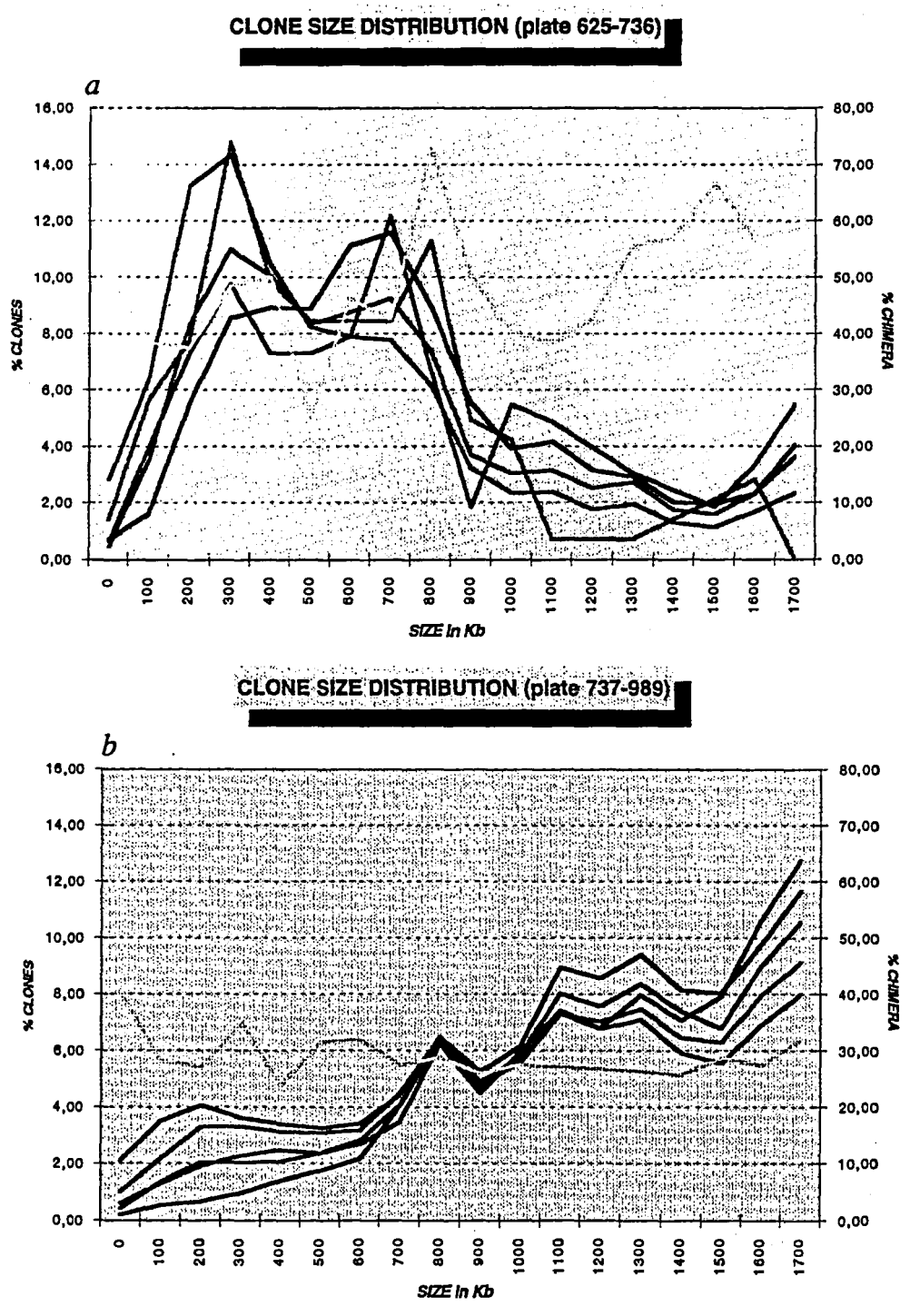


FIG. 1 Clone size distribution. The distribution of different categories of YAC are shown as follows all clones (blue), STS positive YACs (green), Alu-PCR target YACs (grey), Alu-PCR probe YACs (pink), YACs with informative fingerprint (purple). The distribution of chimaeras among A PCR probe YACs is shown in yellow.

was produced with an average insert size of 600 kb. A still larger Mega-YAC library (containing 24,288 clones in plates 737–989) was produced with an average insert size of 1,054 kb.

The YACs used in this project consisted of 10,752 clones from the Mark II library (plates 625–736) and all of the Mega-YAC library, for a total of 34,560 YAC clones providing tenfold coverage of the genome. The size of each clone was determined by field inverted gel electrophoresis (FIGE)[18] followed by Southern blotting and hybridization with a labelled probe containing pBR322 and total human DNA. Under the conditions used, sizes above 1,700 kb could not be accurately resolved. We found that 6% of the clones failed to give a hybridization signal. The size distribution is shown in Fig. 1. Multiple bands were detected in a certain proportion of the YACs (12% from Mark II and 6.8% from the Mega-YACs), which may result from clone rearrangements. In addition to these 34,560 clones used to construct the map, some YACs from Mark I and the first part of Mark II were also used. Specifically, some YACs that had previously been anchored by STS were used as hybridization probes.

STS screening

Methodology. The YAC library was screened with a large collection of PCR-typeable genetic markers, to identify clones containing each locus. To facilitate PCR-based screening of 33,024 clones (plates 625–968), we prepared pools of clones in such a manner as to reduce the number of reactions required by 100-fold, as compared to screening each clone individually[11].

The library was divided into 43 'blocks', each corresponding to eight microtitre plates (containing $8 \times 96 = 768$ clones). For each block we prepared one 'superpool' containing DNA from all the clones and 28 'subpools' prepared by using a three-dimensional pooling system based on the plate, row and column address of each clone (specifically, 8 subpools consisted of all clones residing in a given microtitre plate; 8 subpools consisted of all clones in a given row; and 12 subpools consisted of all clones in a given column). The PCR screening for each STS involved three steps: (1) identifying the positive superpools (43 reactions); (2) for each positive superpool, identifying the positive plate, row and column subpools to obtain the address of the positive clone (28 reactions); and (3) directly confirming the PCR assay on the identified clone (1 reaction). Unique addresses were not obtained when a superpool contained more than one positive clone or when one of the three dimensions failed to amplify; such cases were resolved by testing the candidate addresses consistent with the partial data when less than 16 reactions were required.

The 'complete screening' scheme described above was used in the first part of the project. After this stage we switched to a 'directed screening' strategy, using the links between YACs to further reduce the number of reactions by twofold. The strategy was first to identify positive superpools for a given STS, and screen some subpools until two YACs were identified; three positive superpools were usually necessary for this. Then we used two directed screening methods based on our database of results. The first method involved using the 'LOCUS' function, developed as part of the QUICKMAP software, to display the local contig attached to the STS and the YACs linked to it to identify other clones likely to contain the STS; such YACs were directly tested for the STS. The second method was used for confirmation of the paths. It used the 'CLONESPATH' function of QUICKMAP to construct and display potential paths through adjacent STSs (see sections on construction of the map and representation of the map below). We then tested some clones of the path against both STSs. These directed strategies were very efficient in terms of screening, although did not provide two independent tests for each clone, as in the first strategy. As false positives were highly detrimental to our mapping strategy, we distinguished between: (1) the YACs that were identified by subpool screening and individually confirmed; (2) the YACs that were identified by subpool screening but proved to be negative upon checking; and (3) the YACs that were identified by direct testing. The second case, representing about 3% of the addresses, may correspond to clones which might be genuine positive clones that we failed to detect for technical reasons. During map construction, we used the last two cases more cautiously, checking (when possible) fingerprint or hybridization information before including such YACs in the map. The PCR products were detected by agarose gel electrophoresis, ethidium bromide staining, and ultraviolet illumination. Images were captured by a CCD camera and analysed with semiautomatic software interfaced to a laboratory notebook (using Sybase).

Results. At Centre d'Etude du Polymorphisme Humain (CEPH)/Généthon, we examined a total of 2,890 polymorphic

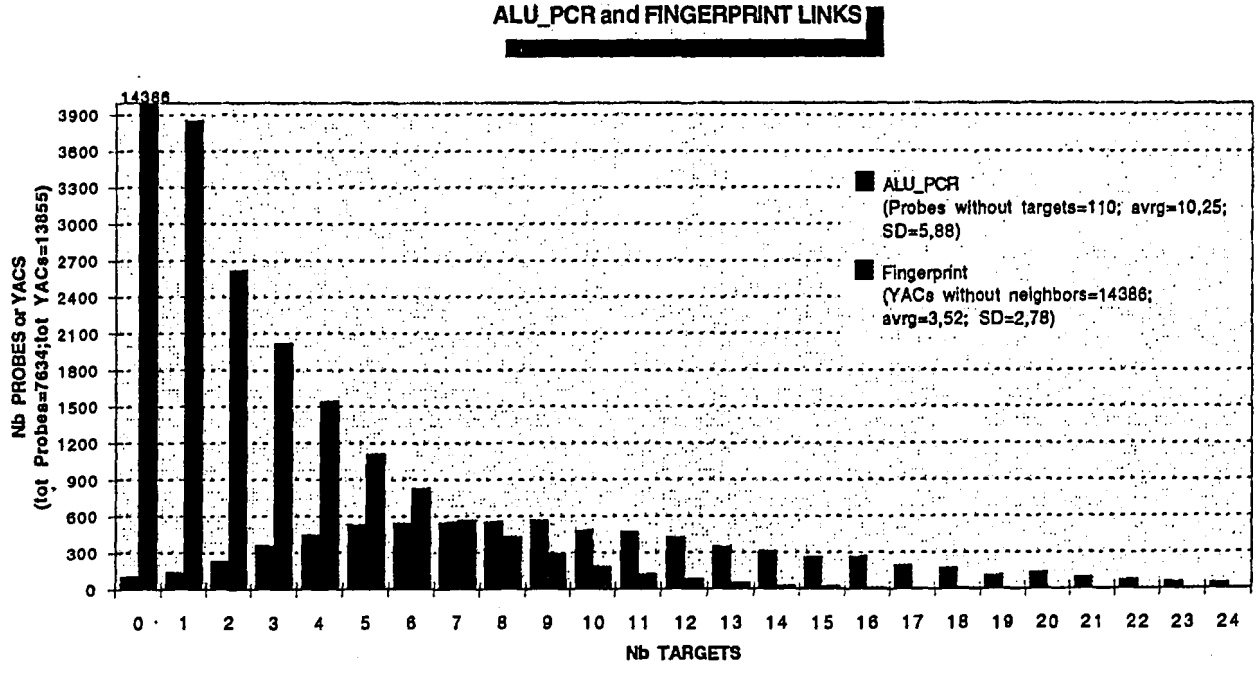


FIG. 2 Alu-PCR and fingerprint links. Blue bars represent Alu-PCR links, red bars represent fingerprint links.

markers, all generated and mapped by genetic linkage analysis as part of the Généthon genetic mapping program[8]. All markers were screened on the 43 superpools to identify the positive blocks. About 5% of the markers failed to work because of poor amplification or high background in the YAC pools. Another 5% gave no signal in the superpools, despite yielding the expected PCR product in a human genomic DNA control. In about 60% of these cases, we were able to detect and confirm a positive signal when the PCR products were electrophoresed, blotted and probed with a $(CA)_{15}$ oligonucleotide (which hybridizes to the CA repeat contained within the polymorphic locus).

Complete screening was performed for the first 814 markers, those of the first Généthon linkage map (1992). Of these, 28 failed to detect any YAC, and 786 identified 5.6 YACs on average. The 2,076 remaining genetic markers were subjected to directed screening. Of these, 261 failed to detect any YAC, and the remaining 1,815 identified an average of 4.9 YACs.

In total, 2,601 genetic markers identified at least one YAC. A total of 289 STSs have no anchored YACs, whether because of PCR-related problems or library-related problems. PCR-related problems are mainly due to sequence-dependent heavy background noise or poor amplification. In some of these cases, the design of another pair of primers from the original sequence data allowed us to obtain positive clones. YAC library-related problems can be due to the absence of clones in certain regions of the genome, either for statistical reasons or for non-clonability of certain human DNA sequences in yeast. The inability to find anchored YACs was more frequent for STSs located in certain regions of the genome, such as 1p, 19, the distal part of 17q, and most of the telomeric regions.

In addition to data generated at CEPH/Généthon, we also used results for 1,500 STSs screened elsewhere. The largest data set came from the Whitehead Institute/MIT Center for Genome Research (WI/MIT). We used the July 1994 release of this data, which contained 3,419 STSs screened with a different technology[19], using the 25,344 clones in plates 709–977 (the current publicly available release contains over 10,000 STSs and can be accessed via the World Wide Web, address 'www.genome.mit.edu'). Among these STSs were 1,128 AFM markers also screened at CEPH/Généthon. Each group found an average of 1 definite YAC address per 2 genome equivalents screened: 5.1 YACs in 10 genome equivalents screened at CEPH/Généthon, and 4.1 YACs in 8 genome equivalents screened at WI/MIT. (Additional incomplete YAC addresses were also obtained, for example, about 1.5 at WI/MIT. These were still being resolved and are not used here.) The combined data provided more complete coverage than either group alone, as roughly two YACs were found in common, three only by CEPH/Généthon, and two only by WI/MIT.

Finally, we also incorporated results from about 370 STSs screened elsewhere and deposited in public databases.

Screening by hybridization

Methodology. We screened the YAC library by hybridization, using individual probes derived from individual YACs to screen the entire Mega-YAC library. To circumvent the tedious process of purifying YAC DNA from the total yeast genomic DNA, and also to increase efficiency of the hybridization, we derived from each YAC a representative set of human-specific DNA fragments by means of inter-Alu PCR, between the ubiquitous Alu repeats spread along the human genome[20,21]. This was achieved by PCR amplification from total yeast clone DNA with a single primer[22] specific for the 3′ part of the Alu repeat sequence. Under our conditions there was no amplification from yeast genomic DNA with this primer; on average 10 different fragments of 300 base pairs (bp) average size were produced from random Mega-YACs.

Alu-PCR products were prepared individually from each YAC to be used as probe or target. To simplify the screening procedure, we used a pooling scheme for the target Alu-PCR products. The pooling procedure was similar to the scheme used for STS screening, but in this case all subpools were simultaneously screened by hybridization. In the pooling scheme, 'blocks' consisted of 4 microtitre plates which were conceptually divided into 8 half-plates. From these 8 half-plates, a total of 22 subpools were prepared, consisting of 8 subpools containing clones in the same half-plate, 8 subpools containing clones in the same row of the half-plate, and 6 subpools containing clones in the same column of the half-plate. As this part of the library (plates 734–989) represents 64 blocks, the total number of subpools to screen is $64 \times 22 = 1,408$.

The pools were spotted at high density onto nylon membranes before hybridization. The addresses of positive candidates were deduced according to which half-plate, row and column pools were found positive for each block. The YACs identified by a single signal in each dimension were called 'unique positives'. If two candidate clones are present in a block, more than three signals will be observed. In general, the addresses of the positive clones cannot be deduced unambiguously under these conditions ('undetermined positives'). However, when such candidates are located on the same row or the same column of a single half-plate, it is possible to determine these positive clones ('determined positives'). Our experience indicates, that these determined positives can be used for the map construction, but rather cautiously, as some (or many) of them are false positives. One possible explanation of this phenomenon is that some of these 'determined' positives appear to be linked to artefactual spots due to hybridization background. Moreover, in some cases some of the three-dimensional signals could not be detected for technical reasons. This could interfere with the deduction of YAC addresses when using undetermined positives.

In addition to the Alu-PCR products from the YAC clones, we also spotted in duplicate Alu-PCR products from a somatic cell hybrid panel consisting of cell lines, each containing only one or two human chromosomes. These hybridization targets provided information about the likely chromosomal localization of the YAC probes. Most of the cell hybrids were obtained from the NIGMS (Coriell Institute of Medical Research, Camden, New Jersey) mapping panel 2 (ref. 23). A chromosome 20-only G418-resistant monosomic cell hybrid DNA was provided by C. Smith. GM10791, a chromosome 7-only somatic cell hybrid DNA was provided by E. Green; and GM06318B, a chromosome X-only somatic cell hybrid DNA was provided by D. Schlessinger. In the second set of membranes used for this project, we also included somatic cell hybrids for chromosomes 1+X, 5, 6, 12 and 19, provided by D. Patterson.

The Alu-PCR products of subpools and somatic cell hybrids were spotted onto membranes together with ϕX DNA for automatic filter identification. This spotting was performed by an automatic replicating device. The membranes were hybridized in the presence of human DNA competitor with ^{33}P-labelled mixture of phage ϕX DNA and Alu PCR products of individual YACs. A high-throughput protocol that included labelling in microplates and washing membranes in batches allowed a team of two people to hybridize 200 YAC probes per day. After washing and exposure, the films were scanned and images were stored on a workstation. After automatic treatment, all images were manually inspected so artefacts could be removed from analysis and the interpreted results checked (positive YACs deduced from the subgroups and chromosomal assignment) during the analysis. The software for this semiautomatic procedure was developed in collaboration with Cose (Paris).

The pilot hybridizations with freshly made membranes indicated that 80% of random YAC probes produced an effective hybridization result. The remaining 20% gave either no signal (4% of the cases) or high background noise. This latter phenomenon is probably associated with middle-frequency repeat sequences included occasionally in inter-Alu PCR amplification products. In most of these cases, we were also unable to determine the chromosomal origin of the probe.

Generally, hybridization to somatic cell hybrid inter-Alu PCR products was less effective than to YAC targets. In pilot experiments, only 80% of successful probes gave a signal to at least one duplicate of the chromosomal inter-Alu PCR products spots. In general, we observed a very good result reproducibility when the same YAC probe was used on different batches of membranes.

Results. We derived inter-Alu PCR products for each of the 24,576 YACs of the Mega-YAC library (about eight human genome equivalents) to be used as targets for hybridization. Probes were selected by various criteria.

The first 2,000 probes were YACs belonging to chromosome-specific sublibraries generated according the procedure described to obtain the chromosome 21-specific YAC subset[22]. Briefly, Alu-PCR products of clones from a four-genome equivalent portion of the Mega-YAC library were individually spotted on membranes and hybridized successively with chromosome-specific probes obtained with inter-Alu PCR DNA products from the panel of somatic cell hybrids.

We also used as probes 200 YACs cloned in Rad52- yeast strain[17]. According to the chromosomal assignment results from hybridization, this set appeared to contain only 8% of chimaeric YACs.

The rest of the probes were chosen using the QUICKMAP software. The first objective was to obtain for each genetic locus two YACs successfully used as probes. For this about 2,500 YACs were chosen with the 'locus' function. We also used the 2,000 largest YACs that were not genetically anchored. Finally, about 2,000 YACs were chosen with the 'CLONESPATH' function during the map confirmation.

In total, 8,785 probes gave interpretable signals in this screening procedure. As expected from the selection process, the size distribution of the probes is shifted towards larger size (Fig. 1). The distribution of the number of targets detected per probe is almost gaussian, with an average of 7 ('unique positives') (Fig. 2) or 10 (when adding 'determined positives'). This is approximately half of that expected with probes spanning 1 Mb of genome. The first reason is that we wanted to avoid false positives, so we kept only the clearest signals during the image analysis. The discrepancy can also be explained by non-random distribution of inter-Alu PCR products and unequal efficiency of their individual hybridization. The distribution of YAC target sizes is also shifted towards the larger size, probably because larger clones are likely to produce more inter-Alu PCR products and so will provide stronger signals. This may also account for the larger size of successful YAC probes. In total, 20,890 (85%) of YACs were linked by hybridization to at least one other YAC. In most cases, a given YAC is detected as a target when it is used as a probe. The signal obtained is generally very intense. However, pools containing adjacent clones in the corresponding plate often produce a signal as well, probably because of minor cross-contamination. These artefacts interfere with the evaluation of positives in the corresponding pools, so targets could appear as 'undetected' in the database.

A total of 7,209 probes were assigned to chromosomes based on hybridization. Although the chromosomal assignment by inter-Alu PCR is simple, care should be taken in interpreting the results. For example, supposedly monochromosomal hybrids often contain insertions of small chromosomal fragments and deletions of other chromosomal regions. This was experimentally confirmed for the NIGMS mapping panel II used in our work. We also found by conducting reciprocal hybridization between these somatic cell hybrids that inter-Alu PCR products from some of them cross-hybridize. The most striking overlap was detected between chromosome 5-'only' and chromosome 6-'only' hybrids, as well as between chromosome 12-'only' and chromosome 6-'only' hybrids. The same pattern of cross-hybridization was observed with YAC probes. This cross-hybridization could, in some cases, be due to repeated or duplicated genomic regions.

In addition to problems with the hybrid cell lines themselves, false chromosomal assignment could result from laboratory error or sequence similarity causing cross-hybridization. Alternatively, false negatives could be due to inefficient hybridization with inter-Alu PCR products from certain YACs, or deletion of the corresponding region in the somatic cell hybrid.

Chromosomal assignment by hybridization assists in the detection of chimaeric YACs, but will obviously miss some chimaeras, including those containing only a small portion from a different chromosome region, those containing a region that is poor in Alu repeats, and those consisting of two fragments of the same chromosome. However, some apparent chimaeras could result from sequence similarity between several chromosomes. Despite these difficulties, we have used this result to analyse the chimaerism rate according to the library origin and the size distribution of the YACs (Fig. 1). The Mark II library contains a greater proportion of chimaeric YACs than the Mega-YAC library. In the Mark II library, the very large YACs seem to be more chimaeric than the smaller ones, but this is not the case for the Mega-YAC library.

Because of these interpretation problems, we treated the chromosomal assignment data with extreme caution in the QUICKMAP software, where the criteria of assignment depended on several parameters which varied according to the genomic region.

Fingerprinting

Methodology. To detect overlaps among YACs, we performed fingerprint analysis as previously described[13]. Each YAC DNA was digested with three enzymes: *Eco*RI, *Pvu*II and PstI. after agarose gel electrophoresis, the fragments were transferred onto nylon membranes using a robot. Membranes were then hybridized successively with two probes: human repeated sequences LINE-1 (L1)[24], and THE-LTR (transposon-like human-element long terminal repeat: THE)[25]. The corresponding patterns were captured automatically after scanning each film. The size of each fragment was extrapolated from the migration length of reference markers with known sizes which were run in parallel.

The L1 and THE probes were selected as they gave 6 and 11 bands per megabase, respectively. We attempted to use other repetitive probes, such as Alu, medium reiteration frequency repeats (MER)[9] and poly(GA), but with little success. The Alu probe patterns were too complex, and the MER and poly(GA) probes gave rather poor patterns with 27.6, 12.8 and 15.6% negative clones for MER 1, MER 10 (ref. 26) and poly(GA), respectively. Promising results were obtained with two probes for two Alu subfamilies, GA.007 (ref. 27) and 5OS (ref. 28), but these were poorly reproducible.

Results. A total of 31,392 YACs were successfully fingerprinted. Of these, 12.5% gave no bands for L1, 7.3% gave no bands for THE, and 4% were negative for both. When hybridized with an Alu consensus probe, one-third of these L1/THE-negative clones gave no Alu bands. The remaining clones (L1/THE-negative clones with Alu bands) contained inserts half the size of L1/THE-positive clones.

Pairwise comparisons were performed among all the fingerprints as described previously[13], and a likelihood of overlap score (LOS) was determined for each pair of clones and for each probe. Only YAC pairs having a LOS value greater than or equal to 70 for both L1 and THE were declared linked. These threshold values were chosen according to criteria based on the analysis of YACs previously mapped on chromosome 21 for which an extensive study had been performed[3]. We considered all possible pairs of YAC probes for which a chromosomal assignment was obtained by hybridization on somatic cell hybrid DNA. In this set, 70% of YAC pairs linked by fingerprint data were assigned to the same chromosome by hybridization (concordant pairs). Similarly, 68% of YAC pairs linked by hybridization showed concordant chromosomal assignment. (Interestingly, YAC pairs with reciprocal links by hybridizaton

showed 82% concordant assignment. As a control, random YAC pairs show only 8% concordance.)

In total, 17,006 YACs with these threshold values were linked to at least one other YAC from the library. On average, each of these YACs was associated with 5.8 YACs. The size distribution of these 17,006 YACs is shifted significantly towards larger sizes (1,119 kb on average). Larger YACs containing more bands would be expected to be more informative. A comparison with STS and hybridization data enabled us to detect 22 plates giving an abnormally high number of links due to a conserved fingerprint pattern in all of them. We suspect that well-to-well contamination occurred during the fingerprint process, and we removed these 22 plates from analysis. The corresponding clones made available in 1992 are free of this contamination.

Construction of the map

The starting point of the map was the framework of STSs given by the Généthon 1993–1994 linkage map[8]. This map contains 2,066 polymorphic markers, ordered in 1,267 genetic loci, each of which corresponds to a bin of 1–7 polymorphic STSs that were not recombinationally resolved. We used the three types of links between YACs (based on STS content, fingerprint and hybridization) to assemble contigs that span the intervals between genetically adjacent STSs. During this process, we integrated new STSs to this map to refine the framework order and strengthen the contigs. The limitations and precautions taken in building consistent contigs are discussed here, as each data type has its own limitations and error rate.

First, we define a minimal path between two STSs, S_1 and S_2 as an ordered list of YACs $(Y_1, \ldots, Y_n)$ that satisfy the following conditions: (1) Y_1 and Y_n contain S_1 and S_2, respectively; (2) for each $i=1, \ldots, n-1$, the YACs Y_i and Y_{i+1} are linked by one of the three mapping methods; and (3) there is no link between YACs that are not consecutive in the list. The number n of YACs in the minimal path is called the level of the path.

For several reasons, minimal paths do not necessarily represent valid 'contigs' of sequences that actually overlap in the human genome. Most importantly, chimaeric YACs artefactually join distant segments of genomic DNA, establishing connections between pairs of distant STSs. Such YACs represent between 30 and 50% of the library, depending on the genomic regions. Similarly, false positive links between clones can also result from hybridization or fingerprinting. Such false positives may make up 5–10% of the links.

Because of these problems, the backbone information from the genetic map is crucial for building accurate contigs. First, we only look for YAC paths connecting nearby STSs. Second, we can exclude some YACs that appear to be chimaeras based on their containing STSs from distinct locations, based both on Alu-PCR and the STS data (see below for more information about elimination of false links).

Contig assembly algorithm. The algorithm for constructing paths between two nearby loci proceeds by the construction of progressively larger 'neighbourhoods' of YACs. For each locus x, the computer can construct the set $N_{x,1}$ of first-degree neighbours consisting of anchored YACs (that is, YACs containing at least one STS in the locus); the set $N_{x,2}$ of second-degree neighbours consisting of YACs linked to those in $N_{x,1}$; the set $N_{x,3}$ of third-degree neighbours consisting of YACs linked to those in $N_{x,2}$; and so on. Any overlap between the neighbourhoods of loci x and y clearly yields a path connecting them. (More precisely, a YAC present in both $N_{x,i}$ and $N_{y,j}$ yields a path of length $i+j-1$.) In practice, the computer program constructs increasing neighbourhoods around both loci, halting as soon as an overlap is found. Examples are illustrated in Figs 3–5.

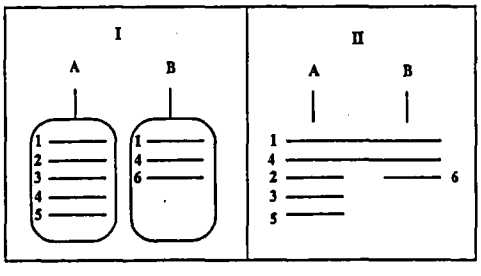


FIG. 3 Construction of a level 1 path between two loci A and B. Stage I is the construction of the first-degree neighbour set for each locus. In stage II, YACs 1 and 4 are found in common. These YACs establish the level 1 path.

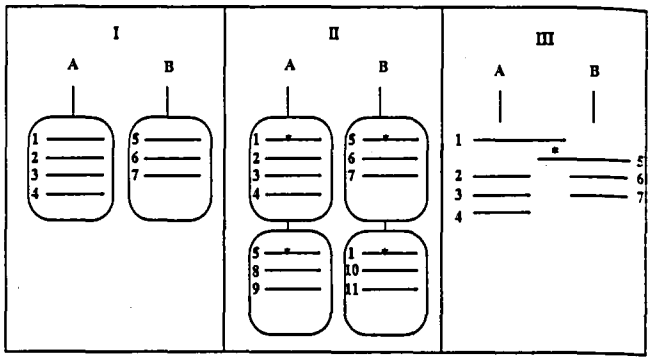


FIG. 4 Construction of a level 2 path. After stage I, no common clones are found. Stage II is the construction of the second-degree neighbour set for each locus. The asterisk documents the link between clones 1 and 5, which establish a level 2 path (stage III).

In attempting to link nearby loci on a given chromosome, we used positional information in an attempt to avoid paths that branch to distant parts of the genome. In forming second-degree and higher neighbourhoods, we excluded YACs exclusively assigned to other chromosomes by Alu-PCR hybridization, and also excluded STS-content links involving STSs from other regions.

Although the genetic linkage map represents the most likely genetic order, some local marker orders may be inverted. Accordingly, we searched not only for paths between immediately consecutive STSs (such as i and $i+1$), but also between nearby but non-consecutive STSs (such as i and $i+2$). For such non-consecutive STSs, the genetic distance and the number of intervening STSs was constrained depending on the level of the path.

Manual inspection of the paths. The map construction algorithm was applied to the whole genome. Each candidate path was then subjected to several types of checking. The first step involved graphical inspection using the 'CLONESPATH' part

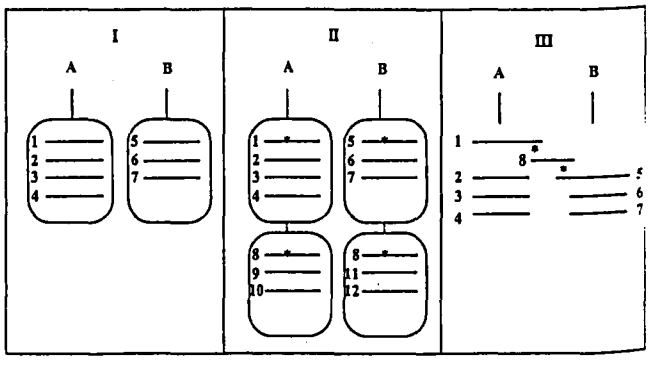


FIG. 5 Construction of a level 3 path. After stage II, clone 8 is found in common between the two second-degree neighbour sets. The level 3 path is represented at stage III.

of the program to evaluate paths based on the following criteria: (1) the number of YACs in the path; (2) The density of links between YACs; and (3) the extent to which YACs in the path were chromosomally assigned (by Alu-PCR hybridization, other STS-content information, or fluorescence *in situ* hybridization (FISH)). Graphical inspection also allowed us to detect and reject cases in which two independent paths linked the two STSs. After such visual inspection, we could reject candidate paths, try to generate a new candidate path (by trying new parameters in the algorithm or changing the order of STSs), or perform additional STS screening to test the paths further.

We tried to improve candidate paths that were judged satisfactory after graphical analysis. We derived new STSs from the ends of internal clones in a path. We also subjected the most critical clones to Alu-PCR hybridization to test their chromosomal assignment and to establish more links between clones in the path. This strategy often shortened paths by indicating overlaps that had not previously been detected because our STS screening was incomplete. To illustrate this point, Fig. 6 shows the result of an incomplete STS screening, and Fig. 7 shows the result of incomplete hybridization data. In particular, paths of level 6 or 7 in our 1993 version were converted to shorter paths. The present map contains now only paths of level 5 or less.

The bins may contain several markers that, although not recombinationally separated, span a certain distance in the physical map. To cover the physical region within the bin, we searched for paths linking different STSs within a bin. Many bins were covered by paths of level 1. For the remaining bins, we tested YACs positive for one of the STSs in a bin with the other STSs in the bin. In some cases, we used the locus program to close gaps between STSs with paths of higher levels. Fewer than 10% of the bins are not completely covered in the present map.

Because we constructed paths between genetic markers that were not necessarily adjacent in the linkage map (see above), we sometimes encountered cases in which the shortest paths connected markers i with $i+2$ and markers $i+1$ with $i+3$. This situation could arise for two reasons: we could have missed actual overlaps in the paths owing to false negative screening results (Figs 6 and 7), or the putative order on the genetic linkage map could be incorrect. To preserve the linearity of the map in these rare cases, we have either inverted the marker order or joined recombinationally separable genetic markers in the same bin. As a result, the physical order of the markers on the summary figures of the atlas (see below) does not perfectly correspond to the Généthon 1994 linkage map.

Integration of other STSs. In addition to the backbone STSs taken from the 1994 Généthon genetic linkage map, we also integrated some additional STSs that improved paths in the map. These markers came from two sources. First, WI/MIT had screened 3,419 STSs against the YACs by June 1994. From this

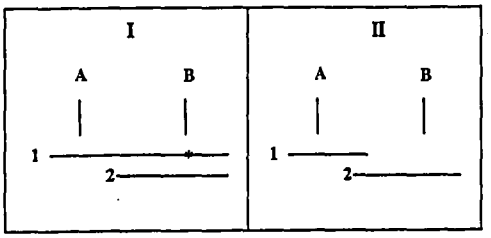


FIG. 6 Part I represents the real disposal of the YACs. If STS B is not tested against YAC 1, the path would appear at level 2 (provided the overlap between the two YACs is detected), as shown in part II. In such a case, we would have tested A against 2 and B against 1 and reduced the level to the actual value.

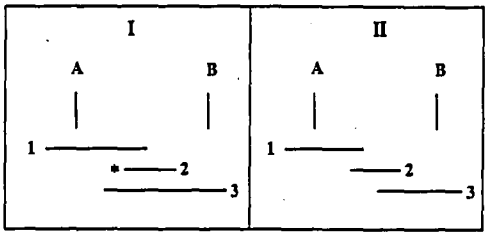


FIG. 7 Part I represents the real disposal of the YACs. If neither YAC 1 nor YAC 3 is used as Alu-PCR probe, the hybridization between them cannot be detected. In this case, YAC 2 was used as probe and detected the YACs 1 and 3. The path appears to be level 3, as shown in part II. This situation can be resolved by testing either 1 or 3 as Alu-PCR probe.

STS-content data, we selected 173 STSs (including 76 non-AFM genetic markers) that significantly improved paths. Second, the CEPH/Généthon group screened STSs from 445 unpublished genetic markers from Généthon (C. Dib *et al.*, manuscript submitted). Where known, chromosomal assignment or approximate map position was used for both sets of markers. In most cases the integration produced denser contigs and decreased the level of paths (see Fig. 6).

FISH mapping

A total of 650 genetically anchored YACs, approximately one every 5–10 cM, were selected and used as probes for fluorescent *in situ* hybridization (FISH) on metaphase chromosomes. The chimaerism rate detected by this method was 46%. Based on the comparison of cytogenetic and genetic localizations, there appear to be higher frequencies of recombination near telomeres and lower frequencies near centromeres. For example, the genetic distance between the centromeric markers D1S440 (at 163 cM) and D1S484 (at 182 cM) represents 6.5% of the genetic length of this chromosome, but 17% of the fractional cytogenetic length of the chromosome. Similarly, the interval between the centromeric markers D6S272 (at 75 cM) and D6S421 (at 86 cM) represents 5.3% of the genetic map but 17% of the fractional length of chromosome 6. In contrast, the telomeric loci D6S411 (at 173 cM) and D6S281 (at 207 cM) are separated by 16.4% of the genetic length of the chromosome, but the interval between D6S411 to the telomere is only 4% of the fractional cytogenetic distance. The FISH analysis indicates that there are no genetic markers on 13p, 14p and 15p, and that the terminal region of chromosome 20q is not contained in the genetic map. Thus there is no coverage by YAC of these regions in our physical map.

Presentation of the map

For each chromosome, the atlas following this paper shows: (1) a summary map of each chromosome, showing the cytogenetic representation, and the scales of the physical and genetic maps, together with the indication of the regions covered in contigs; and (2) a map of detailed contigs for each chromosome.

Summary figure. Each chromosome is presented at the same genetic scale as an ideogram[29] at the left side of each drawing. At the right of this ideogram is the physical map scale, showing the location of the bins. In parentheses are two numbers, separated by a semicolon: the first number is the number of STSs in the bin, the second is the number of YACs that are anchored to the bin. The links between the physical map scale and the cytogenetic scale are established through YACs that are anchored to the bins and have been used in FISH hybridization. Each of those YACs establishes a connection between the bin and an interval in the ideogram. Note that the FISH measurements have been made in terms of fractional length of the whole chromosome, and that the size of heterochromatic and centromeric regions may vary between individuals. As a consequence, a slight distortion can occur in our figures after these regions, especially for chromosomes containing entirely heterochromatic p arms

(acrocentrics). At the right of the physical map, the intervals covered in contigs are represented with coloured rectangles. The different colours represent the different levels of the paths. Finally, the correspondence between the physical map scale and the Généthon linkage map, used as a backbone for the bin locations, is shown at the right of these rectangles. The positions in the genetic map are expressed in Morgans from the most distal marker of the p arm of the chromosome.

Detailed contigs. Contigs are presented for each chromosome from pter to qter. They correspond to a succession of paths, represented by rectangles on the summary figure. Each path is a collection of clones, ordered in stacks. The number of stacks in a path corresponds to the level of the path. The graphical presentation of paths provides the following characteristics of the clones: STS-content information for the YACs; sizes of the YACs; overlap relationships between YACs based on Alu-PCR hybridization and fingerprint data; chromosomal assignment for YACs used as probes for Alu-PCR hybridization; and indication of YACs used as FISH probes.

Each locus is indicated by a white rectangle that indicates its chromosome and position. STSs located in the bin are displayed above the rectangle. These STSs are numbered within the bin and are displayed in a beige rectangle. For example, the bin located at position 1.00 on chromosome 1 contains two STSs: AFM120xd4 (D1S209), and AFM286xd9 (D1S473). This bin is thus presented as:

1: AFM120xd4 (D1S209)
2: AFM286xd9 (D1S473)
CHR 1 position 1.00

The clone stacks displayed under each bin represent the anchored YACs (that is, the YACs that contain at least one STS of the bin), and the stacks that are between two bins represent the ordered groups of clones internal to paths of level 3 and above. Within a stack, the YACs are displayed according their order in the library, from top to bottom. Each YAC is represented by a yellow box with a horizontal bar in the middle. The YAC name and its size in kilobases are represented from left to right above the bar. A '+' sign after the size means that multiple bands were detected; only the biggest size is displayed. The names of YACs used for FISH hybridization appear in a box (for example, YAC 763B12, anchored to position 1.00 on chromosome 1). Chromosomal assignment (for clones used as Alu-PCR hybridization probe) and the STS content of the YAC are represented from left to right under the bar.

Chromosomal assignment is made based on the results of hybridization with somatic cell hybrids. Because of the problems described above, chromosomal assignments were sometimes ambiguous. The assignments are represented by the following code: (1) one white dot: a probe that was not assigned; (2) two blue dots: a probe assigned only to the chromosome under consideration; (3) one blue dot, one orange dot: a probe assigned to the chromosome under consideration, as well as to one or more additional chromosomes; and (4) two orange dots: a probe that is not assigned to the chromosome under consideration, but that is assigned to one or more other chromosomes.

The display of the STS information differs between stacks composed of anchored YACs and stacks located between two bins. For anchored YACs, the stack shows the clone numbers of the STSs in the bins. For example, the bin at position 1.00 on chromosome 1 shows YACs 631C9, 732A10 and 752E3; they contain, respectively, the second, the first, and both STSs, and are given the lists '2', '1' and '1 2'. For stacks between bins, we represent the position of the STS for which the YAC is positive. For example, '1-0.87' means position 0.87 on chromosome 1, '1-?' means chromosome 1 but position unknown on this chromosome, and '?' means that no positional information is known. If the YAC is positive for several STSs, located at different places, then asterisks are displayed.

The relationship between clones in adjacent stacks is shown as follows. For paths of level 1, the path is established through the presence of one or more clone in the adjacent anchored stacks, with a thick bar (yellow and black) displayed between the two stacks. For example, paths of level 1 are established between the loci 1.00 and 1.02 on chromosome 1 by the YACs 752E3, 763B12, 830E7 and 940C1. No Alu-PCR hybridization or fingerprint linkage is involved in establishing paths of level 1.

For level 2 and higher, an array of one or more columns is displayed between the stacks, representing the fingerprint and Alu-PCR relationships that link the stacks. Each column of the array is composed of a black box and 3 subcolumns. The horizontal position of the black box relative to the column gives the orientation for reading the columns. All columns within an array have the same orientation. If the box lies on the left (respectively, right) of the column, this column refers to the clone of the left (respectively, right) stack that is vertically in the same place as the box. We call this clone the attached clone. The three subcolumns contain symbols (dots and triangles) that refer in this case to the clones of the right (respectively, left) stack that are vertically in front of them. The subcolumn that is just to the side of the black box can be either yellow or pale blue. It is yellow if the attached clone was not used as Alu-PCR target (does not belong to plates 734–989). If this clone was used as an Alu-PCR target, this subcolumn is pale blue and the triangles in it refer to the Alu-PCR probes that hit this clone by hybridization. The middle subcolumn is yellow if the attached clone was not used as Alu-PCR probe. This subcolumn is blue otherwise, and the triangles in it refer to the targets hit by this clone. The third subcolumn is yellow if the attached YAC was not fingerprinted. If this clone was fingerprinted, it is green and contains black dots that refer to overlapping YACs by fingerprint. This two-colour presentation allows the reader to distinguish for example between a clone that was not used as an Alu-PCR probe from one that was used as an Alu-PCR probe but did not hit any YAC in the adjacent stack. It also provides a very quick way of highlighting the clones with the most overlap information, which are the ones linked to the region with the highest probability.

As an example, in the path between 1.08 and 1.10 on chromosome 1, the first column refers to the YAC 912G11 and the second to the YAC 957A9, because the black box lies on the left of the column. The YAC 912G11 was used as a target for Alu-PCR hybridization, was not used as an Alu-PCR probe, and was fingerprinted. As a target for Alu-PCR hybridization, it is hit by YACs 774C4, 800E10 and 943A2 as hybridization probes. It overlaps by fingerprint data with YAC 895B12. The YAC 957A9 was used as a target for Alu-PCR hybridization, was used as an Alu-PCR probe, and was fingerprinted. As an Alu-PCR target, it is hit by the probe 927C3. As an Alu-PCR probe, it hit YAC 927C3. It also overlaps by fingerprint data with YAC 927C3.

Map reliability and coverage

All YAC paths covering genetic intervals have been inspected and checked, as described above. Contigs cover 75% of the genetic intervals, which together comprise 66% of the total genetic length of the genome (based on the sex-averaged meiotic map). The proportion of the genetic length covered in paths of level 1 is 26%, of level 2 is 17%, of level 3 is 15%, of level 4 is 5%, and of level 5 is 2%. These numbers are calculated on the basis of all chromosomes except 3, 12, 21, 22 and Y, which were either previously mapped or subjected to more intensive mapping by groups presenting their maps elsewhere in this volume. The chromosomes that are covered for more than 66% of their genetic length are: 4, 5, 7, 8, 9, 11, 14, 15, 16, 18 and 20. Low coverage of chromosome X (23% of its total genetic length) is expected, both because the YAC library underrepresents the

X chromosome (being derived from a male) and because the genetic map of X is sparse. In addition, the screening efficiency with backbone STSs from 1p, 19 and 17 was particularly low, resulting in poor coverage of these regions. However, chromosome 18 is almost entirely covered because we made a special effort to fill gaps by more intensive hybridization screening effort with YAC probes derived from path extremities. In many cases, new STSs were derived from those YAC probes creating new paths.

Inferring the actual proportion of the physical length of the genome covered is not entirely straightforward. The proportion of the genetic length covered (66%) may overestimate the actual proportion covered because it neglects the physical distances within the bins of recombinationally inseparable markers. But it may underestimate the coverage because the density of YACs appears to be sparsest in the telomeric regions, which are precisely those regions in which the ratio of genetic to physical distance appears to be greatest.

If the AFM markers were randomly distributed they would occur with a random spacing of about 1.2 Mb, just slightly larger than the average size of our YAC clones. We would thus expect to cover much of the genome in paths of level 1, 2 or 3. The observed proportion of intervals covered by such paths agree very well with expectation based on computer simulations. Moreover, mathematical analysis predicts that the YACs identified by the genetic markers would contain about half of the genome[30]. Although the AFM markers are known to be not completely randomly distributed, the overall effect of non-randomness appears not to be severe.

To evaluate the reliability of the YAC contig map, we examined 161 non-AFM genetic markers from a recently published collaborative genetic map[31] of the human genome (omitting markers for chromosomes 3, 12, 21 and 22) that were screened against the YAC library at WI/MIT. Of the STSs, the number detecting 1, 2 or at least 3 YACs was 20, 14 and 127, respectively 1, 2 and 3. In 60, 78 and 88% of the cases, respectively, the YACs containing the markers had been assigned to the expected location (based on the known genetic location of the marker). In the remaining cases, the corresponding YACs were not found on the expected chromosome but were sometimes found on other chromosomes (possibly resulting from chimaerism).

These results also indicate that the map covers most of the human genome. However, they do not provide a direct estimate of coverage because only loci that detected at least one YAC were considered, and the genetic markers tested may tend to lie in the same regions as the genetic markers used to construct the map. Mitigating against this concern is that the genetic markers used to assess coverage were predominantly tetra-nucleotide repeats, whose regional biases may differ from the CA repeats in the Généthon genetic map.

Given the results above, it seems reasonable to estimate that the physical map covers about 75% of the genome in 225 contigs having an average size of about 10 Mb.

Public availability of the map

Clone availability. Primary copies of CEPH YAC library were distributed to following centres:

Whitehead Institute/MIT Center for Genome Research, Cambridge, Massachusetts 02142, USA; E. S. Lander and T. Hudson; e-mail: lander@genome.wi.mit.edu.

The Reference Library DataBase (RLDB), MPI for Molecular Genetics, Ihnestrasse 73, 14195 Berlin-Dahlen, Germany; H. Lehrach; tel: (49) 30 8413 1627; fax: (49) 30 8413 1395.

3-1-1 Koyadai, Tsukuba, Ibaraki 305, Japan; K. Yokoyama; tel: (81) 298 36 3612; fax: (81) 298 36 9120.

Human Genome Center, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato, Tokyo 108, Japan; Y. Nakamura; tel: (81) 3 5449 5372; fax: (81) 3 5449 5433.

Shanghai Institute of Hematology, Rui-Jin Hospital, Shanghai Second Medical University, Shanghai 200025, China; Z. Chen; tel: (86) 21 3180 300; fax: (86) 21 4743 206.

GBE, CNR, via Abbiategrasso 207, 27100 Pavia, Italy; D. Toniolo; tel: (39) 382 546 340; fax: (39) 382 422 286.

YAC Screening Centro, Leiden University, Department of Human Genetics Wassenaarseweg 72, 2333 Al Leiden, The Netherlands; G. J. B. van Ommon; tel: (31) 71 276 081; fax: (31) 71 276 075.

Human Genome Mapping Project Resource Centre, (HGMP) Hinxton Hall, Hinxton, Cambridge CB10 1RQ, U.K.; K. Gibson; tel: (44) 1223 494 500; fax: (44) 1223 494 512.

Clones can be obtained also from Foundation Jean-Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France; D. Le Paslier; e-mail: denis@ceph.cephb.fr.

Data distribution

Anonymous ftp server: ftp://ceph-genethon-map.cephb.fr/pub/ceph-genethon-map.

World Wide Web server: URL address: http://www.cephb.fr/bio/ceph-genethon-map.html.

Mail server: ceph-genethon-map@cephb.fr.

How to use it: $ mail ceph-genethon-map@cephb.fr. Subject: infoclone. 755_f_4 672_a_3 D12S76.

other YAC or STS names.

QUICKMAP (developed by P. Rigault and E. Poullier at CEPH) is a mapping tool containing all the CEPH/Généthon screening data. It was designed to manage the production of STS screening and hybridization data, using the results analysed on a daily basis to suggest new tests. It was then modified to make CEPH/Généthon data accessible to the scientific community. QUICKMAP allows both navigation within CEPH/Généthon map and dynamic construction of contigs to integrate further datasets. QUICKMAP has been publicly available since February 1993 on our ftp site. □

Received 29 September 1994; revised 6 July 1995; accepted 17 July 1995.

1. McKusick, V. A. *Mendelian Inheritance In Man* 5th edn. (John Hopkins Univ. Press, Baltimore, 1978).
2. Burke, D. T., Carle, G. F. & Olson, M. V. *Science* **236,** 806–812 (1987).
3. Chumakov, I. *et al. Nature* **359,** 380–386 (1992).
4. Foote, S., Vollrath, D., Hilton, A. & Page, D. *Science* **258,** 60–66 (1992).
5. Miki, Y. *et al. Science* **266,** 66–71 (1994).
6. The Huntington Disease Collaboration Research Group *Cell* **72,** 971–983 (1993).
7. Collins, F. S. *Nature Genet.* **1,** 3–6 (1992).
8. Gyapay, G. *et al. Nature Genet.* **7,** 246–339 (1994).
9. Litt, M. & Luty, J. A. *Am. J. hum. Genet.* **44,** 397–401 (1989).
10. Weber, J. L. & May, P. E. *Am. J. hum. Genet.* **44,** 388–396 (1989).
11. Green, E. D., Riethman, H. C., Dutchik, J. E. & Olson, M. V. *Genomics* **11,** 658–659 (1991).
12. Olson, M. V., Hood, L., Cantor, C. R. & Botstein, D. *Science* **245,** 1434–1435 (1989).
13. Bellanne-Chantelot, C. *et al. Cell* **70,** 1059–1068 (1992).
14. Cohen, D., Chumakov, I. & Weissenbach, J. *Nature* **366,** 698–701 (1993).
15. Yang, S. Y. in *Immunobiology of HLA*, Vol. 1 (Springer, New York, 1989).
16. Albertsen, H. *et al. Proc. natn. Acad. Sci. U.S.A.* **87,** 4256–4260 (1990).
17. Haldi, M. *et al. Genomics* **24,** 478–484 (1994).
18. Carle, G. F., Franck, M. & Olson, M. V. *Science* **232,** 65–68 (1986).
19. Bell, C. J. *et al. Hum. molec. Genet.* **4,** 59–69 (1995).
20. Nelson, D. L. *et al. Proc. natn. Acad. Sci. U.S.A.* **86,** 6686–6690 (1989).
21. de Jong, P. J. *et al. Cytogenet. Cell Genet.* **51,** 985 (1989).
22. Chumakov, I. M. *et al. Nature Genet.* **1,** 222–225 (1992).
23. Mullivor, R. A., Greene, A. E., Drwinga, H. L., Toji, L. H. & Klm, C. *Am J. hum. Genet.* (Suppl.) **49,** 370 (1991).
24. Shafit-Zagardo, B., Maio, J. J. & Brown, F. L. *Nucleic Acids Res.* **10,** 3175–3193 (1982).
25. Fields, C. A., Grady, D. L. & Moyzis, R. K. Genomics **13,** 431–436 (1992).
26. Kaplan, D. J., Jurka, J., Solus, J. F. & Duncan, C. H. *Nucleic Acids Res.* **19,** 4731–4738 (1991).
27. Matera, G. A., Hellmann, U., Hintz, M. F. & Schmid, C. W. *Nucleic Acids Res.* **18,** 6019–6023 (1990).
28. Jurka, J. & Milosavljevic, J. *Molec. Evol.* **32,** 105–121 (1991).
29. Francke, U. *Cytogenet. Cell Genet.* **68,** 206–219 (1994).
30. Aratia, R., Lander, E. S., Tavare, S. & Waterman, M. S. *Genomics* **11,** 806–827 (1991).
31. Murray, J. C. *et al. Science* **265,** 2049–2054 (1994).

ACKNOWLEDGEMENTS. This work was supported by Association Francaise contre les Myopathies (AFM), Ministere de la Recherche et l'Enseignement Superieure, Groupement de Recherche et d'Etude des Genomes, la Ligue contre le Cancer, the European Economic Community program DGXII and the National Center for Human Genome Research of the US NIH.

A YAC contig map of the human genome

Ilya M. Chumakov*, Philippe Rigault*, Isabelle Le Gall*, Christine Bellanné-Chantelot*, Alain Billault*, Sophie Guillou*, Pascal Soularue*, Ghislaine Guasconi*, Eric Pouilier*, Isabelle Gros*, Maria Belova*, Jean-Luc Sambucy*, Laurent Susini*, Patricia Gervy*, Fabrice Gilbert*, Sandrine Beaufils*, Hung Bui*, Catherine Massart*, Marie-France De Tand*, Frédérique Dukasz*, Sandrine Lecouliant*, Pierre Ougen*, Virginie Perrot*, Martial Saumier*, Catherine Soravito*, Rita Bahouayila*, Annick Cohen-Akenine*, Emmanuel Barillot†, Stéphane Bertrand†, Jean-Jacques Codani†, Dominique Caterina*, Isabelle Georges†, Bruno Lacroix*, Georges Lucotte*, Mourad Sahbatou*, Christian Schmit*, Muriel Sangouard*, Emmanuel Tubacher*, Colette Dib†, Sabine Fauré†, Cécile Fizames†, Gabor Gyapay†, Philippe Millasseau†, Simon NGuyen†, Delphine Muselet†, Alain Vignal†, Jean Morissette†¶, Joan Menninger‡, Jonathan Lieman‡, Trushna Desai‡, Amy Banks‡, Patricia Bray-Ward‡, David Ward‡, Thomas Hudson§, Sebastian Gerety§, Simon Foote§, Lincoln Stein§, David C. Page§||, Eric S. Lander§||, Jean Weissenbach†, Denis Le Paslier* & Daniel Cohen*

* Fondation Jean Dausset Centre d'Etude du Polymorphisme Humain, 27 rue Juliette Dodu, 75010 Paris, France
† Généthon, 1 rue de l'Internationale, 91000 Evry, France
‡ Department of Genetics, Yale University School of Medicine, New Haven, Connecticut CT06510, USA
§ Whitehead Institute/MIT Center for Genome Research, Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, Massachusetts 02142, USA
|| Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA
¶ Centre de Recherche du Centre Hospitalier, Université Laval, Quebec, G1V 4G2, Canada

A yeast artificial chromosome library containing 33,000 clones with an average insert size of one megabase of human genomic DNA was extensively analysed by several different procedures for detecting overlaps and positional information. We developed an analysis strategy that resulted, after confirmatory tests, in a YAC contig map reliably covering about 75% of the human genome in 225 contigs having an average size of about ten megabases.

PHYSICAL maps of the human genome are essential tools for unravelling the genetic basis of disease[1], localizing the complete inventory of human genes, understanding the principles of genome organization and achieving other objectives of the Human Genome Project. Physical maps consist of ordered, overlapping cloned fragments of genomic DNA covering each chromosome.

Given the large size of the mammalian genomes, physical mapping of the entire human genome requires using clones with extremely large inserts, of the order of 1 megabase (Mb). Yeast artificial chromosomes (YACs)[2] are currently the only cloning system capable of propagating such large DNA fragments. Indeed, YACs have provided the basis for the first two physical maps of entire human chromosomes: 21q (ref. 3) and Y (ref. 4). More generally, YACs have been crucial tools in cloning disease genes based on their chromosomal location[5,6]. Such positional cloning[7] projects begin by genetically mapping a disease gene to a region of a few centiMorgans by tracing its inheritance relative to polymorphic DNA markers, a task made feasible by the recent availability of a complete genetic map[8] containing thousands of highly polymorphic, polymerase chain reaction (PCR)-typeable markers[9,10] known as microsatellites, simple tandem repeat polymorphisms or simple sequence length polymorphisms. One must then analyse the entire chromosomal region between the closest flanking genetic markers to identify the disease gene. YACs are invaluable for the purpose of covering such large regions, although their utility for detailed genomic analysis is somewhat limited by problems of infidelity–notably, a high frequency of chimaeric clones[11] (containing fragments from more than one genomic region)—and instability of some regions. In addition, YAC-based physical maps are important intermediates in producing a 'sequence-ready' physical map consisting of smaller and more stable clones.

Here we report our progress towards making a physical map of the human genome consisting of overlapping YACs anchored to a comprehensive set of genetic markers.

General strategy

To construct a physical map, we analysed a large-insert YAC library providing tenfold coverage of the human genome by three different experimental procedures: (1) sequence-tagged sites (STSs)[12] content mapping, involving PCR-based screening with genetically-mapped microsatellite markers: YACs identified as containing such markers were referred to as 'genetically-anchored YACs'; (2) cross-hybridization, involving hybridizing the library with probes derived from individual YACs; and (3) fingerprinting, involving characterizing each YAC in terms of the pattern of restriction fragments detected by two human repetitive sequence probes.

These three procedures provide different ways of establishing 'links', representing potential overlaps between clones. In the case of STS content mapping and cross-hybridization, the experiment yields a binary result from which links can be immediately deduced. In the case of fingerprinting, links between YACs are inferred statistically[13] when the fingerprint patterns are sufficiently similar, as described below.

It is not possible to construct a physical map based solely on the complete collection of links: most YACs aggregate into a few huge, branched, artefactual contigs. This can be expected because of the high rate of YAC chimaerism (40–50%), intra- or interchromosomal sequence similarities in the human genome, and the possibility of laboratory errors.

To circumvent this problem, we sought to build only short 'paths' between genetically anchored YACs. Paths connecting nearby points are less likely to be affected by false connections within or between the intervening YACs (such a false connection would require two chimaeric clones: one leaving away from the region and another returning to it). We also obtained partial information about the chromosomal origin(s) of many YACs through our cross-hybridization procedure and used this information to choose between paths.

We have previously given a brief description of this general strategy[14] and reported that an automatic computer implementa-

tion appeared to cover most of the human genome, but we did not provide a detailed map. We have since inspected each interval and performed confirmatory tests when necessary and more YAC links have been established. Here we describe the specific methodology of the map construction and discuss the reliability of the procedures. We also present the improved map and evaluate its coverage of the genome.

The CEPH YAC library

The entire CEPH YAC library comprises 98,208 clones representing about 17 genome equivalents. It was derived from a human male lymphoblastoid cell line, Boleth[15], and is arrayed in 1,023 96-well microtitre plates. Inserts consist of *Eco*RI partial digested human genome fragments cloned into the pYAC-vector[2] and transfected into the host strain AB1380, as previously described[16]. (The sole exception is a set of 237 clones, in plates 2001–3, for which a recombination deficient host Rad52-3a was used[17].)

The first portion of the library, termed Mark I (containing 52,992 clones in plates 1–551) has an average insert size of 43 kilobases (kb). By using different size fractionation conditions a Mark II library (containing 17,760 clones in plates 552–736

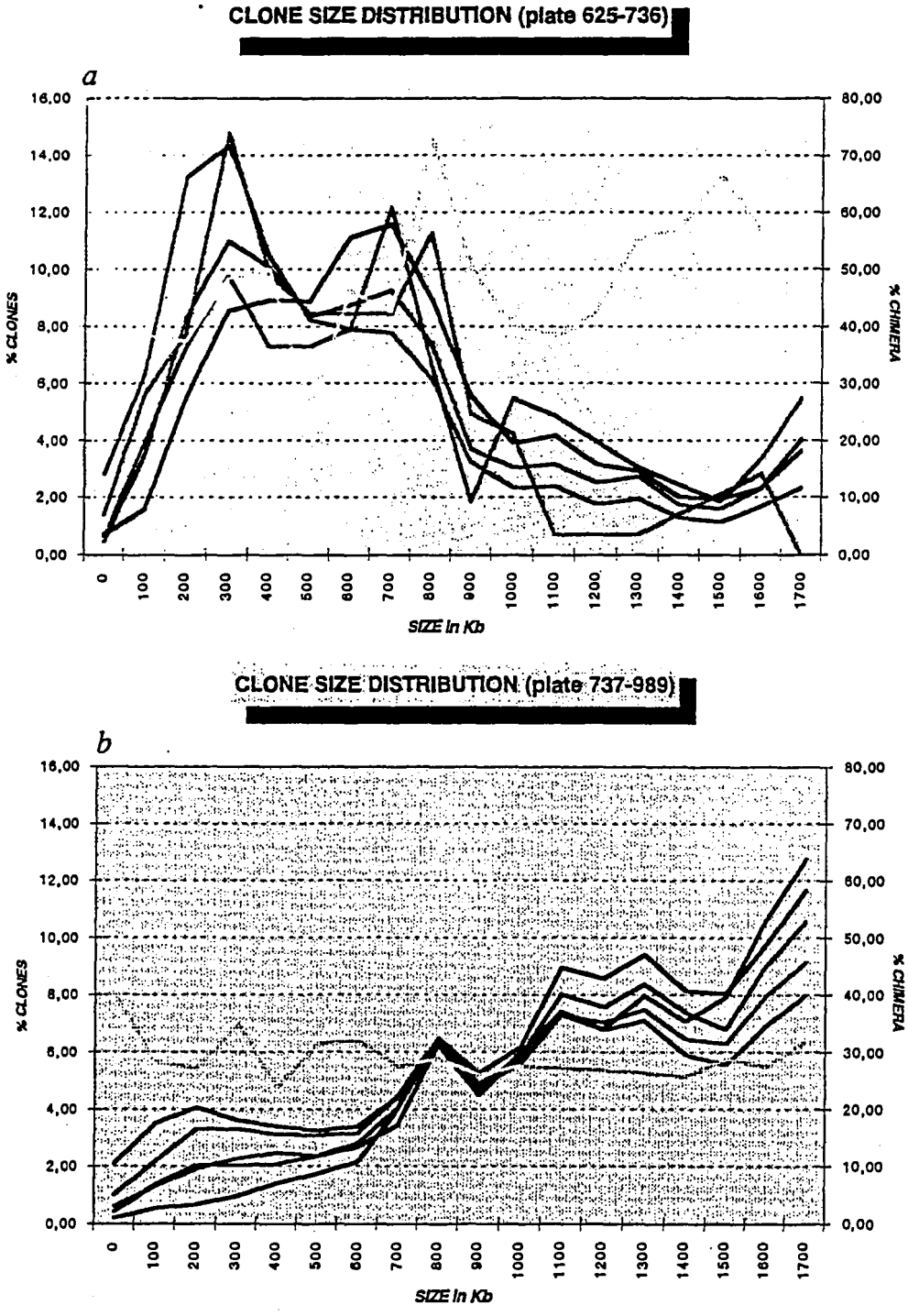


FIG. 1 Clone size distribution. The distribution of different categories of YAC are shown as follows all clones (blue), STS positive YACs (green), Alu-PCR target YACs (grey), Aln-PCR probe YACs (pink), YACs with informative fingerprint (purple). The distribution of chimaeras among PCR probe YACs is shown in yellow.

was produced with an average insert size of 600 kb. A still larger Mega-YAC library (containing 24,288 clones in plates 737–989) was produced with an average insert size of 1,054 kb.

The YACs used in this project consisted of 10,752 clones from the Mark II library (plates 625–736) and all of the Mega-YAC library, for a total of 34,560 YAC clones providing tenfold coverage of the genome. The size of each clone was determined by field inverted gel electrophoresis (FIGE)[18] followed by Southern blotting and hybridization with a labelled probe containing pBR322 and total human DNA. Under the conditions used, sizes above 1,700 kb could not be accurately resolved. We found that 6% of the clones failed to give a hybridization signal. The size distribution is shown in Fig. 1. Multiple bands were detected in a certain proportion of the YACs (12% from Mark II and 6.8% from the Mega-YACs), which may result from clone rearrangements. In addition to these 34,560 clones used to construct the map, some YACs from Mark I and the first part of Mark II were also used. Specifically, some YACs that had previously been anchored by STS were used as hybridization probes.

STS screening

Methodology. The YAC library was screened with a large collection of PCR-typeable genetic markers, to identify clones containing each locus. To facilitate PCR-based screening of 33,024 clones (plates 625–968), we prepared pools of clones in such a manner as to reduce the number of reactions required by 100-fold, as compared to screening each clone individually[11].

The library was divided into 43 'blocks', each corresponding to eight microtitre plates (containing $8 \times 96 = 768$ clones). For each block we prepared one 'superpool' containing DNA from all the clones and 28 'subpools' prepared by using a three-dimensional pooling system based on the plate, row and column address of each clone (specifically, 8 subpools consisted of all clones residing in a given microtitre plate; 8 subpools consisted of all clones in a given row; and 12 subpools consisted of all clones in a given column). The PCR screening for each STS involved three steps: (1) identifying the positive superpools (43 reactions); (2) for each positive superpool, identifying the positive plate, row and column subpools to obtain the address of the positive clone (28 reactions); and (3) directly confirming the PCR assay on the identified clone (1 reaction). Unique addresses were not obtained when a superpool contained more than one positive clone or when one of the three dimensions failed to amplify; such cases were resolved by testing the candidate addresses consistent with the partial data when less than 16 reactions were required.

The 'complete screening' scheme described above was used in the first part of the project. After this stage we switched to a 'directed screening' strategy, using the links between YACs to further reduce the number of reactions by twofold. The strategy was first to identify positive superpools for a given STS, and screen some subpools until two YACs were identified; three positive superpools were usually necessary for this. Then we used two directed screening methods based on our database of results. The first method involved using the 'LOCUS' function, developed as part of the QUICKMAP software, to display the local contig attached to the STS and the YACs linked to it to identify other clones likely to contain the STS; such YACs were directly tested for the STS. The second method was used for confirmation of the paths. It used the 'CLONESPATH' function of QUICKMAP to construct and display potential paths through adjacent STSs (see sections on construction of the map and representation of the map below). We then tested some clones of the path against both STSs. These directed strategies were very efficient in terms of screening, although did not provide two independent tests for each clone, as in the first strategy. As false positives were highly detrimental to our mapping strategy, we distinguished between: (1) the YACs that were identified by subpool screening and individually confirmed; (2) the YACs that were identified by subpool screening but proved to be negative upon checking; and (3) the YACs that were identified by direct testing. The second case, representing about 3% of the addresses, may correspond to clones which might be genuine positive clones that we failed to detect for technical reasons. During map construction, we used the last two cases more cautiously, checking (when possible) fingerprint or hybridization information before including such YACs in the map. The PCR products were detected by agarose gel electrophoresis, ethidium bromide staining, and ultraviolet illumination. Images were captured by a CCD camera and analysed with semiautomatic software interfaced to a laboratory notebook (using Sybase).

Results. At Centre d'Etude du Polymorphisme Humain (CEPH)/Généthon, we examined a total of 2,890 polymorphic

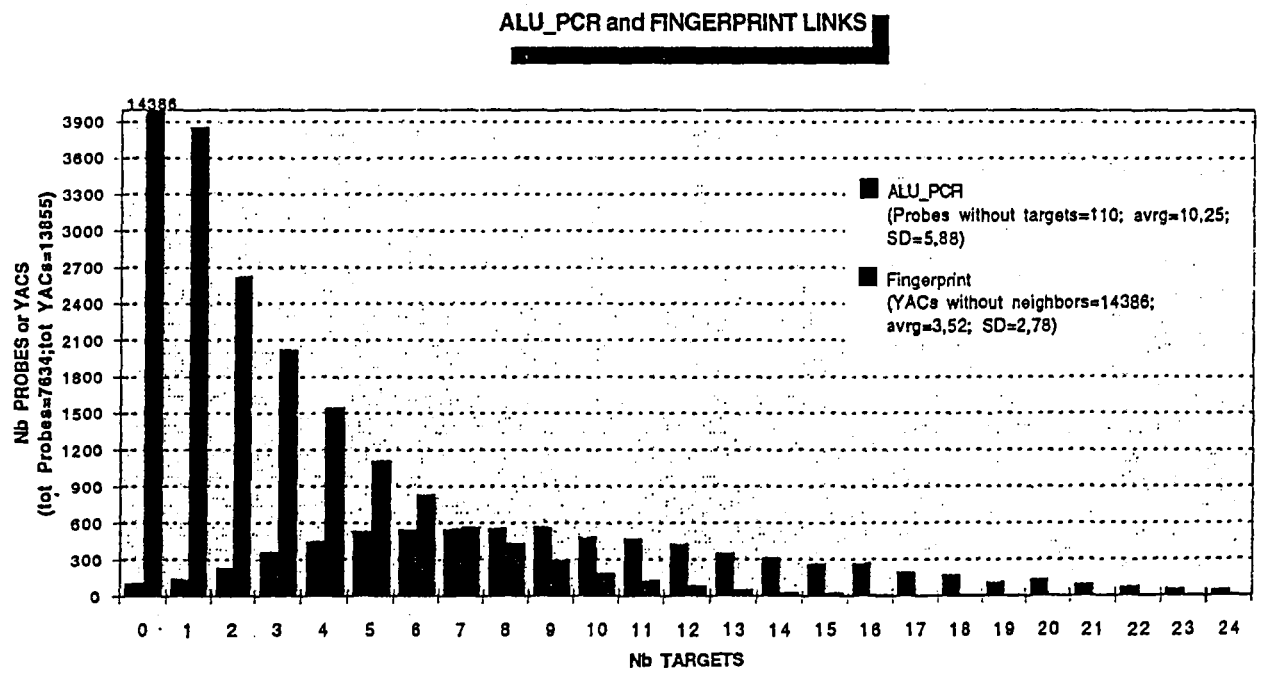


FIG. 2 Alu-PCR and fingerprint links. Blue bars represent Alu-PCR links, red bars represent fingerprint links.

markers, all generated and mapped by genetic linkage analysis as part of the Généthon genetic mapping program[3]. All markers were screened on the 43 superpools to identify the positive blocks. About 5% of the markers failed to work because of poor amplification or high background in the YAC pools. Another 5% gave no signal in the superpools, despite yielding the expected PCR product in a human genomic DNA control. In about 60% of these cases, we were able to detect and confirm a positive signal when the PCR products were electrophoresed, blotted and probed with a $(CA)_{15}$ oligonucleotide (which hybridizes to the CA repeat contained within the polymorphic locus).

Complete screening was performed for the first 814 markers, those of the first Généthon linkage map (1992). Of these, 28 failed to detect any YAC, and 786 identified 5.6 YACs on average. The 2,076 remaining genetic markers were subjected to directed screening. Of these, 261 failed to detect any YAC, and the remaining 1,815 identified an average of 4.9 YACs.

In total, 2,601 genetic markers identified at least one YAC. A total of 289 STSs have no anchored YACs, whether because of PCR-related problems or library-related problems. PCR-related problems are mainly due to sequence-dependent heavy background noise or poor amplification. In some of these cases, the design of another pair of primers from the original sequence data allowed us to obtain positive clones. YAC library-related problems can be due to the absence of clones in certain regions of the genome, either for statistical reasons or for non-clonability of certain human DNA sequences in yeast. The inability to find anchored YACs was more frequent for STSs located in certain regions of the genome, such as 1p, 19, the distal part of 17q, and most of the telomeric regions.

In addition to data generated at CEPH/Généthon, we also used results for 1,500 STSs screened elsewhere. The largest data set came from the Whitehead Institute/MIT Center for Genome Research (WI/MIT). We used the July 1994 release of this data, which contained 3,419 STSs screened with a different technology[19], using the 25,344 clones in plates 709–977 (the current publicly available release contains over 10,000 STSs and can be accessed via the World Wide Web, address 'www.genome.-mit.edu'). Among these STSs were 1,128 AFM markers also screened at CEPH/Généthon. Each group found an average of 1 definite YAC address per 2 genome equivalents screened: 5.1 YACs in 10 genome equivalents screened at CEPH/Généthon, and 4.1 YACs in 8 genome equivalents screened at WI/MIT. (Additional incomplete YAC addresses were also obtained, for example, about 1.5 at WI/MIT. These were still being resolved and are not used here.) The combined data provided more complete coverage than either group alone, as roughly two YACs were found in common, three only by CEPH/Généthon, and two only by WI/MIT.

Finally, we also incorporated results from about 370 STSs screened elsewhere and deposited in public databases.

Screening by hybridization

Methodology. We screened the YAC library by hybridization, using individual probes derived from individual YACs to screen the entire Mega-YAC library. To circumvent the tedious process of purifying YAC DNA from the total yeast genomic DNA, and also to increase efficiency of the hybridization, we derived from each YAC a representative set of human-specific DNA fragments by means of inter-Alu PCR, between the ubiquitous Alu repeats spread along the human genome[20,21]. This was achieved by PCR amplification from total yeast clone DNA with a single primer[22] specific for the 3′ part of the Alu repeat sequence. Under our conditions there was no amplification from yeast genomic DNA with this primer; on average 10 different fragments of 300 base pairs (bp) average size were produced from random Mega-YACs.

Alu-PCR products were prepared individually from each YAC to be used as probe or target. To simplify the screening procedure, we used a pooling scheme for the target Alu-PCR products. The pooling procedure was similar to the scheme used for STS screening, but in this case all subpools were simultaneously screened by hybridization. In the pooling scheme, 'blocks' consisted of 4 microtitre plates which were conceptually divided into 8 half-plates. From these 8 half-plates, a total of 22 subpools were prepared, consisting of 8 subpools containing clones in the same half-plate, 8 subpools containing clones in the same row of the half-plate, and 6 subpools containing clones in the same column of the half-plate. As this part of the library (plates 734–989) represents 64 blocks, the total number of subpools to screen is $64 \times 22 = 1,408$.

The pools were spotted at high density onto nylon membranes before hybridization. The addresses of positive candidates were deduced according to which half-plate, row and column pools were found positive for each block. The YACs identified by a single signal in each dimension were called 'unique positives'. If two candidate clones are present in a block, more than three signals will be observed. In general, the addresses of the positive clones cannot be deduced unambiguously under these conditions ('undetermined positives'). However, when such candidates are located on the same row or the same column of a single half-plate, it is possible to determine these positive clones ('determined positives'). Our experience indicates, that these determined positives can be used for the map construction, but rather cautiously, as some (or many) of them are false positives. One possible explanation of this phenomenon is that some of these 'determined' positives appear to be linked to artefactual spots due to hybridization background. Moreover, in some cases some of the three-dimensional signals could not be detected for technical reasons. This could interfere with the deduction of YAC addresses when using undetermined positives.

In addition to the Alu-PCR products from the YAC clones we also spotted in duplicate Alu-PCR products from a somatic cell hybrid panel consisting of cell lines, each containing only one or two human chromosomes. These hybridization targets provided information about the likely chromosomal localization of the YAC probes. Most of the cell hybrids were obtained from the NIGMS (Coriell Institute of Medical Research, Camden New Jersey) mapping panel 2 (ref. 23). A chromosome 20-only G418-resistant monosomic cell hybrid DNA was provided by C. Smith. GM10791, a chromosome 7-only somatic cell hybrid DNA was provided by E. Green; and GM06318B, a chromosome X-only somatic cell hybrid DNA was provided by D Schlessinger. In the second set of membranes used for this project, we also included somatic cell hybrids for chromosome 1+X, 5, 6, 12 and 19, provided by D. Patterson.

The Alu-PCR products of subpools and somatic cell hybrids were spotted onto membranes together with ϕX DNA for automatic filter identification. This spotting was performed by an automatic replicating device. The membranes were hybridized in the presence of human DNA competitor with ^{33}P-labelled mixture of phage ϕX DNA and Alu PCR products of individual YACs. A high-throughput protocol that included labelling in microplates and washing membranes in batches allowed a team of two people to hybridize 200 YAC probes per day. After washing and exposure, the films were scanned and images were stored on a workstation. After automatic treatment, all images were manually inspected so artefacts could be removed from analysis and the interpreted results checked (positive YACs deduced from the subgroups and chromosomal assignment) during the analysis. The software for this semiautomatic procedure was developed in collaboration with Cose (Paris).

The pilot hybridizations with freshly made membranes indicated that 80% of random YAC probes produced an effective hybridization result. The remaining 20% gave either no signal (4% of the cases) or high background noise. This latter phenomenon is probably associated with middle-frequency repeat sequences included occasionally in inter-Alu PCR amplification products. In most of these cases, we were also unable to determine the chromosomal origin of the probe.

Generally, hybridization to somatic cell hybrid inter-Alu PCR products was less effective than to YAC targets. In pilot experiments, only 80% of successful probes gave a signal to at least one duplicate of the chromosomal inter-Alu PCR products spots. In general, we observed a very good result reproducibility when the same YAC probe was used on different batches of membranes.

Results. We derived inter-Alu PCR products for each of the 2[illegible]576 YACs of the Mega-YAC library (about eight human genome equivalents) to be used as targets for hybridization. Probes were selected by various criteria.

The first 2,000 probes were YACs belonging to chromosome-specific sublibraries generated according the procedure described to obtain the chromosome 21-specific YAC subset[22]. Briefly, Alu-PCR products of clones from a four-genome equivalent portion of the Mega-YAC library were individually spotted on membranes and hybridized successively with chromosome-specific probes obtained with inter-Alu PCR DNA products from the panel of somatic cell hybrids.

We also used as probes 200 YACs cloned in Rad52- yeast strain[17]. According to the chromosomal assignment results from hybridization, this set appeared to contain only 8% of chimaeric YACs.

The rest of the probes were chosen using the QUICKMAP software. The first objective was to obtain for each genetic locus two YACs successfully used as probes. For this about 2,500 YACs were chosen with the 'locus' function. We also used the 2,000 largest YACs that were not genetically anchored. Finally, about 2,000 YACs were chosen with the 'CLONESPATH' function during the map confirmation.

In total, 8,785 probes gave interpretable signals in this screening procedure. As expected from the selection process, the size distribution of the probes is shifted towards larger size (Fig. 1). The distribution of the number of targets detected per probe is almost gaussian, with an average of 7 ('unique positives') (Fig. 2) or 10 (when adding 'determined positives'). This is approximately half of that expected with probes spanning 1 Mb of genome. The first reason is that we wanted to avoid false positives, so we kept only the clearest signals during the image analysis. The discrepancy can also be explained by non-random distribution of inter-Alu PCR products and unequal efficiency of their individual hybridization. The distribution of YAC target sizes is also shifted towards the larger size, probably because larger clones are likely to produce more inter-Alu PCR products and so will provide stronger signals. This may also account for the larger size of successful YAC probes. In total, 20,890 (85%) of YACs were linked by hybridization to at least one other YAC. In most cases, a given YAC is detected as a target when it is used as a probe. The signal obtained is generally very intense. However, pools containing adjacent clones in the corresponding plate often produce a signal as well, probably because of minor cross-contamination. These artefacts interfere with the evaluation of positives in the corresponding pools, so targets could appear as 'undetected' in the database.

A total of 7,209 probes were assigned to chromosomes based on hybridization. Although the chromosomal assignment by inter-Alu PCR is simple, care should be taken in interpreting the results. For example, supposedly monochromosomal hybrids often contain insertions of small chromosomal fragments and deletions of other chromosomal regions. This was experimentally confirmed for the NIGMS mapping panel II used in our work. We also found by conducting reciprocal hybridization between these somatic cell hybrids that inter-Alu PCR products from some of them cross-hybridize. The most striking overlap was detected between chromosome 5-'only' and chromosome 6-'only' hybrids, as well as between chromosome 12-'only' and chromosome 6-'only' hybrids. The same pattern of cross-hybridization was observed with YAC probes. This cross-hybridization could, in some cases, be due to repeated or duplicated genomic regions.

In addition to problems with the hybrid cell lines themselves, false chromosomal assignment could result from laboratory error or sequence similarity causing cross-hybridization. Alternatively, false negatives could be due to inefficient hybridization with inter-Alu PCR products from certain YACs, or deletion of the corresponding region in the somatic cell hybrid.

Chromosomal assignment by hybridization assists in the detection of chimaeric YACs, but will obviously miss some chimaeras, including those containing only a small portion from a different chromosome region, those containing a region that is poor in Alu repeats, and those consisting of two fragments of the same chromosome. However, some apparent chimaeras could result from sequence similarity between several chromosomes. Despite these difficulties, we have used this result to analyse the chimaerism rate according to the library origin and the size distribution of the YACs (Fig. 1). The Mark II library contains a greater proportion of chimaeric YACs than the Mega-YAC library. In the Mark II library, the very large YACs seem to be more chimaeric than the smaller ones, but this is not the case for the Mega-YAC library.

Because of these interpretation problems, we treated the chromosomal assignment data with extreme caution in the QUICKMAP software, where the criteria of assignment depended on several parameters which varied according to the genomic region.

Fingerprinting

Methodology. To detect overlaps among YACs, we performed fingerprint analysis as previously described[13]. Each YAC DNA was digested with three enzymes: *Eco*RI, *Pvu*II and PstI. after agarose gel electrophoresis, the fragments were transferred onto nylon membranes using a robot. Membranes were then hybridized successively with two probes: human repeated sequences LINE-1 (LI)[24], and THE-LTR (transposon-like human-element long terminal repeat: THE)[25]. The corresponding patterns were captured automatically after scanning each film. The size of each fragment was extrapolated from the migration length of reference markers with known sizes which were run in parallel.

The L1 and THE probes were selected as they gave 6 and 11 bands per megabase, respectively. We attempted to use other repetitive probes, such as Alu, medium reiteration frequency repeats (MER)[9] and poly(GA), but with little success. The Alu probe patterns were too complex, and the MER and poly(GA) probes gave rather poor patterns with 27.6, 12.8 and 15.6% negative clones for MER 1, MER 10 (ref. 26) and poly(GA), respectively. Promising results were obtained with two probes for two Alu subfamilies, GA.007 (ref. 27) and 5OS (ref. 28), but these were poorly reproducible.

Results. A total of 31,392 YACs were successfully fingerprinted. Of these, 12.5% gave no bands for L1, 7.3% gave no bands for THE, and 4% were negative for both. When hybridized with an Alu consensus probe, one-third of these L1/THE-negative clones gave no Alu bands. The remaining clones (L1/THE-negative clones with Alu bands) contained inserts half the size of L1/THE-positive clones.

Pairwise comparisons were performed among all the fingerprints as described previously[13], and a likelihood of overlap score (LOS) was determined for each pair of clones and for each probe. Only YAC pairs having a LOS value greater than or equal to 70 for both L1 and THE were declared linked. These threshold values were chosen according to criteria based on the analysis of YACs previously mapped on chromosome 21 for which an extensive study had been performed[3]. We considered all possible pairs of YAC probes for which a chromosomal assignment was obtained by hybridization on somatic cell hybrid DNA. In this set, 70% of YAC pairs linked by fingerprint data were assigned to the same chromosome by hybridization (concordant pairs). Similarly, 68% of YAC pairs linked by hybridization showed concordant chromosomal assignment. (Interestingly, YAC pairs with reciprocal links by hybridizaton

showed 82% concordant assignment. As a control, random YAC pairs show only 8% concordance.)

In total, 17,006 YACs with these threshold values were linked to at least one other YAC from the library. On average, each of these YACs was associated with 5.8 YACs. The size distribution of these 17,006 YACs is shifted significantly towards larger sizes (1,119 kb on average). Larger YACs containing more bands would be expected to be more informative. A comparison with STS and hybridization data enabled us to detect 22 plates giving an abnormally high number of links due to a conserved fingerprint pattern in all of them. We suspect that well-to-well contamination occurred during the fingerprint process, and we removed these 22 plates from analysis. The corresponding clones made available in 1992 are free of this contamination.

Construction of the map

The starting point of the map was the framework of STSs given by the Généthon 1993–1994 linkage map[3]. This map contains 2,066 polymorphic markers, ordered in 1,267 genetic loci, each of which corresponds to a bin of 1–7 polymorphic STSs that were not recombinationally resolved. We used the three types of links between YACs (based on STS content, fingerprint and hybridization) to assemble contigs that span the intervals between genetically adjacent STSs. During this process, we integrated new STSs to this map to refine the framework order and strengthen the contigs. The limitations and precautions taken in building consistent contigs are discussed here, as each data type has its own limitations and error rate.

First, we define a minimal path between two STSs, S_1 and S_2 as an ordered list of YACs $(Y_1, \ldots, Y_n)$ that satisfy the following conditions: (1) Y_1 and Y_n contain S_1 and S_2, respectively; (2) for each $i=1, \ldots, n-1$, the YACs Y_i and Y_{i+1} are linked by one of the three mapping methods; and (3) there is no link between YACs that are not consecutive in the list. The number n of YACs in the minimal path is called the level of the path.

For several reasons, minimal paths do not necessarily represent valid 'contigs' of sequences that actually overlap in the human genome. Most importantly, chimaeric YACs artefactually join distant segments of genomic DNA, establishing connections between pairs of distant STSs. Such YACs represent between 30 and 50% of the library, depending on the genomic regions. Similarly, false positive links between clones can also result from hybridization or fingerprinting. Such false positives may make up 5–10% of the links.

Because of these problems, the backbone information from the genetic map is crucial for building accurate contigs. First, we only look for YAC paths connecting nearby STSs. Second, we can exclude some YACs that appear to be chimaeras based on their containing STSs from distinct locations, based both on Alu-PCR and the STS data (see below for more information about elimination of false links).

Contig assembly algorithm. The algorithm for constructing paths between two nearby loci proceeds by the construction of progressively larger 'neighbourhoods' of YACs. For each locus

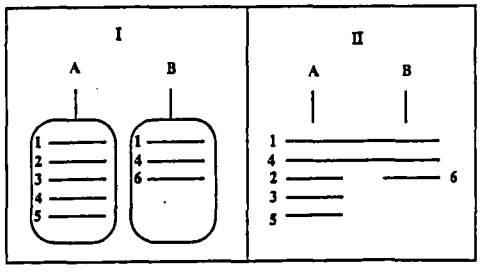


FIG. 3 Construction of a level 1 path between two loci A and B. Stage I is the construction of the first-degree neighbour set for each locus. In stage II, YACs 1 and 4 are found in common. These YACs establish the level 1 path.

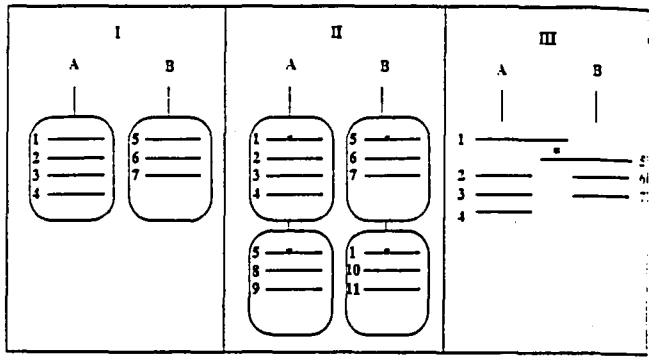


FIG. 4 Construction of a level 2 path. After stage I, no common clones are found. Stage II is the construction of the second-degree neighbour set for each locus. The asterisk documents the link between clones 1 and 5, which establish a level 2 path (stage III).

x, the computer can construct the set $N_{x,1}$ of first-degree neighbours consisting of anchored YACs (that is, YACs containing at least one STS in the locus); the set $N_{x,2}$ of second-degree neighbours consisting of YACs linked to those in $N_{x,1}$; the set $N_{x,3}$ of third-degree neighbours consisting of YACs linked to those in $N_{x,2}$; and so on. Any overlap between the neighbourhoods of loci x and y clearly yields a path connecting them. (More precisely, a YAC present in both $N_{x,i}$ and $N_{y,j}$ yields a path of length $i+j-1$.) In practice, the computer program constructs increasing neighbourhoods around both loci, halting as soon as an overlap is found. Examples are illustrated in Figs 3–5.

In attempting to link nearby loci on a given chromosome, we used positional information in an attempt to avoid paths that branch to distant parts of the genome. In forming second-degree and higher neighbourhoods, we excluded YACs exclusively assigned to other chromosomes by Alu-PCR hybridization, and also excluded STS-content links involving STSs from other regions.

Although the genetic linkage map represents the most likely genetic order, some local marker orders may be inverted. Accordingly, we searched not only for paths between immediately consecutive STSs (such as i and $i+1$), but also between nearby but non-consecutive STSs (such as i and $i+2$). For such non-consecutive STSs, the genetic distance and the number of intervening STSs was constrained depending on the level of the path.

Manual inspection of the paths. The map construction algorithm was applied to the whole genome. Each candidate path was then subjected to several types of checking. The first step involved graphical inspection using the 'CLONESPATH' par

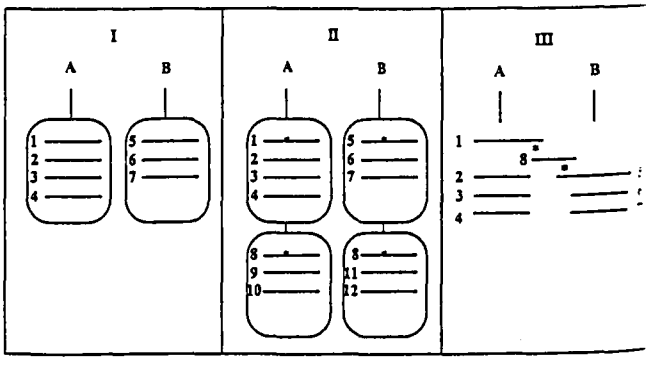


FIG. 5 Construction of a level 3 path. After stage II, clone 8 is found in common between the two second-degree neighbour sets. The level 3 path is represented at stage III.

of the program to evaluate paths based on the following criteria: (1) the number of YACs in the path; (2) The density of links between YACs; and (3) the extent to which YACs in the path were chromosomally assigned (by Alu-PCR hybridization, other STS-content information, or fluorescence *in situ* hybridization (FISH)). Graphical inspection also allowed us to detect and reject cases in which two independent paths linked the two STSs. After such visual inspection, we could reject candidate paths, try to generate a new candidate path (by trying new parameters in the algorithm or changing the order of STSs), or perform additional STS screening to test the paths further.

We tried to improve candidate paths that were judged satisfactory after graphical analysis. We derived new STSs from the ends of internal clones in a path. We also subjected the most critical clones to Alu-PCR hybridization to test their chromosomal assignment and to establish more links between clones in the path. This strategy often shortened paths by indicating overlaps that had not previously been detected because our STS screening was incomplete. To illustrate this point, Fig. 6 shows the result of an incomplete STS screening, and Fig. 7 shows the result of incomplete hybridization data. In particular, paths of level 6 or 7 in our 1993 version were converted to shorter paths. The present map contains now only paths of level 5 or less.

The bins may contain several markers that, although not recombinationally separated, span a certain distance in the physical map. To cover the physical region within the bin, we searched for paths linking different STSs within a bin. Many bins were covered by paths of level 1. For the remaining bins, we tested YACs positive for one of the STSs in a bin with the other STSs in the bin. In some cases, we used the locus program to close gaps between STSs with paths of higher levels. Fewer than 10% of the bins are not completely covered in the present map.

Because we constructed paths between genetic markers that were not necessarily adjacent in the linkage map (see above), we sometimes encountered cases in which the shortest paths connected markers i with $i+2$ and markers $i+1$ with $i+3$. This situation could arise for two reasons: we could have missed actual overlaps in the paths owing to false negative screening results (Figs 6 and 7), or the putative order on the genetic linkage map could be incorrect. To preserve the linearity of the map in these rare cases, we have either inverted the marker order or joined recombinationally separable genetic markers in the same bin. As a result, the physical order of the markers on the summary figures of the atlas (see below) does not perfectly correspond to the Généthon 1994 linkage map.

Integration of other STSs. In addition to the backbone STSs taken from the 1994 Généthon genetic linkage map, we also integrated some additional STSs that improved paths in the map. These markers came from two sources. First, WI/MIT had screened 3,419 STSs against the YACs by June 1994. From this

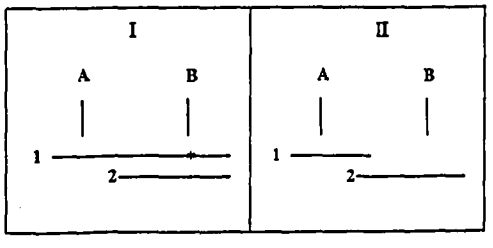


FIG. 6 Part I represents the real disposal of the YACs. If STS B is not tested against YAC 1, the path would appear at level 2 (provided the overlap between the two YACs is detected), as shown in part II. In such a case, we would have tested A against 2 and B against 1 and reduced the level to the actual value.

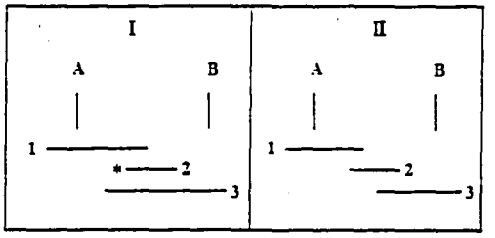


FIG. 7 Part I represents the real disposal of the YACs. If neither YAC 1 nor YAC 3 is used as Alu-PCR probe, the hybridization between them cannot be detected. In this case, YAC 2 was used as probe and detected the YACs 1 and 3. The path appears to be level 3, as shown in part II. This situation can be resolved by testing either 1 or 3 as Alu-PCR probe.

STS-content data, we selected 173 STSs (including 76 non-AFM genetic markers) that significantly improved paths. Second, the CEPH/Généthon group screened STSs from 445 unpublished genetic markers from Généthon (C. Dib *et al.*, manuscript submitted). Where known, chromosomal assignment or approximate map position was used for both sets of markers. In most cases the integration produced denser contigs and decreased the level of paths (see Fig. 6).

FISH mapping

A total of 650 genetically anchored YACs, approximately one every 5–10 cM, were selected and used as probes for fluorescent *in situ* hybridization (FISH) on metaphase chromosomes. The chimaerism rate detected by this method was 46%. Based on the comparison of cytogenetic and genetic localizations, there appear to be higher frequencies of recombination near telomeres and lower frequencies near centromeres. For example, the genetic distance between the centromeric markers D1S440 (at 163 cM) and D1S484 (at 182 cM) represents 6.5% of the genetic length of this chromosome, but 17% of the fractional cytogenetic length of the chromosome. Similarly, the interval between the centromeric markers D6S272 (at 75 cM) and D6S421 (at 86 cM) represents 5.3% of the genetic map but 17% of the fractional length of chromosome 6. In contrast, the telomeric loci D6S411 (at 173 cM) and D6S281 (at 207 cM) are separated by 16.4% of the genetic length of the chromosome, but the interval between D6S411 to the telomere is only 4% of the fractional cytogenetic distance. The FISH analysis indicates that there are no genetic markers on 13p, 14p and 15p, and that the terminal region of chromosome 20q is not contained in the genetic map. Thus there is no coverage by YAC of these regions in our physical map.

Presentation of the map

For each chromosome, the atlas following this paper shows: (1) a summary map of each chromosome, showing the cytogenetic representation, and the scales of the physical and genetic maps, together with the indication of the regions covered in contigs; and (2) a map of detailed contigs for each chromosome.

Summary figure. Each chromosome is presented at the same genetic scale as an ideogram[29] at the left side of each drawing. At the right of this ideogram is the physical map scale, showing the location of the bins. In parentheses are two numbers, separated by a semicolon: the first number is the number of STSs in the bin, the second is the number of YACs that are anchored to the bin. The links between the physical map scale and the cytogenetic scale are established through YACs that are anchored to the bins and have been used in FISH hybridization. Each of those YACs establishes a connection between the bin and an interval in the ideogram. Note that the FISH measurements have been made in terms of fractional length of the whole chromosome, and that the size of heterochromatic and centromeric regions may vary between individuals. As a consequence, a slight distortion can occur in our figures after these regions, especially for chromosomes containing entirely heterochromatic p arms

(acrocentrics). At the right of the physical map, the intervals covered in contigs are represented with coloured rectangles. The different colours represent the different levels of the paths. Finally, the correspondence between the physical map scale and the Généthon linkage map, used as a backbone for the bin locations, is shown at the right of these rectangles. The positions in the genetic map are expressed in Morgans from the most distal marker of the p arm of the chromosome.

Detailed contigs. Contigs are presented for each chromosome from pter to qter. They correspond to a succession of paths, represented by rectangles on the summary figure. Each path is a collection of clones, ordered in stacks. The number of stacks in a path corresponds to the level of the path. The graphical presentation of paths provides the following characteristics of the clones: STS-content information for the YACs; sizes of the YACs; overlap relationships between YACs based on Alu-PCR hybridization and fingerprint data; chromosomal assignment for YACs used as probes for Alu-PCR hybridization; and indication of YACs used as FISH probes.

Each locus is indicated by a white rectangle that indicates its chromosome and position. STSs located in the bin are displayed above the rectangle. These STSs are numbered within the bin and are displayed in a beige rectangle. For example, the bin located at position 1.00 on chromosome 1 contains two STSs: AFM120xd4 (D1S209), and AFM286xd9 (D1S473). This bin is thus presented as:

1: AFM120xd4 (D1S209)
2: AFM286xd9 (D1S473)
CHR 1 position 1.00

The clone stacks displayed under each bin represent the anchored YACs (that is, the YACs that contain at least one STS of the bin), and the stacks that are between two bins represent the ordered groups of clones internal to paths of level 3 and above. Within a stack, the YACs are displayed according their order in the library, from top to bottom. Each YAC is represented by a yellow box with a horizontal bar in the middle. The YAC name and its size in kilobases are represented from left to right above the bar. A '+' sign after the size means that multiple bands were detected; only the biggest size is displayed. The names of YACs used for FISH hybridization appear in a box (for example, YAC 763B12, anchored to position 1.00 on chromosome 1). Chromosomal assignment (for clones used as Alu-PCR hybridization probe) and the STS content of the YAC are represented from left to right under the bar.

Chromosomal assignment is made based on the results of hybridization with somatic cell hybrids. Because of the problems described above, chromosomal assignments were sometimes ambiguous. The assignments are represented by the following code: (1) one white dot: a probe that was not assigned; (2) two blue dots: a probe assigned only to the chromosome under consideration; (3) one blue dot, one orange dot: a probe assigned to the chromosome under consideration, as well as to one or more additional chromosomes; and (4) two orange dots: a probe that is not assigned to the chromosome under consideration, but that is assigned to one or more other chromosomes.

The display of the STS information differs between stacks composed of anchored YACs and stacks located between two bins. For anchored YACs, the stack shows the clone numbers of the STSs in the bins. For example, the bin at position 1.00 on chromosome 1 shows YACs 631C9, 732A10 and 752E3; they contain, respectively, the second, the first, and both STSs, and are given the lists '2', '1' and '1 2'. For stacks between bins, we represent the position of the STS for which the YAC is positive. For example, '1-0.87' means position 0.87 on chromosome 1, '1-?' means chromosome 1 but position unknown on this chromosome, and '?' means that no positional information is known. If the YAC is positive for several STSs, located at different places then asterisks are displayed.

The relationship between clones in adjacent stacks is shown as follows. For paths of level 1, the path is established through the presence of one or more clone in the adjacent anchored stacks, with a thick bar (yellow and black) displayed between the two stacks. For example, paths of level 1 are established between the loci 1.00 and 1.02 on chromosome 1 by the YAC 752E3, 763B12, 830E7 and 940C1. No Alu-PCR hybridization or fingerprint linkage is involved in establishing paths of level 1.

For level 2 and higher, an array of one or more columns is displayed between the stacks, representing the fingerprint and Alu-PCR relationships that link the stacks. Each column of the array is composed of a black box and 3 subcolumns. The horizontal position of the black box relative to the column gives the orientation for reading the columns. All columns within an array have the same orientation. If the box lies on the left (respectively, right) of the column, this column refers to the clone of the left (respectively, right) stack that is vertically in the same place as the box. We call this clone the attached clone. The three subcolumns contain symbols (dots and triangles) that refer in this case to the clones of the right (respectively, left) stack that are vertically in front of them. The subcolumn that is just to the side of the black box can be either yellow or pale blue. It is yellow if the attached clone was not used as Alu-PCR target (does not belong to plates 734–989). If this clone was used as an Alu-PCR target, this subcolumn is pale blue and the triangles in it refer to the Alu-PCR probes that hit this clone by hybridization. The middle subcolumn is yellow if the attached clone was not used as Alu-PCR probe. This subcolumn is blue otherwise and the triangles in it refer to the targets hit by this clone. The third subcolumn is yellow if the attached YAC was not fingerprinted. If this clone was fingerprinted, it is green and contains black dots that refer to overlapping YACs by fingerprint. This two-colour presentation allows the reader to distinguish for example between a clone that was not used as an Alu-PCR probe from one that was used as an Alu-PCR probe but did not hit any YAC in the adjacent stack. It also provides a very quick way of highlighting the clones with the most overlap information, which are the ones linked to the region with the highest probability.

As an example, in the path between 1.08 and 1.10 on chromosome 1, the first column refers to the YAC 912G11 and the second to the YAC 957A9, because the black box lies on the left of the column. The YAC 912G11 was used as a target for Alu-PCR hybridization, was not used as an Alu-PCR probe and was fingerprinted. As a target for Alu-PCR hybridization it is hit by YACs 774C4, 800E10 and 943A2 as hybridization probes. It overlaps by fingerprint data with YAC 895B12. The YAC 957A9 was used as a target for Alu-PCR hybridization, was used as an Alu-PCR probe, and was fingerprinted. As an Alu-PCR target, it is hit by the probe 927C3. As an Alu-PCR probe, it hit YAC 927C3. It also overlaps by fingerprint data with YAC 927C3.

Map reliability and coverage

All YAC paths covering genetic intervals have been inspected and checked, as described above. Contigs cover 75% of the genetic intervals, which together comprise 66% of the total genetic length of the genome (based on the sex-averaged meiotic map). The proportion of the genetic length covered in paths of level 1 is 26%, of level 2 is 17%, of level 3 is 15%, of level 4 is 5%, and of level 5 is 2%. These numbers are calculated on the basis of all chromosomes except 3, 12, 21, 22 and Y, which were either previously mapped or subjected to more intensive mapping by groups presenting their maps elsewhere in this volume. The chromosomes that are covered for more than 66% of their genetic length are: 4, 5, 7, 8, 9, 11, 14, 15, 16, 18 and 20. Low coverage of chromosome X (23% of its total genetic length) is expected, both because the YAC library underrepresents the

X chromosome (being derived from a male) and because the genetic map of X is sparse. In addition, the screening efficiency with backbone STSs from 1p, 19 and 17 was particularly low, resulting in poor coverage of these regions. However, chromosome 18 is almost entirely covered because we made a special effort to fill gaps by more intensive hybridization screening effort with YAC probes derived from path extremities. In many cases, new STSs were derived from those YAC probes creating new paths.

Inferring the actual proportion of the physical length of the genome covered is not entirely straightforward. The proportion of the genetic length covered (66%) may overestimate the actual proportion covered because it neglects the physical distances within the bins of recombinationally inseparable markers. But it may underestimate the coverage because the density of YACs appears to be sparsest in the telomeric regions, which are precisely those regions in which the ratio of genetic to physical distance appears to be greatest.

If the AFM markers were randomly distributed they would occur with a random spacing of about 1.2 Mb, just slightly larger than the average size of our YAC clones. We would thus expect to cover much of the genome in paths of level 1, 2 or 3. The observed proportion of intervals covered by such paths agree very well with expectation based on computer simulations. Moreover, mathematical analysis predicts that the YACs identified by the genetic markers would contain about half of the genome[30]. Although the AFM markers are known to be not completely randomly distributed, the overall effect of non-randomness appears not to be severe.

To evaluate the reliability of the YAC contig map, we examined 161 non-AFM genetic markers from a recently published collaborative genetic map[31] of the human genome (omitting markers for chromosomes 3, 12, 21 and 22) that were screened against the YAC library at WI/MIT. Of the STSs, the number detecting 1, 2 or at least 3 YACs was 20, 14 and 127, respectively 1, 2 and 3. In 60, 78 and 88% of the cases, respectively, the YACs containing the markers had been assigned to the expected location (based on the known genetic location of the marker). In the remaining cases, the corresponding YACs were not found on the expected chromosome but were sometimes found on other chromosomes (possibly resulting from chimaerism).

These results also indicate that the map covers most of the human genome. However, they do not provide a direct estimate of coverage because only loci that detected at least one YAC were considered, and the genetic markers tested may tend to lie in the same regions as the genetic markers used to construct the map. Mitigating against this concern is that the genetic markers used to assess coverage were predominantly tetra-nucleotide repeats, whose regional biases may differ from the CA repeats in the Généthon genetic map.

Given the results above, it seems reasonable to estimate that the physical map covers about 75% of the genome in 225 contigs having an average size of about 10 Mb.

Public availability of the map

Clone availability. Primary copies of CEPH YAC library were distributed to following centres:

Whitehead Institute/MIT Center for Genome Research, Cambridge, Massachusetts 02142, USA; E. S. Lander and T. Hudson; e-mail: lander@genome.wi.mit.edu.

The Reference Library DataBase (RLDB), MPI for Molecular Genetics, Ihnestrasse 73, 14195 Berlin-Dahlen, Germany; H. Lehrach; tel: (49) 30 8413 1627; fax: (49) 30 8413 1395.

3-1-1 Koyadai, Tsukuba, Ibaraki 305, Japan; K. Yokoyama; tel: (81) 298 36 3612; fax: (81) 298 36 9120.

Human Genome Center, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato, Tokyo 108, Japan; Y. Nakamura; tel: (81) 3 5449 5372; fax: (81) 3 5449 5433.

Shanghai Institute of Hematology, Rui-Jin Hospital, Shanghai Second Medical University, Shanghai 200025, China; Z. Chen; tel: (86) 21 3180 300; fax: (86) 21 4743 206.

GBE, CNR, via Abbiategrasso 207, 27100 Pavia, Italy; D. Toniolo; tel: (39) 382 546 340; fax: (39) 382 422 286.

YAC Screening Centro, Leiden University, Department of Human Genetics Wassenaarseweg 72, 2333 Al Leiden, The Netherlands; G. J. B. van Ommon; tel: (31) 71 276 081; fax: (31) 71 276 075.

Human Genome Mapping Project Resource Centre, (HGMP) Hinxton Hall, Hinxton, Cambridge CB10 1RQ, U.K.; K. Gibson; tel: (44) 1223 494 500; fax: (44) 1223 494 512.

Clones can be obtained also from Foundation Jean-Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France; D. Le Paslier; e-mail: denis@ceph.cephb.fr.

Data distribution

Anonymous ftp server: ftp://ceph-genethon-map.cephb.fr/pub/ceph-genethon-map.

World Wide Web server: URL address: http://www.cephb.fr/bio/ceph-genethon-map.html.

Mail server: ceph-genethon-map@cephb.fr.

How to use it: $ mail ceph-genethon-map@cephb.fr. Subject: infoclone. 755_f_4 672_a_3 D12S76.

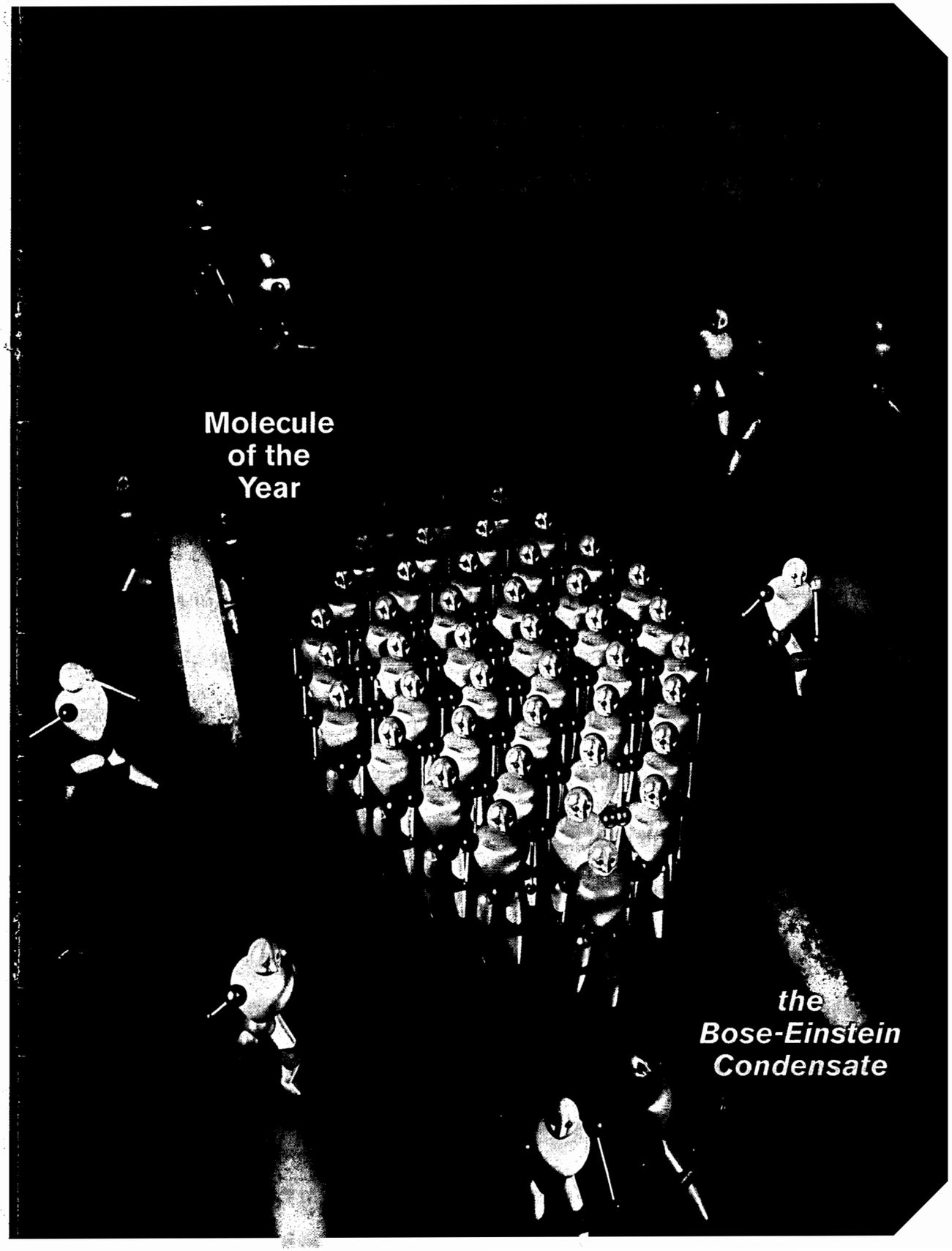
other YAC or STS names.

QUICKMAP (developed by P. Rigault and E. Poullier at CEPH) is a mapping tool containing all the CEPH/Généthon screening data. It was designed to manage the production of STS screening and hybridization data, using the results analysed on a daily basis to suggest new tests. It was then modified to make CEPH/Généthon data accessible to the scientific community. QUICKMAP allows both navigation within CEPH/Généthon map and dynamic construction of contigs to integrate further datasets. QUICKMAP has been publicly available since February 1993 on our ftp site. □

Received 29 September 1994; revised 6 July 1995; accepted 17 July 1995.

1. McKusick, V. A. *Mendelian Inheritance in Man* 5th edn. (John Hopkins Univ. Press, Baltimore, 1978).
2. Burke, D. T., Carle, G. F. & Olson, M. V. *Science* **236**, 806–812 (1987).
3. Chumakov, I. *et al. Nature* **359**, 380–386 (1992).
4. Foote, S., Vollrath, D., Hilton, A. & Page, D. *Science* **258**, 60–66 (1992).
5. Miki, Y. *et al. Science* **266**, 66–71 (1994).
6. The Huntington Disease Collaboration Research Group *Cell* **72**, 971–983 (1993).
7. Collins, F. S. *Nature Genet.* **1**, 3–6 (1992).
8. Gyapay, G. *et al. Nature Genet.* **7**, 246–339 (1994).
9. Litt, M. & Luty, J. A. *Am. J. hum. Genet.* **44**, 397–401 (1989).
10. Weber, J. L. & May, P. E. *Am. J. hum. Genet.* **44**, 388–396 (1989).
11. Green, E. D., Riethman, H. C., Dutchik, J. E. & Olson, M. V. *Genomics* **11**, 658–659 (1991).
12. Olson, M. V., Hood, L., Cantor, C. R. & Botstein, D. *Science* **245**, 1434–1435 (1989).
13. Bellanne-Chantelot, C. *et al. Cell* **70**, 1059–1068 (1992).
14. Cohen, D., Chumakov, I. & Weissenbach, J. *Nature* **366**, 698–701 (1993).
15. Yang, S. Y. in *Immunobiology of HLA*, Vol. 1 (Springer, New York, 1989).
16. Albertsen, H. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 4256–4260 (1990).
17. Haldi, M. *et al. Genomics* **24**, 478–484 (1994).
18. Carle, G. F., Franck, M. & Olson, M. V. *Science* **232**, 65–68 (1986).
19. Bell, C. J. *et al. Hum. molec. Genet.* **4**, 59–69 (1995).
20. Nelson, D. L. *et al. Proc. natn. Acad. Sci. U.S.A.* **86**, 6686–6690 (1989).
21. de Jong, P. J. *et al. Cytogenet. Cell Genet.* **51**, 985 (1989).
22. Chumakov, I. M. *et al. Nature Genet.* **1**, 222–225 (1992).
23. Mullivor, R. A., Greene, A. E., Drwinga, H. L., Toji, L. H. & Kim, C. *Am J. hum. Genet.* (Suppl.) **49**, 370 (1991).
24. Shafit-Zagardo, B., Maio, J. J. & Brown, F. L. *Nucleic Acids Res.* **10**, 3175–3193 (1982).
25. Fields, C. A., Grady, D. L. & Moyzis, R. K. *Genomics* **13**, 431–436 (1992).
26. Kaplan, D. J., Jurka, J., Solus, J. F. & Duncan, C. H. *Nucleic Acids Res.* **19**, 4731–4738 (1991).
27. Matera, G. A., Hellmann, U., Hintz, M. F. & Schmid, C. W. *Nucleic Acids Res.* **18**, 6019–6023 (1990).
28. Jurka, J. & Milosavljevic, J. *Molec. Evol.* **32**, 105–121 (1991).
29. Francke, U. *Cytogenet. Cell Genet.* **65**, 206–219 (1994).
30. Aratia, R., Lander, E. S., Tavare, S. & Waterman, M. S. *Genomics* **11**, 806–827 (1991).
31. Murray, J. C. *et al. Science* **265**, 2049–2054 (1994).

ACKNOWLEDGEMENTS. This work was supported by Association Francaise contre les Myopathies (AFM), Ministere de la Recherche et l'Enseignement Superieure, Groupement de Recherche et d'Etude des Genomes, la Ligue contre le Cancer, the European Economic Community program DGXII and the National Center for Human Genome Research of the US NIH.



COVER

Artist's conception of supercooled atoms in an exotic phase of matter called the Bose-Einstein condensate, which is Molecule of the Year for 1995. Each of the atoms in the condensate (in blue) has the same quantum mechanical wave function, and so they all move as one. Atoms outside the condensate move faster and in all directions. See Editorial on page 1901 and Molecule of the Year section beginning on page 1902. [Illustration: Steve Keller]

An STS-Based Map of the Human Genome

Thomas J. Hudson,* Lincoln D. Stein, Sebastian S. Gerety, Junli Ma, Andrew B. Castle, James Silva, Donna K. Slonim, Rafael Baptista, Leonid Kruglyak, Shu-Hua Xu, Xintong Hu, Angela M. E. Colbert, Carl Rosenberg, Mary Pat Reeve-Daly, Steve Rozen, Lester Hui, Xiaoyun Wu, Christina Vestergaard, Kimberly M. Wilson, Jane S. Bae, Shanak Maitra, Soula Ganiatsas, Cheryl A. Evans, Margaret M. DeAngelis, Kimberly A. Ingalls, Robert W. Nahf, Lloyd T. Horton Jr., Michele Oskin Anderson, Alville J. Collymore, Wenjuan Ye, Vardouhie Kouyoumjian, Irena S. Zemsteva, James Tam, Richard Devine, Dorothy F. Courtney, Michelle Turner Renaud, Huy Nguyen, Tara J. O'Connor, Cécile Fizames, Sabine Fauré, Gabor Gyapay, Colette Dib, Jean Morissette, James B. Orlin, Bruce W. Birren, Nathan Goodman, Jean Weissenbach, Trevor L. Hawkins, Simon Foote, David C. Page, Eric S. Lander*

A physical map has been constructed of the human genome containing 15,086 sequence-tagged sites (STSs), with an average spacing of 199 kilobases. The project involved assembly of a radiation hybrid map of the human genome containing 6193 loci and incorporated a genetic linkage map of the human genome containing 5264 loci. This information was combined with the results of STS-content screening of 10,850 loci against a yeast artificial chromosome library to produce an integrated map, anchored by the radiation hybrid and genetic maps. The map provides radiation hybrid coverage of 99 percent and physical coverage of 94 percent of the human genome. The map also represents an early step in an international project to generate a transcript map of the human genome, with more than 3235 expressed sequences localized. The STSs in the map provide a scaffold for initiating large-scale sequencing of the human genome.

A physical map affording ready access to all chromosomal regions is an essential prerequisite for the international effort to sequence the entire human genome. In the shorter term, it is also a key tool for positional cloning of disease genes and for studies of genome organization. Physical maps have evolved over the past decade from their initial conception as a set of overlapping clones (*1*) to the more recent idea of a well-spaced collection of unique landmarks called sequence-tagged sites (STSs), each defined by a polymerase chain reaction (PCR) assay (*2*–*4*). The U.S. Human Genome Project, for example, has set a target of a physical map consisting of 30,000 STSs spaced at intervals of about 100 kb (*5*).

By focusing on STS landmarks, genome researchers sought to insure against the inevitable problems inherent in any given clone library (*2*). The wisdom of this approach was borne out as it emerged that yeast artificial chromosomes (YACs), the best clones for covering large distances, suffer from high rates of chimerism and rearrangement and thus are unsuitable for genomic sequencing (*6*, *7*). STS-based maps sidestep this problem by having a sufficiently high density of landmarks that one can rapidly regenerate physical coverage of any region by PCR-based screening of clones appropriate for sequencing—such as cosmids, bacterial artificial chromosomes, and P1-artificial chromosomes (*8*).

STS-based physical maps with extensive long-range continuity have been constructed for only a handful of human chromosomes: 3, 12, 16, 21, 22, and Y (*3*, *4*, *9*, *10*). These combined maps cover just less than 20% of the genome with about 1600 STSs, and the average spacing on most of these chromosomes is about 250 kb. Projects are also underway for a few additional chromosomes (*11*). An international collaboration among the Centre d'Etude du Polymorphisme Humain (CEPH), Généthon, and Whitehead genome centers has also produced a clone-based physical map estimated to cover up to 75% of the genome in overlapping YAC clones (*7*). The map is clone-based, rather than STS-based, because it was primarily assembled by detecting physical overlaps among the clones themselves (by means of cross-hybridization and fingerprinting methods), with only a sparse set of STS landmarks used as anchors (786 loci fully screened and 1815 loci partially screened on YACs). The map is quite valuable for positional cloning projects, but it does not provide a scaffold for sequencing the human genome: The YAC clones themselves are not suitable for sequencing, and the STS coverage is too sparse to regenerate substantial physical coverage.

Here, we report the construction of an STS-based physical map of the human genome containing more than 15,000 loci, with an average spacing of 199 kb. The map covers the vast majority of the human genome and provides a scaffold for initiating large-scale sequencing.

Basic Strategy

We used three mapping methods to gain information about the proximity of STS loci within the human genome.

1) STS-content mapping. YAC libraries are screened by PCR to identify all clones containing a given locus (*12*). Nearby loci tend to be present in many of the same clones, allowing proximity to be inferred. STS-content linkage can be detected over distances of about 1 Mb, given the average insert size of the YAC library used here.

2) Radiation hybrid (RH) mapping. Hybrid cell lines, each containing many large chromosomal fragments produced by radiation of the relevant region on a high-resolution RH panel in parallel with screening them on the BAC library. As a simple test, we scored the STSs from a 3-Mb region on chromosome 6 on the G3 RH panel and were able to readily infer the fine-structure order of nearly all the loci with high confidence (*45*).

The use of STS-based maps as a scaffold for large-scale sequencing has several advantages: It can be initiated now with the existing STS-based map; it automatically anchors sequences in the genome; it does not require chromosome-specific libraries, which involve specialized preparation procedures and often have cryptic biases; it allows improved libraries to be substituted as they become available; and it promotes decentralization by allowing sequencing efforts to focus on regions of any given size, in contrast to entire chromosomes.

In summary, the physical map must still be refined but is already adequate to allow initiation of the international project to sequence the entire human genome—a landmark effort that will set the stage for the biology of the next century.

T. J. Hudson, S. S. Gerety, J. Ma, A. B. Castle, J. Silva, D. K. Slonim, R. Baptista, L. Kruglyak, S.-H. Xu, X. Hu, A. M. E. Colbert, C. Rosenberg, M. P. Reeve-Daly, S. Rozen, L. Hui, X. Wu, C. Vestergaard, K. M. Wilson, J. S. Bae, S. Maitra, S. Ganiatsas, C. A. Evans, M. M. DeAngelis, K. A. Ingalls, R. W. Nahf, L. T. Horton Jr., M. O. Anderson, A. J. Collymore, W. Ye, V. Kouyoumjian, I. S. Zemsteva, J. Tam, R. Devine, D. F. Courtney, M. T. Renaud, H. Nguyen, T. J. O'Connor, B. W. Birren, N. Goodman, T. L. Hawkins, and S. Foote, Whitehead–Massachusetts Institute of Technology (MIT) Center for Genome Research, Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA. L. D. Stein, Whitehead–MIT Center for Genome Research, Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA, and Department of Pathology, Brigham and Women's Hospital, Boston, MA 02115, USA. C. Fizames, S. Fauré, G. Gyapay, C. Dib, and J. Weissenbach, Généthon, CNRS URA 1922, 1 rue de l'Internationale, 91000 Evry, France. J. Morissette, Généthon, CNRS URA 1922, 1 rue de l'Internationale, 91000 Evry, France, and Centre de Recherche du Centre Hospitalier de l'Université Laval, 2705 Boulevard Laurier, Ste-Foy, Québec G1V 4G2, Canada. J. B. Orlin, Sloan School of Management, MIT, Cambridge, MA 02139, USA. D. C. Page, Whitehead–MIT Center for Genome Research, Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA; Howard Hughes Medical Institute, MIT, 9 Cambridge Center, Cambridge, MA 02139, USA; and Department of Biology, MIT, 9 Cambridge Center, Cambridge, MA 02139, USA. E. S. Lander, Whitehead–MIT Center for Genome Research, Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA, and Department of Biology, MIT, 9 Cambridge Center, Cambridge, MA 02139, USA.

*To whom correspondence should be addressed.

REFERENCES AND NOTES

1. M. Olson *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 7826 (1986); A. Coulson *et al.*, *ibid.*, p. 7821; Y. Kohara *et al.*, *Cell* **50**, 495 (1987).
2. M. Olson, L. Hood, C. Cantor, D. Botstein, *Science* **245**, 1434 (1989).
3. S. Foote, D. Vollrath, A. Hilton, D. C. Page, *ibid.* **258**, 60 (1992).
4. I. Chumakov *et al.*, *Nature* **359**, 380 (1992).
5. F. Collins and D. Galas, *Science* **262**, 43 (1993).
6. L. Selleri *et al.*, *Genomics* **14**, 536 (1992); M. Haldi *et al.*, *ibid.* **24**, 478 (1994).
7. I. Chumakov *et al.*, *Nature* **377**, S175 (1995).
8. H. Shizuya *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 8794 (1992); P. A. Ioannous *et al.*, *Nature Genet.* **6**, 84 (1994).
9. R. M. Gemmill *et al.*, *Nature* **377**, S299 (1995); N. A. Doggett *et al.*, *ibid.*, p. S335; J. E. Collins *et al.*, *ibid.*, p. S367.
10. K. Krauter *et al.*, *ibid.*, p. S321.
11. E. D. Green *et al.*, *Hum. Mol. Genet.* **3**, 489 (1994); J. Quackenbush *et al.*, *Genomics* **29**, 512 (1995); T. Alitalo *et al.*, *ibid.* **25**, 691 (1995).
12. E. D. Green and M. V. Olson, *Science* **250**, 94 (1990).
13. D. R. Cox, M. Burmeister, E. Roydon Price, S. Kim, R. M. Myers, *ibid.*, p. 245; M. A. Walter *et al.*, *Nature Genet.* **7**, 22 (1994).
14. D. Botstein *et al.*, *Am. J. Hum. Genet.* **32**, 314 (1980); J. L. Weber and P. E. May, *ibid.* **44**, 388 (1989).
15. J. Murray *et al.*, *Science* **265**, 2049 (1994).
16. J. Ott, *Analysis of Human Genetic Linkage* (Johns Hopkins Press, Baltimore, MD, 1991).
17. R. Arratia *et al.*, *Genomics* **11**, 806 (1991).
18. Loci were used only if they produced a single clear band visible by ethidium bromide staining, except for genetic markers, which were used even when they produced more than one band on an agarose gel (in view of their value in providing top-down orientation). Assays meeting this criterion are more likely, although not certain, to represent single unique loci in the genome.
19. Sequence data were analyzed with the Whitehead/MIT STS Pipeline software, which removes vector sequences, identifies duplicate sequences, and uses sequence similarity programs (FASTA and BLASTN) to eliminate known repeat sequences. Primers were chosen with PRIMER (M. J. Daly, S. Lincoln, E. S. Lander, Whitehead Institute) having the desired T_m (temperature at which 50% of double-stranded DNA is denatured) for primers set at 58°C.
20. M. Boguski and G. Schuler, *Nature Genet.* **10**, 369 (1995).
21. Nonredundant ESTs were part of the UniEST set prepared by M. Boguski and G. Schuler at the National Center for Biotechnology Information, derived from the dbEST database. These ESTs were taken from the Washington University–Merck EST project and the GenExpress project; R. Houlgatte *et al.*, *Genome Res.* **5**, 272 (1995).
22. R. Berry *et al.*, *Nature Genet.* **10**, 415 (1995); M. D. Adams *et al.*, *Nature* **377**, S3 (1995).
23. To select primers from ESTs, we modified the STS pipeline (*19*) to select shorter PCR products of 100 to 150 bp near the end of the 3′-UTRs [but 20 bp away from the polyadenylate [poly(A)] tract] in a region of high sequence quality.
24. J. Weissenbach *et al.*, *Nature* **359**, 794 (1992); G. Gyapay *et al.*, *Nature Genet.* **7**, 247 (1994); C. Dib *et al.*, *Nature*, in press.
25. C. J. Bell *et al.*, *Hum. Mol. Genet.* **4**, 59 (1995).
26. STSs were kindly shared by D. Cox and R. Myers, Stanford University, Stanford, CA, and J. Gastier, Harvard University, Cambridge, MA.
27. Some incomplete addresses could be resolved by a simple band-matching test with complete addresses by using CEPH fingerprint data, as described (*25*). Others could be resolved by virtue of comparison with complete addresses for nearby STSs.
28. Dot-blotted PCR products were initially detected by using ECL kits (Amersham), as described (*25*). We later switched to overnight hybridization with a biotinylated oligonucleotide probe to an internal sequence, followed by chemiluminescent detection with a peroxidase catalyzed luminol reaction, as described [R. P. M. Gijlswijk *et al.*, *Mol. Cell. Probes* **6**, 223 (1992)]. STSs known to contain an internal repeat sequence such as CA or AGAT were probed with an oligonucleotide for the repeat. Other STSs were probed with a specific internal oligonucleotide, having a T_m of 58°C. Computer images of each hybridization were obtained with a CCD camera. VIEW software (C. Rosenberg; Whitehead Institute) was used to locate and determine the intensity of positive dots. A small proportion of STSs were screened by standard agarose gel stained with ethidium bromide.
29. It is not possible to draw conclusions about library coverage from the overall number of STSs with no definite addresses, because many of these represented weak PCR assays that sometimes worked on human control DNAs but failed on YAC pools.
30. The probability that a unique sequence would occur more than 15 times in a random library with 8.4-fold coverage is about 1%. Some of these STSs may thus be unique loci, but they were excluded to guard against repeats.
31. G. Gyapay *et al.*, *Hum. Mol. Genet.*, in press.
32. Dot-blotted PCR products were detected as for the YAC screening, except that STSs containing CA repeats were screened with oligonucleotides containing unique internal sequence rather than $(CA)_n$, because the latter produced high background.
33. H. L. Drwinga *et al.*, *Genomics* **16**, 311 (1993); T. J. Hudson *et al.*, *ibid.* **13**, 622 (1992). In 300 cases, ambiguous or conflicting assignments were resolved by using the NIGMS Human/Rodent Somatic Cell Hybrid Panel #2, described in B. L. Dubois and S. Naylor *et al.*, *ibid.* **16**, 315 (1993).
34. In 151 cases, STSs were chromosomally assigned by virtue of having at least three single links to other markers on a chromosome and no links to any loci on any other chromosome.
35. About 200 such conflicts were resolved. Half were resolved by repeating the typing of the somatic cell hybrid panel. For the other half, an STS was demonstrated to amplify products from more than one chromosome. Such STSs were discarded.
36. RHMAPPER (L. Kruglyak, D. K. Slonim, L. D. Stein, E. S. Lander, unpublished data) uses a hidden Markov model to account for breaks in diploid DNA and for false positives and negatives, as in E. S. Lander and P. Green, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 2363 (1987); and E. S. Lander and S. Lincoln, *Genomics* **14**, 604 (1992). Framework maps were initiated and extended by a greedy algorithm and then subjected to local permutation tests, thereby allowing for efficient exploration of a vast space of possible orders.
37. In most cases, frameworks for the two chromosomal arms were constructed separately and then oriented and joined by using information from the genetic map. There is significant pairwise RH linkage (at lod > 5.0) between framework markers on opposite sides of the centromere on nine of the final framework maps, but not on the other 14.
38. M. R. James *et al.*, *Nature Genet.* **8**, 70 (1994).
39. For each pair of consecutive STSs, a positive score a_{++} was added for each YAC containing both, a negative score a_{+-} for each YAC containing one but not the other, a positive score b_{++} for each hybrid containing both, and a negative score b_{+-} for each hybrid containing one but not the other. Letting x_{11}, x_{10}, and x_{00} denote the probabilities that two STSs separated by about 500 kb would be observed to be both present, both absent, or one present and one absent in a randomly chosen YAC, we set $a_{++} = \log(x_{11}/x_{00})$ and $a_{+-} = \log(x_{10}/x_{00})$. The constants b_{++} and b_{+-} were defined similarly with respect to individual RH cell lines. The various probabilities were calculated on the basis of the distribution of fragment sizes and the inferred false positive and false negative rates. These four parameters, a_{++}, a_{+-}, b_{++}, and b_{+-}, were thus determined directly from the data and were not optimized. The two weighting parameters for conflicts with the genetic and framework RH maps were chosen by optimization in test cases.
40. S. Saccone *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 4913 (1992).
41. Three markers mapped into large centromeric intervals on the correct chromosome; they had high lod scores but were about 30 cR away from the closest marker. All were confirmed by double-linkage with YACs. For three other markers, chromosomal assignment could not be obtained from polychromosomal hybrid panels because of rodent background.
42. For one of these four loci, there was a (presumably chimeric) single YAC link to a marker on the same chromosome but located 70 cR from the correct location.
43. Three of the loci belonged to doubly linked contigs that were anchored by virtue of a CHLC genetic marker.
44. If gene promoters on chromosome X have the same average expression level as on autosomes, then the fact that only one X chromosome is active (due to hemizygosity in males and X inactivation in females) would cause transcripts from X-linked genes to be half as abundant. Because half of the cDNAs came from nonnormalized libraries and half from normalized libraries, the occurrence of ESTs in the relatively small set examined will partly reflect abundance. This issue will recede when enough ESTs have been isolated to overcome issues related to message levels. Underrepresentation of chromosome X could also conceivably represent some other systematic bias of which we are not aware.
45. T. J. Hudson *et al.*, data not shown.
46. J. S. Heslop-Harrison *et al.*, *Hum. Genet.* **84**, 27 (1989).

47. We thank A. Kaufman, O. Merport, and J. Spencer for technical assistance; L. Bennett for computer system administration; G. Rogers and M. Foley for assistance with media preparation and glass washing; S. Gordon, A. Christopher, P. Mansfield, and others at Intelligent Automation Systems for assistance in design, construction, and maintenance of automation equipment; D. Cox, R. Myers, J. Sikela, M. Adams, J. Murray, and K. Buetow for sharing data including sequences and markers; M. Boguski and G. Schuler for assistance in analysis of EST sequences; and D. Cohen and I. Chumakov for sharing the CEPH YAC library in 1992 and for public distribution of their STS-content, Alu-PCR, and fingerprint data. Supported by NIH award HG00098 to E.S.L. and by the Whitehead Institute for Biomedical Research. T.J.H. was a recipient of a Clinician Scientist Award from the Medical Research Council of Canada. L.K. is a recipient of a Special Emphasis Research Career Award (HG00017) from the National Center for Human Genome Research.

17 October 1995; accepted 8 November 1995

genetic maps. By measuring the frequency of such occurrences as a function of the distance between the loci in the STS map, we estimated that about 0.5% of the loci may be significantly misplaced in the maps.

In summary, the local order in the map must be regarded as uncertain. There will surely be many errors requiring attention and correction. The effective resolution of the map is certainly lower than the average spacing between loci and may be about 1 Mb. To improve the local accuracy of the maps, investigators interested in particular regions would be well advised to retest the STSs against an RH panel with higher resolution [such as the G3 panel developed by D. Cox and R. Myers, in which the fragments are about 1/10 those in the GeneBridge 4 panel] and against regional YAC panels, as described above. In this fashion, the map provides the tools for its own refinement.

Finally, we note that direct comparison of our STS-based map with the recently reported YAC-based map (*7*) is difficult, because of the very different natures of the maps. For example, it is not meaningful to compare the STS orders in the two maps: The YAC-based map almost exclusively involved genetic markers and provided no independent information about locus order, but instead simply adopted the genetic order. It is also problematic to compare the specific YACs identified, because the YAC-based map involved only partial screening of most STSs and did not fully distinguish clones actually belonging to paths through a region from those representing false positive hybridization. At the grossest level, it is possible to compare the coverage of the maps: The current map appears to cover about 95% of the genome (the precise amount depends on the type of mapping information used), whereas the other map was reported to cover about 75%. More detailed comparison would be worthwhile, as it would likely lead to improvements in both maps.

Table 4. STS-content mapping of YACs.

Chr.	STS-content mapped loci	STS spacing (kb/STS)	No. of YACs*	Contigs			YAC hits per STS‖	STSs per contig¶
				Before gap closure†	After gap closure‡	Avg. size (Mb)§		
1	1,048	237	1,393	49	34	7.3	6.7	30.8
2	933	258	1,469	56	20	12.0	7.3	46.7
3	791	255	1,192	46	30	6.7	7.5	26.4
4	718	267	1,272	42	11	17.4	8.2	65.3
5	651	281	1,163	35	19	9.6	7.9	34.3
6	641	269	1,091	40	24	7.2	7.8	26.7
7	559	288	942	39	13	12.4	7.8	43.0
8	552	265	945	23	11	13.3	8.0	50.2
9	394	347	675	28	12	11.4	8.0	32.8
10	519	262	750	36	26	5.2	6.8	20.0
11	490	277	696	23	14	9.7	7.2	35.0
12	509	265	842	29	16	8.4	7.4	31.8
13	300	308	556	12	5	18.5	8.0	60.0
14	352	249	593	9	6	14.6	8.0	58.7
15	301	279	439	16	10	8.4	7.0	30.1
16	255	362	308	26	16	5.8	6.0	15.9
17	267	325	330	27	17	5.1	5.8	15.7
18	315	254	478	16	8	10.0	7.6	39.4
19	79	800	76	17	15	4.2	4.7	5.3
20	266	255	328	15	10	6.8	6.6	26.6
21	113	325	182	4	2	18.4	8.0	56.5
22	182	223	134	11	11	3.7	5.4	16.5
X	408	379	406	53	46	3.4	4.7	8.9
Y#	207	128	234	1	1	26.4	4.1	207.0
Total	10,850	276	16,494	653	377	8.0	7.3	28.8

*Number of YACs hit by at least two STSs on the chromosome. YACs hit by only one STSs are omitted. †Number of contigs based only on STS-content data. ‡Number of contigs after gap closure by Alu-PCR hybridization and fingerprint data. §Average contig size estimated by assuming that 94% of the chromosome length is covered and dividing by the number of contigs. ‖Includes all YAC hits in the library screen (not limited to YACs having multiple STS hits on the chromosome) and thus reflects coverage by the library. ¶Reflects contigs after gap closure. #From our previously reported work (*3*).

Distribution of Genes

The map also sheds light on the organization of the human genome. By comparing the chromosomal distribution of the expressed sequences to the chromosomal distribution of the random single-copy sequences (both determined in the same manner), one can draw inferences about the density of genes on different chromosomes. We compared the observed number on each chromosome to the expected number, assuming that expressed sequences have the same distribution as random STSs (Table 2). Chromosomes 1, 11, 17, 19, and 22 showed a statistically significant excess of expressed sequences ($P = 0.001$ after correction for multiple testing). Chromosomes 17, 19, and 22, which showed the greatest excess, have been previously suggested to have a high density of genes on the basis of indirect evidence (*40*). Chromosome X was the only chromosome to show a statistically significant deficit of expressed sequences—only about half as many as expected. This would suggest that there is a low gene density on this sex chromosome, although alternative explanations are possible (*44*). We also analyzed the raw data from two recent papers reporting chromosomal assignment of expressed sequence tags (ESTs) (*21*, *22*) and found a similar deficit of X-linked loci.

A Scaffold for Sequencing the Genome

As genetic and physical maps approach their intended goals, attention is turning to the challenge of sequencing the entire human genome. A key issue is how to obtain the required sequence-ready clones. STS-based maps provide a general solution by making it possible to generate extensive physical coverage of a region by screening a single high-quality human genomic library.

One could, for example, proceed as follows: Screen the STSs in a region against a bacterial artificial chromosome (BAC) library having 150-kb inserts and 10-fold coverage, use a simple fingerprinting scheme to detect overlaps among adjacent clones, and select a minimally overlapping set for sequencing. Given a physical map containing 30,000 ordered STSs, one would screen about 100 STSs and fingerprint about 520 BACs to cover a 10-Mb region; this task could be readily accomplished in a few days with modest automation and would not contribute significantly to the cost of sequencing. The resulting BACs would be expected to cover about 95% of the region in ordered sequence contigs (*17*). The region could then be closed by straightforward walking—that is, serially screening the BAC library with STSs derived from sequences at the ends of each contig.

The current map falls short in terms of marker density and local order, but neither shortcoming poses a serious obstacle for initiating large-scale sequencing now. With the 15,000 STSs currently available, one should cover about 75% by direct screening, 90% by one round of walking, and more than 95% with two rounds (*17*). The desired map with 30,000 STSs will likely be available within the next 2 years, through current projects underway at several centers including our own. Uncertainties about locus order can be overcome simply by scoring the STSs from

tion breakage, are screened by PCR to identify those hybrids that have retained a given locus (*13*). Nearby loci tend to show similar retention patterns, allowing proximity to be inferred. RH linkage can be detected for distances of about 10 Mb, given the average fragment size of the RH panel used here.

3) Genetic mapping. A locus that is polymorphic in the human population can be screened by PCR to determine its inheritance patterns in families (*14*, *15*). Nearby loci tend to show similar inheritance patterns, allowing proximity to be inferred. Genetic linkage can be reliably detected over distances of about 30 Mb, given the recombination rate of human chromosomes (*16*).

These three methods were used to produce independent maps and then combined to produce an integrated map. Because RH mapping and genetic mapping can detect linkage over large regions (0.3 to 1% of the genome), comprehensive RH and genetic maps spanning all chromosomes can be assembled with a few thousand loci. The order of loci can be inferred from the extent of correlation in the retention or inheritance patterns, although estimates of fine-structure order are not precise. These methods can thus provide "top-down" information about global position in the genome.

In contrast, STS-content mapping provides "bottom-up" information. It reveals tight physical linkage among loci but is useful only over short distances and does not provide extensive long-range connectivity across chromosomes (*17*). Two STSs are said to be singly linked if they share at least one YAC in common and doubly linked if they share at least two YACs (*17*). Single linkage is an inadequate criterion for declaring adjacency of STSs, because of the high rate of YAC chimerism (about 50%) and the possibility of laboratory error. Double linkage, however, turns out to be a reliable indication, because two genomic regions are unlikely to be juxtaposed in multiple independent YACs. Accordingly, a three-step procedure was used. (i) STSs were assembled into doubly linked contigs (groups of STSs connected by double linkage). (ii) The doubly linked contigs were localized within the genome on the basis of RH and genetic map information about loci in the contig. (iii) Single linkage was then used to join contigs localized to the same small genomic region. The overall strategy is illustrated in Fig. 1. We now describe the data generation, map construction, and map analysis in greater detail.

Data Generation

Marker development. Over the course of the project, we tested 20,795 distinct PCR assays. These candidate STSs were initially characterized to see whether they were likely to detect a unique genomic locus (*18*) and whether they consistently yielded correct results on control samples under uniform production conditions. A total of 16,239 STSs met these stringent criteria and were used for mapping. The STSs fell into one of the following four categories.

1) Random loci. We generated 3027 working STSs by sequencing random human genomic clones and discarding those that appeared to contain repetitive sequences (*19*).

2) Expressed sequences. We developed 921 STSs from complete complementary DNA (cDNA) sequences in GenBank, taken from the Unigene collection (*20*). Another 3349 STSs were developed from expressed sequence tags (ESTs). Of these, 71% came from the dbEST database (*21*), 13% from the laboratory of Jim Sikela, 9% from the Institute for Genomic Research, and 7% from various other sources (*22*). We found that the success rate for STSs derived from the last 200 base pairs (bp) of 3′-untranslated regions (UTRs) of cDNAs was similar to that for STSs derived from random genomic DNA, consistent with the idea that introns rarely occur near the ends of 3′-UTRs (*23*). The results indicate that PCR assays can be readily derived for the vast majority of cDNAs.

3) Genetic markers. A total of 6986 loci were used, consisting of 5264 polymorphic loci developed at Généthon (primarily CA repeats) (*24*) and 1722 loci developed by the Cooperative Human Linkage Center (CHLC) (primarily tri- and tetranucleotide repeats) (*15*).

4) Other loci. A total of 1956 STSs were developed from various sources. These included 1091 CA-repeat loci developed at Généthon that were not sufficiently polymorphic to be useful for genetic mapping, as well as 865 loci from chromosome 22–specific and chromosome Y–specific libraries and gifts from other laboratories (*3*, *25*, *26*).

A total of 15,086 STSs appear in the final maps. The number of markers of each type appearing in the final STS-content, RH, and genetic maps is shown in Table 1.

STS-content mapping: Methodology. STSs were screened against 25,344 clones from plates 709 to 972 of the CEPH mega-YAC library (*7*), estimated to have an average insert size of 1001 kb and to provide roughly 8.4-fold coverage of the genome. To facilitate screening, we used a hierarchical pooling system. The library was divided into 33 "blocks," each corresponding to eight microtiter plates or roughly 0.25 genome equivalent. For each block, we prepared one "superpool" containing DNA from all the clones and 28 "subpools" by using a three-dimensional pooling system based on the row, plate, and column address of each clone. Specifically, there were 8, 8, and 12 subpools consisting of YACs in the same plate, row, and column, respectively. There was thus a total of 957 super- and subpools.

For blocks with a single positive YAC, the row, column, and plate subpools should specify the precise address of the YAC ("definite addresses"). If a block contained two or more positive YACs or if one of the three subpool dimensions did not yield a positive, partial information was obtained ("incomplete addresses") (*27*). Such incomplete addresses could consist of up to 12 possible addresses (for example, in the case that a column address was missing). Incomplete addresses were not used in initial map assembly but were used at the final stages to detect connections between nearby loci. Definite addresses composed 88% of the total hits.

Half of the markers were screened by a two-level procedure, in which we first identified the positive superpools and then tested only the corresponding subpools. The other half were screened by a one-level procedure, bypassing the superpools and directly screening all subpools. Although the latter procedure involves more reactions,

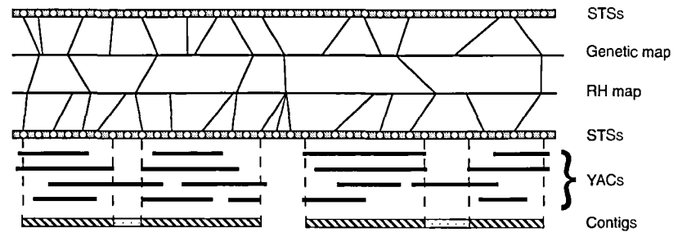


Fig. 1. Schematic diagram of the STS-based map. STSs are shown as circles on the first and fourth line. Loci that are genetically mapped or RH mapped are connected to the appropriate position on these maps, with connections between these maps in the cases of loci present in both maps. YACs containing STSs are shown below. The STSs fall into two singly linked contigs (stippled rectangles) and four doubly linked contigs (striped rectangles). Single linkage is not reliable for connecting arbitrary doubly linked contigs, but it is reliable in the case of anchored doubly linked contigs known to be adjacent on the genetic or RH map, as in the figure.

each locus is treated in an identical manner, which offers advantages for automation. In both procedures, we identified the positive pools by spotting the PCR reactions on membranes, hybridizing them to a chemiluminescent probe specific for each STS, capturing the resulting signal directly by a charge-coupled device (CCD) camera, and up-loading the results into our database (28); this approach proved to be much more efficient than the traditional detection procedure of gel electrophoresis.

Because the project involved processing more than 15 million reactions, laboratory automation was essential. We collaborated with an engineering firm, Intelligent Automation Systems, Incorporated, (IAS) of Cambridge, Massachusetts, to design and build various special-purpose machines to accelerate STS-based mapping.

The two-level screening procedure was carried out with a large robotic liquid-pipetting workstation and two custom-designed thermocyclers (Fig. 2). A laboratory information management system used the superpool results to automatically program the robotic workstation to set up the appropriate subpool screens. The system has a maximal throughput of 6144 PCR reactions per run.

The one-level screening procedure was made feasible by the development of a massively parallel factory-style automation system nicknamed the Genomatron (Fig. 2). The Genomatron was also developed in collaboration with IAS and consists of three stations. The first station assembles PCR reactions in custom-fabricated 1536-well microtiter "cards" and seals the wells by welding a thin plastic film across the card. The second station thermocycles the reactions by transporting the cards over three chambers that force temperature-controlled water to flow uniformly between the cards. The third station transfers the reactions from one microtiter card onto a hybridization membrane affixed to the bottom of a second microtiter card by piercing the first card with a bed of 1536 hypodermic needles and sucking the reactions downward with a vacuum plenum. These "filter cards" were then manually hybridized with a chemiluminescent probe and read by the CCD camera. The stations were computer controlled, and the microtiter cards were assigned a bar code to facilitate sample tracking. Each station was designed to process 96 microtiter cards, providing a throughput of nearly 150,000 reactions per run.

STS-content mapping: Results. A total of 11,750 STSs yielded from 1 to 15 definite YAC addresses and were considered successfully screened (29); typical loci yielded approximately one additional incomplete address. STSs having more than 15 definite hits were excluded as likely to detect multiple genomic loci (30).

The successfully screened loci produced an average of 6.4 YACs per STS, considering only definite addresses. A total of 18,879 YACs were hit by at least one STS. For these YACs, the average hit rate was 3.8 STSs per YAC. The average size of the YACs hit by the STSs was about 1.1 Mb (~10% greater than for the library as a

Table 1. Overview of mapped STSs.

STSs on final map	No. of loci
STS-content map	10,850
RH map	6,193
Genetic map	5,264
Intersection of	
STS-content and RH maps	4,036
STS-content and genetic maps	3,106
RH and genetic maps	887
All three maps	807
Total loci	15,086

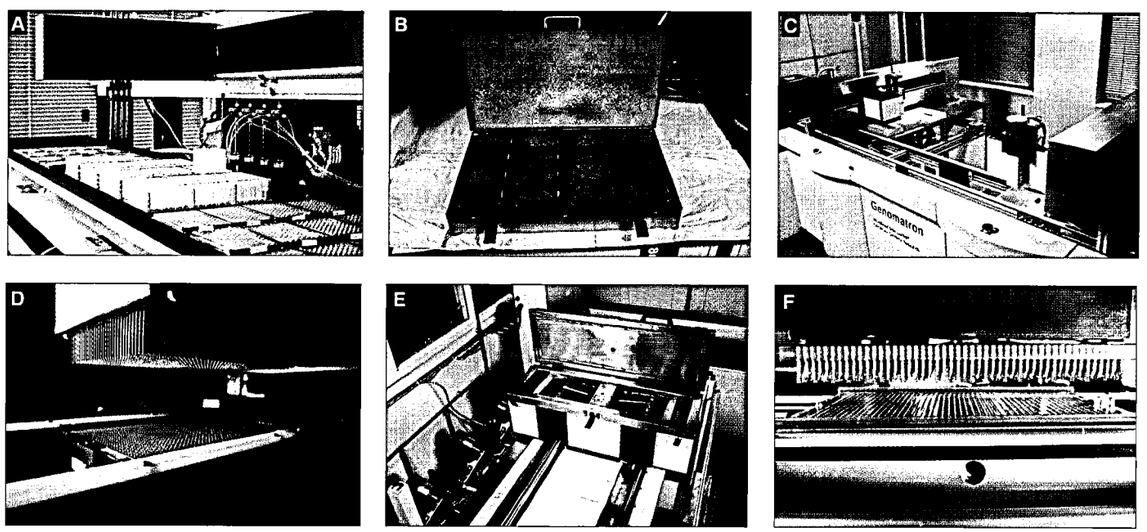


Fig. 2. The first automated system developed for the project was (**A**) a robotic station to set up PCR reactions and (**B**) custom-built "waffle iron" thermocyclers accomodating 16 192-well microtiter plates; the system has a capacity of 6144 PCR reactions per run. The second automated system was the Genomatron, which consists of three robotic stations. PCR reactions are set up in 1536-well microtiter cards (consisting of 15 cm by 24 cm injection molded plastic cards with 1536 holes, to the bottom of which a plastic film is heat-sealed to create wells). The first station (**C**) assembles the PCR reactions. Each run can process up to 96 cards per run, providing a capacity of nearly 150,000 wells. Cards are dispensed by a coining mechanism and travel along a conveyor belt to substations containing a bar code reader; a 1536-head pipettor (**D**) that dispenses template DNAs to be screened; a 48-head pipettor that dispenses PCR primer mixes, including polymerase; a plate sealer that heat-seals a plastic film on the top of the card to create separate reaction chambers; and a refrigerated storage station. The second station is a thermocycler (**E**) that uses three large waterbaths. Up to 96 sealed cards containing PCR reactions are placed in a chamber that travels over the water baths, which pump water at the appropriate denaturing, annealing, and extension temperature. The third station is a parallel "spotting" device that transfers PCR reactions from a card to a nylon filter affixed to the bottom of a second card. After the two cards are aligned, a bed of 1536 hypodermic needles (**F**) pierces a sealed card containing the reactions while a vacuum manifold draws the reaction mixtures down onto the membrane on the second. The filter cards are manually hybridized and subjected to a chemiluminescent detection protocol. Light signals are recorded with a cooled CCD camera.

could thus not be localized on the STS-content map (42). These four STSs appear to be in regions of low YAC coverage, inasmuch as they hit one, one, one, and two YACs, respectively. The remaining 94 STSs could all be localized on the STS content map [with 91 being doubly linked and three being singly linked to existing contigs anchored in the correct chromosomal region in the top-down map (43)]. The 100 loci detected an average of 6.5 YACs.

The map covers the vast majority of the human genome. We estimate that 99% of random STSs can be readily positioned on the RH map, and 94% can be positioned on the STS-content map relative to YAC clones.

The physical map thus fills a major need in human genetics, providing a general method by which an investigator can map a locus in the human genome by screening readily available RH or YAC pools and comparing the resulting pattern with the map. To make this information easily accessible to the scientific community, we have written a "map server." The server reports the likely position of an STS given information about its YACs, RH pattern, and chromosomal assignment. The server is freely available via our World Wide Web site.

Fig. 3 (previous pages). Integrated map of human chromosome 14q. Long vertical lines represent the STS-content map (first and fourth lines, in black), genetic map (second line, in blue), and RH map (third line, in orange), in the same fashion as the diagram in Fig. 1. All three maps are drawn to equal length. The four columns of STS names correspond to the four lines. For the STS-content map, intermarker distance is not known and loci are displayed as equally spaced. For genetic and RH maps, loci are indicated at positions spaced proportionally along the map according to the respective metrics. Loci in common between two maps are connected by black lines. Loci belonging to the RH framework map (in which the relative ordering is supported by lod > 2.5) are shown in bold type and with thicker connecting lines. Loci derived from expressed sequences are shown in purple. YACs are displayed as black rectangles, to the right of the STSs that were found to be contained in the clone. YAC names are shown to the top right. Unfilled portions of YACs represent assays that were negative. Thin red lines in some YACs represent incomplete addresses that were resolved by virtue of overlap with addresses from a nearby locus. Gaps between contigs are shown as horizontal lines separating groups of YACs. Gaps that were likely to be undetected overlaps based on Alu-PCR hybridization or fingerprint information (see text) are shown in yellow; gaps for which there is no evidence of overlap are shown in gray. Vertical dotted gray bars indicate STSs with identical data for given mapping method. YACs detected by only a single STS were omitted from this display. These YAC addresses can be obtained from the Whitehead Institute–MIT Center for Genome Research World Wide Web server at URL http://www-genome.wi.mit.edu/. Figure represents slightly earlier version of the map, from the 14,000-marker stage.

Accuracy

Although the long-range order of the map is reliable because of top-down anchoring, precise local orders must be regarded as only approximate. Local ordering depends on the position of loci with respect to individual breakpoints, that is, the ends of YAC or RH fragments. The accuracy of such inference is limited by the presence of false positives and false negatives in our data, as well as by the presence of internal deletions in YACs. Whereas the long-range order tends to be over-determined in genomic maps, several alternative local orders may be reasonably compatible with the data. The "best" order may change with the alteration of a few data points.

We used three approaches to evaluate the accuracy of the data and the map.

1) Rescreening of loci on chromosome 14. Chromosome 14 was divided into 16 regions and regional YAC panels were defined, consisting of all clones hit by one or more loci in the region. For each regional YAC panel, individual DNAs were prepared from each clone. We tested 112 STSs against their corresponding panels to directly compare the results from high-throughput screening of pools with the screening of individual clones. We found a false positive rate of 5.5% and a false negative rate of 19.5% in our high-throughput screening data, both of which were consistent with our earlier indirect estimates. We constructed a new STS map of the chromosome using these more complete data; the new map showed about six instances of local reorderings involving two to five loci.

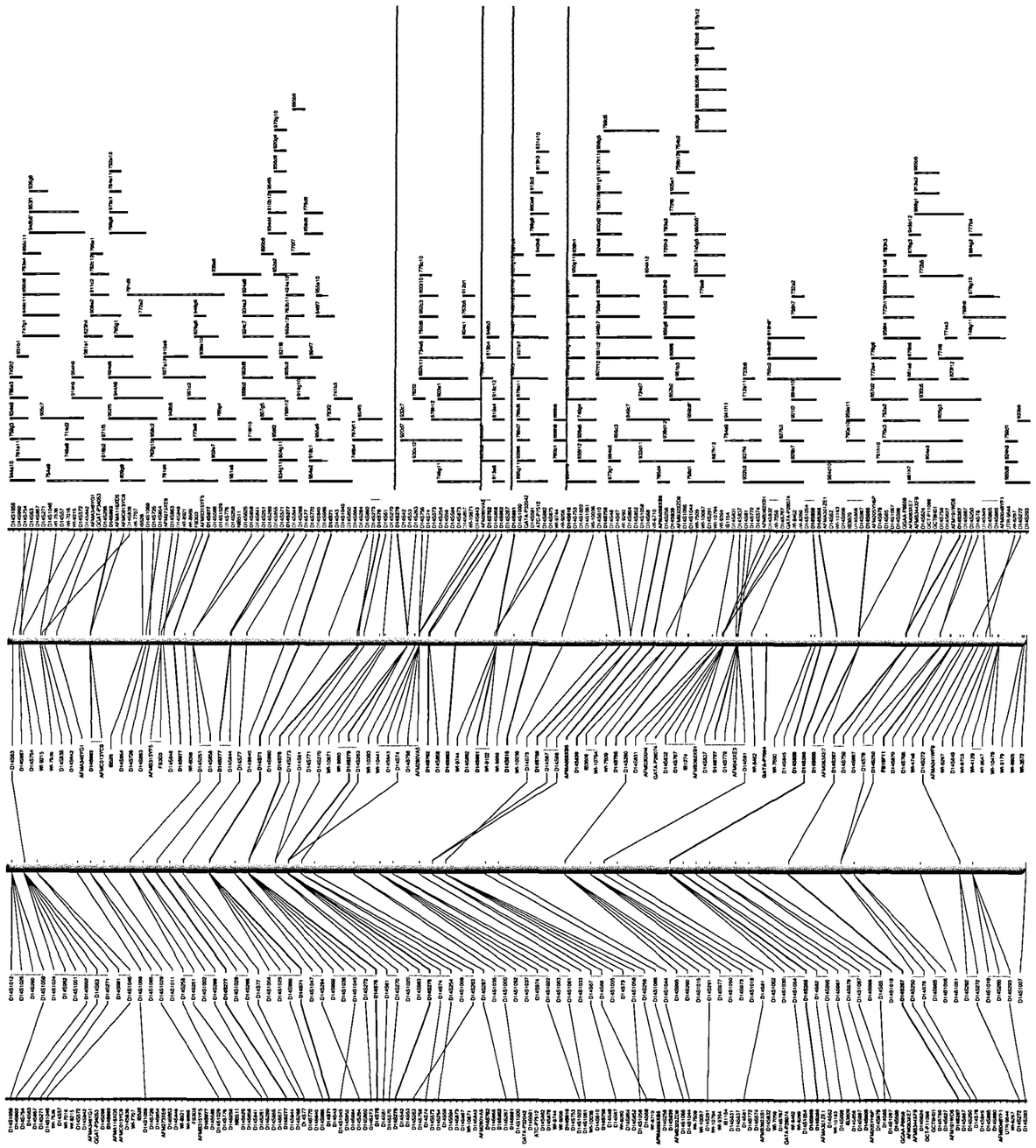
2) Comparison with an independently constructed map of chromosome 12. We compared our map with a recently reported map of this chromosome (10) containing enough loci in common to provide a meaningful test. Of 171 loci in common, there were about a dozen instances of small local inversions involving two to three adjacent markers. A substantial difference in position was seen for only a single marker, AFM263WH1. Our map shows tight STS-content linkage of this locus to genetic markers at 91 cM on the Généthon map, whereas the other map places it near genetic markers at 105 cM. In fact, the position on our map agrees well with the reported genetic map location for this locus (at 93 cM), so we believe it to be correct. In any case, the two maps showed relatively few conflicts.

3) Internal consistency checking. We looked for instances in which pairs of loci occurred in an order on the final STS map that was strongly disfavored by the RH or

Table 3. Genetic and RH maps. Dashes indicate not applicable.

Chr.	Physical length (Mb)*	Genetic map			RH map			
		No. of loci	Length (cM)	Genetic vs. physical (cM/Mb)	Framework map No. of loci	Total RH map	RH length (cR)†	RH vs. Physical (cR/Mb)
1	248	461	293	1.2	107	559	743	3.0
2	240	452	277	1.2	119	532	977	4.1
3	202	353	233	1.2	95	475	801	4.0
4	191	280	212	1.1	80	370	552	2.9
5	183	312	198	1.1	60	339	508	2.8
6	173	311	201	1.2	97	374	739	4.3
7	161	272	184	1.1	63	360	591	3.7
8	146	249	166	1.1	77	264	711	4.9
9	137	189	166	1.2	75	260	440	3.2
10	136	281	182	1.3	71	297	599	4.4
11	136	273	156	1.1	66	302	515	3.8
12	135	249	169	1.3	58	294	565	4.2
13	92	164	117	1.3	46	169	309	3.3
14	88	162	129	1.5	38	210	319	3.6
15	84	145	110	1.3	41	185	342	4.1
16	92	180	131	1.4	33	186	235	2.5
17	87	186	129	1.5	34	156	347	4.0
18	80	136	124	1.5	52	175	450	5.6
19	63	121	110	1.7	21	107	221	3.5
20	68	144	96	1.4	30	157	265	3.9
21	37	61	60	1.6	15	61	151	4.1
22	41	67	58	1.4	15	89	141	3.5
X	155	216	198	1.3	46	272	521	3.4
Y	26	–	–	–	–	–	–	–
Total	3,000	5,264	3,699	1.2	1,339	6,193	11,042	3.7

*Physical lengths were calculated on the basis of a genome of 3000 Mb, with proportional lengths of chromosomes as reported (46). †Total length of the RH framework map, omitting the large interval at the centromere.



whole), corresponding to 6.9-fold coverage of the genome. Some 78% of the STSs showed double-linkage to at least one other STS.

The false positive rate was investigated by regrowing and testing individual YACs. Several thousand addresses were tested, and 95% could be directly confirmed, with the remainder constituting actual false positives, deletions during regrowth, or technical failures during retesting. The false positive rate is thus at most 5% of definite addresses, and the chance of any particular YAC occurring as a false positive in a given screen is about 1.5×10^{-5}. False positive addresses thus will rarely create false links among STSs known to lie in the same genomic region. The false negative rate cannot be computed directly, but the fact that an average of 6.4 hits was seen in 8.4 genome equivalents suggests a rate of about 20%. False negatives pose a less serious problem than false positives (which join incorrect genomic regions), but they can lead to incorrect local ordering of STSs. The false positive and negative rates were reinvestigated once the maps were constructed, as discussed below.

Radiation hybrid mapping. STSs were screened against the GeneBridge 4 whole-genome radiation hybrid panel, consisting of 91 human-on-hamster somatic hybrid cell lines. Each line retains about one-third of the human genome in fragments of about 10 Mb in size. The GeneBridge 4 panel (Research Genetics, Huntsville, Alabama) was developed in the laboratory of P. Goodfellow and distributed to the scientific community as a resource for the mapping of expressed sequences. As part of a separate project, the panel has been characterized for more than 500 well-spaced genetic markers to confirm that substantial linkage can be obtained across the genome (*31*).

RH mapping was performed with essentially the same protocol as for the YAC screening: PCR reactions were set up either by the Genomatron (with each 1536-well microtiter card containing reactions for eight loci) or by the robotic workstation (by using 192-well microtiter plates), spotted on membranes, hybridized to a chemiluminescent probe, and detected by a CCD camera (*32*).

Scoring results from RH panels requires considerable caution. Human chromosomal fragments are present at various molarities among the hybrid cell lines; thus, the ability to detect their presence may vary with the sensitivity of each PCR assay. As a result, STSs that are immediately adjacent in the genome could conceivably give somewhat different retention patterns, which would limit the ability to determine fine-structure order. To minimize discrepancies due to assays near the limit of detection, we performed all assays in duplicate. Hybrids were scored if the two duplicates gave concordant positive or negative results but were recorded as "discrepant" if the duplicates were discordant. The mean discrepancy rate was 1.2%; loci with a discrepancy rate exceeding 4.5% were eliminated as unreliable.

A total of 6469 STSs were successfully screened on the GeneBridge 4 RH panel. The overall retention rate of the panel was 32% (or about 18% per haploid genome from the diploid donor cell).

Genetic mapping. Genetic linkage information was used from the recent Généthon linkage map of the human genome, containing 5264 polymorphic markers (*24*). Genetic linkage information was not incorporated for the 1722 CHLC genetic markers studied.

Chromosomal assignment. Before undertaking map construction, we attempted to assign all loci to specific chromosomes by multiple, independent methods. Most STSs were screened against the NIGMS 1 polychromosomal hybrid panel (*33*), resulting in unambiguous chromosomal assignment in about 75% of the cases (with the remainder having high background from the host genome or poor signal). STSs defining genetic markers typically had chromosomal assignments on the basis of linkage analysis. STSs were also assigned to chromosomes if they were tightly linked by RH screening or doubly linked by YAC screening to chromosomally assigned loci (*34*).

Some 96% of the loci could be chromosomally assigned, with the majority of these being assigned by at least two independent methods. Conflicting assignments were noted in a small proportion of cases (2%); these were subjected to intense scrutiny and resolved in the majority of cases (*35*). Loci that could not be reliably assigned to a chromosome were omitted from map construction, to avoid problems associated with chimeric linkages.

Personnel. The project was carried out during a period of 2.5 years by a team at Whitehead having an average of 16 people involved in mapping, three people involved in sequencing, and five people involved in data management and computational analysis.

Map Construction

Top-down maps. The genetic and RH maps are top-down maps, which provide a global framework and offer many tests of internal consistency. The first step in constructing an RH linkage map was to make high-quality "framework" maps across each chromosome. For this purpose, we included only loci with independent chromosomal assignments and with retention rates in the range of 10 to 60% (unusually high or low retention rates can produce spurious linkage). We wrote a computer package, RHMAPPER, that implements RH mapping for hybrids constructed from diploid sources and incorporates probabilistic error detection and error correction (*36*). Using this program, we generated a framework map—that is, an ordered set of markers such that each consecutive pair was linked with a lod score > 10 (lod score is the logarithm of the likelihood ratio for linkage), and the order was better than all local alternatives by a lod score > 2.5. The framework map included 1339 loci and provided complete connectivity across each chromosome arm with no gaps over 30 centiRays (cR) (cR is a measure of distance that is analogous to centimorgans but depends on the radiation dose). There were, however, large intervals across most centromeres (*37*), a phenomenon that has been previously seen for chromosome 11 (*38*). The total length of the map is 11,042 cR (omitting the centromeric intervals), corresponding to a fairly uniform average of about 300 kb/cR across most chromosomes.

We then localized the remaining markers relative to the framework map. These loci could not be uniquely ordered, either because of close proximity to a framework marker (loci with identical retention patterns cannot be ordered with respect to one another) or because of potentially erroneous typing results (that cause apparent "double-breaks" regardless of the interval in which the marker is placed). RHMAPPER allowed for the possibility of false positive and false negative typings and flagged probable errors (about two-thirds of which were found to be real errors in cases that were subsequently retested). The nonframework markers were estimated by the computer analysis to have an average residual error rate of just less than 1%. To reflect the uncertainty in order, each locus was assigned to the collection of intervals for which the lod score was within three of the optimal position. Loci were not included if they mapped more than 15 cR from a framework marker (that is, past the end of the map or in a large centromeric gap), because such positions could result from a high proportion of errors. In all, 6193 of 6469 loci tested were placed in the RH map.

Together, the two top-down maps contained a total of 10,572 loci. The reliability of the maps can be assessed by studying the loci in common. For loci present in both the genetic map and the framework RH map, there were only four conflicts in order; the loci involved were separated by 1 centimorgan (cM) in three cases and 3 cM in one case. The close agreement between the maps suggests that they correctly reflect the global order of loci in the genome.

Bottom-up map. Using the STS-content data, we assembled doubly linked contigs and checked that they did not connect loci known to map in different chromosomal regions. We then noted information about

single linkages among loci, which could provide connections between nearby doubly linked contigs in the course of integrating the top-down and bottom-up maps. Of the 11,750 STSs successfully screened against the YAC library, 10,850 (92%) showed single linkage to other STSs on the same chromosome. The remaining 8% were not included in the STS-content map.

Integrated map. We next sought to construct an integrated map by combining the STS-content, RH, and genetic linkage information. Each chromosome was treated separately: Only loci that had been assigned to the chromosome were used. Possible orders for the loci were compared by means of a linear scoring function, with the following three components: (i) continuity of STS content, reflecting whether the loci were present in the same YACs; (ii) continuity of RH linkage, reflecting whether the loci were present in the same RH hybrids; and (iii) consistency with top-down maps, incorporating a modest penalty for each violation of the genetic order or RH framework order. The specific parameters were chosen on the basis of the expected chance of concordance and discordance for nearby loci, so that the overall scoring function approximated a logarithm-likelihood for the order (39). The "optimal" order for the loci was found by combinatorial search through simulated annealing. Once the basic orders were established, incomplete addresses were used to identify additional links between nearby loci. The orders were then subjected to local optimization, manual inspection, and refinements where appropriate.

Gap closure. Loci fell into contigs of consecutive STSs connected by YACs and separated by gaps with no apparent YAC connection. Many of these apparent gaps are likely to be undetected overlaps; theoretical considerations would suggest that most gaps should actually be closed (17). We attempted to close these gaps by using non–STS-based information from the recent CEPH physical mapping project (7), inferring YAC overlaps on the basis of fingerprint analysis and Alu-PCR hybridization. Because the Alu-PCR hybridization data have a high false positive rate, gaps were closed only when there were at least seven hybridization links between adjacent contigs. Such closures should usually be correct, because only 3% of pairs of distant contigs meet this criterion. The data indicate overlap for about 50% of adjacent contigs. These gaps were declared tentatively closed, pending direct evaluation.

Description of the Map

The final map contains 15,086 loci, distributed across the 22 autosomes and two sex chromosomes (Tables 2, 3, and 4). The

Table 2. Types of STSs. Chr., chromosome.

Chr.	Total STSs	Random STS*	Genes		Genetic markers		Other loci	ESTs (obs/exp)†
			ESTs	GenBank	Généthon	CHLC		
1	1,374	252	275	106	460	153	128	1.4
2	1,275	307	181	67	452	146	122	0.8
3	1,097	269	181	64	353	134	96	0.9
4	919	210	112	45	281	121	150	0.7
5	858	196	125	30	312	97	98	0.8
6	858	181	114	39	312	108	104	0.8
7	781	168	141	39	272	83	78	1.1
8	739	183	104	35	248	104	65	0.7
9	577	132	106	30	188	68	53	1.1
10	719	154	131	26	281	60	67	1.1
11	706	122	140	42	272	64	66	1.5
12	707	132	104	64	250	91	66	1.0
13	418	102	48	13	164	54	37	0.6
14	489	106	95	27	163	53	45	1.2
15	428	97	97	22	145	30	37	1.3
16	435	87	79	18	180	32	39	1.2
17	447	66	97	39	186	34	25	1.9
18	403	91	46	18	136	64	48	0.7
19	246	23	45	20	121	15	22	2.6
20	386	84	68	26	144	32	32	1.1
21	156	28	18	12	61	13	24	0.8
22	274	19	38	17	67	12	122	2.6
X	587	145	63	28	216	28	107	0.6
Y‡	207	0	0	0	0	0	207	
Total	15,086	3,154	2,408	827	5,264	1,595	1,838	1.0

*Unbiased STSs, generated by sequencing from a random genomic library. †Ratio of observed (obs) number of ESTs divided by expected (exp) number, assuming that ESTs follow the same distribution as random STSs. ‡From our previously reported work (3).

10,850 loci mapped on YACs fall into 653 contigs connecting an average of 17 STSs each before gap closure and 377 contigs with an average of 29 STSs after gap closure. We examined the local density of YAC hits and contigs across the length of each chromosome. The results were relatively similar across the genome, with the notable exception of the chromosomes 1p36, 19, 22, and X. The map has less continuity in these regions, apparently because of systematic underrepresentation in the CEPH Mega-YAC library (see YAC density in Table 4), a problem that has been previously noted (7). Chromosome X is underrepresented because the library was made from a male cell line. The autosomal deficits could reflect cloning biases of the yeast host, inasmuch as these are all regions of high GC content (40).

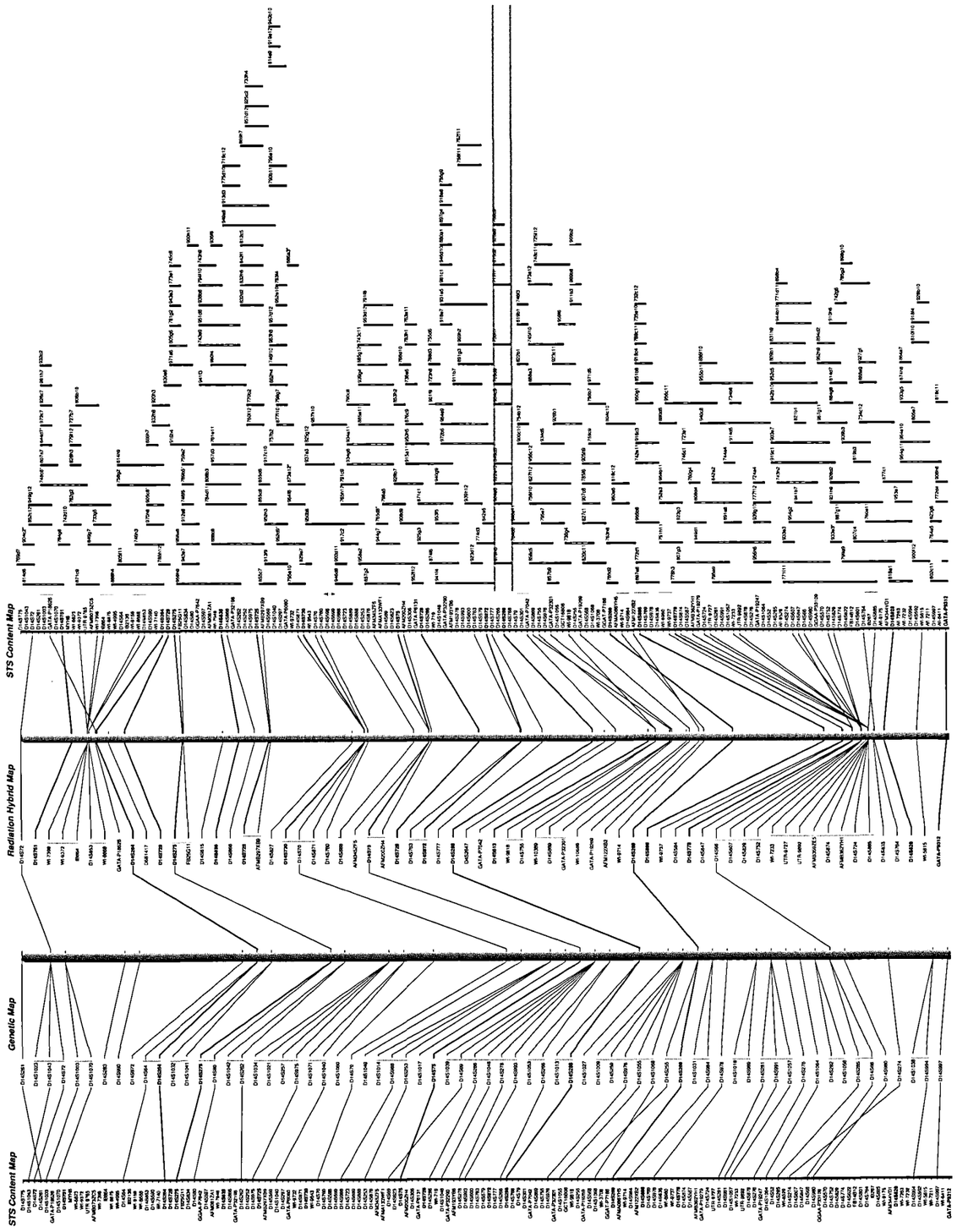
The physical map contains a wealth of information, which is ill-suited for presentation in traditional printed form. The complete physical map—including the STS sequences, RH retention patterns, YAC addresses, and order of loci—would require more than 900 journal pages to display. A compressed view of chromosome 14 is shown in Fig. 3, to illustrate the general nature of the map. The complete data for the map can be freely accessed through a World Wide Web server at the Whitehead Institute (http://www-genome.wi.mit.edu/), which includes various tools for analysis.

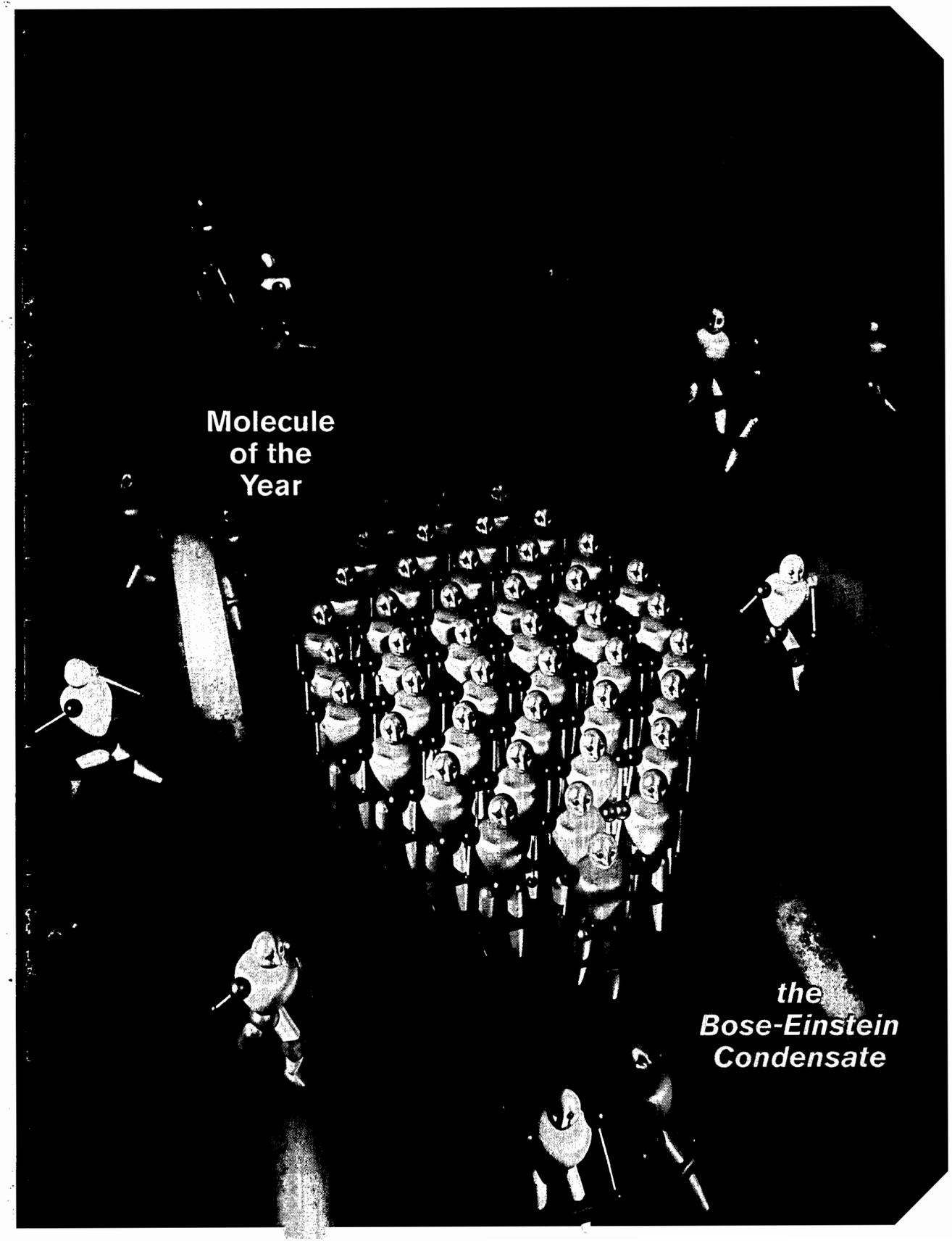
Coverage

We sought to determine how much of the human genome is covered by the physical map. For this purpose, we derived a new collection of random STSs—by sequencing random clones from an M13 library, selecting PCR primers, and retaining those loci that gave consistent amplification of a single fragment in control experiments. The first 100 STSs produced in this fashion were then screened against the NIGMS 1 hybrid panel, the RH panel, and the YAC library. Because the goal was to obtain an unbiased assessment of coverage, special efforts were made to obtain complete data for each locus.

RH data was obtained for all 100 STSs. (In six cases, it was necessary to resort to acrylamide gel electrophoresis of radioactively labeled products to circumvent problems posed by rodent background.) All 100 loci could be positioned on the RH map with a lod ≥8, on the correct chromosome as determined by the polychromosomal hybrid panel (41). The RH map thus appears to cover the vast majority of the human genome.

YAC screening data was also obtained for all 100 STSs. Two STSs detected no YACs in the library, consistent with previous observations that about 2% of DNA sequences appear to be absent from the CEPH Mega-YAC library (7). Four STSs detected YAC hits, but none with links to another STS in the correct chromosomal region; these loci





COVER

Artist's conception of supercooled atoms in an exotic phase of matter called the Bose-Einstein condensate, which is Molecule of the Year for 1995. Each of the atoms in the condensate (in blue) has the same quantum mechanical wave function, and so they all move as one. Atoms outside the condensate move faster and in all directions. See Editorial on page 1901 and Molecule of the Year section beginning on page 1902. [Illustration: Steve Keller]

An STS-Based Map of the Human Genome

Thomas J. Hudson,* Lincoln D. Stein, Sebastian S. Gerety, Junli Ma, Andrew B. Castle, James Silva, Donna K. Slonim, Rafael Baptista, Leonid Kruglyak, Shu-Hua Xu, Xintong Hu, Angela M. E. Colbert, Carl Rosenberg, Mary Pat Reeve-Daly, Steve Rozen, Lester Hui, Xiaoyun Wu, Christina Vestergaard, Kimberly M. Wilson, Jane S. Bae, Shanak Maitra, Soula Ganiatsas, Cheryl A. Evans, Margaret M. DeAngelis, Kimberly A. Ingalls, Robert W. Nahf, Lloyd T. Horton Jr., Michele Oskin Anderson, Alville J. Collymore, Wenjuan Ye, Vardouhie Kouyoumjian, Irena S. Zemsteva, James Tam, Richard Devine, Dorothy F. Courtney, Michelle Turner Renaud, Huy Nguyèn, Tara J. O'Connor, Cécile Fizames, Sabine Fauré, Gabor Gyapay, Colette Dib, Jean Morissette, James B. Orlin, Bruce W. Birren, Nathan Goodman, Jean Weissenbach, Trevor L. Hawkins, Simon Foote, David C. Page, Eric S. Lander*

A physical map has been constructed of the human genome containing 15,086 sequence-tagged sites (STSs), with an average spacing of 199 kilobases. The project involved assembly of a radiation hybrid map of the human genome containing 6193 loci and incorporated a genetic linkage map of the human genome containing 5264 loci. This information was combined with the results of STS-content screening of 10,850 loci against a yeast artificial chromosome library to produce an integrated map, anchored by the radiation hybrid and genetic maps. The map provides radiation hybrid coverage of 99 percent and physical coverage of 94 percent of the human genome. The map also represents an early step in an international project to generate a transcript map of the human genome, with more than 3235 expressed sequences localized. The STSs in the map provide a scaffold for initiating large-scale sequencing of the human genome.

A physical map affording ready access to all chromosomal regions is an essential prerequisite for the international effort to sequence the entire human genome. In the shorter term, it is also a key tool for positional cloning of disease genes and for studies of genome organization. Physical maps have evolved over the past decade from their initial conception as a set of overlapping clones (*1*) to the more recent idea of a well-spaced collection of unique landmarks called sequence-tagged sites (STSs), each defined by a polymerase chain reaction (PCR) assay (*2*–*4*). The U.S. Human Genome Project, for example, has set a target of a physical map consisting of 30,000 STSs spaced at intervals of about 100 kb (*5*).

By focusing on STS landmarks, genome researchers sought to insure against the inevitable problems inherent in any given clone library (*2*). The wisdom of this approach was borne out as it emerged that yeast artificial chromosomes (YACs), the best clones for covering large distances, suffer from high rates of chimerism and rearrangement and thus are unsuitable for genomic sequencing (*6*, *7*). STS-based maps sidestep this problem by having a sufficiently high density of landmarks that one can rapidly regenerate physical coverage of any region by PCR-based screening of clones appropriate for sequencing—such as cosmids, bacterial artificial chromosomes, and P1-artificial chromosomes (*8*).

STS-based physical maps with extensive long-range continuity have been constructed for only a handful of human chromosomes: 3, 12, 16, 21, 22, and Y (*3*, *4*, *9*, *10*). These combined maps cover just less than 20% of the genome with about 1600 STSs, and the average spacing on most of these chromosomes is about 250 kb. Projects are also underway for a few additional chromosomes (*11*). An international collaboration among the Centre d'Etude du Polymorphisme Humain (CEPH), Généthon, and Whitehead genome centers has also produced a clone-based physical map estimated to cover up to 75% of the genome in overlapping YAC clones (*7*). The map is clone-based, rather than STS-based, because it was primarily assembled by detecting physical overlaps among the clones themselves (by means of cross-hybridization and fingerprinting methods), with only a sparse set of STS landmarks used as anchors (786 loci fully screened and 1815 loci partially screened on YACs). The map is quite valuable for positional cloning projects, but it does not provide a scaffold for sequencing the human genome: The YAC clones themselves are not suitable for sequencing, and the STS coverage is too sparse to regenerate substantial physical coverage.

Here, we report the construction of an STS-based physical map of the human genome containing more than 15,000 loci, with an average spacing of 199 kb. The map covers the vast majority of the human genome and provides a scaffold for initiating large-scale sequencing.

Basic Strategy

We used three mapping methods to gain information about the proximity of STS loci within the human genome.

1) STS-content mapping. YAC libraries are screened by PCR to identify all clones containing a given locus (*12*). Nearby loci tend to be present in many of the same clones, allowing proximity to be inferred. STS-content linkage can be detected over distances of about 1 Mb, given the average insert size of the YAC library used here.

2) Radiation hybrid (RH) mapping. Hybrid cell lines, each containing many large chromosomal fragments produced by radiation of the relevant region on a high-resolution RH panel in parallel with screening them on the BAC library. As a simple test, we scored the STSs from a 3-Mb region on chromosome 6 on the G3 RH panel and were able to readily infer the fine-structure order of nearly all the loci with high confidence (*45*).

The use of STS-based maps as a scaffold for large-scale sequencing has several advantages: It can be initiated now with the existing STS-based map; it automatically anchors sequences in the genome; it does not require chromosome-specific libraries, which involve specialized preparation procedures and often have cryptic biases; it allows improved libraries to be substituted as they become available; and it promotes decentralization by allowing sequencing efforts to focus on regions of any given size, in contrast to entire chromosomes.

In summary, the physical map must still be refined but is already adequate to allow initiation of the international project to sequence the entire human genome—a landmark effort that will set the stage for the biology of the next century.

T. J. Hudson, S. S. Gerety, J. Ma, A. B. Castle, J. Silva, D. K. Slonim, R. Baptista, L. Kruglyak, S.-H. Xu, X. Hu, A. M. E. Colbert, C. Rosenberg, M. P. Reeve-Daly, S. Rozen, L. Hui, X. Wu, C. Vestergaard, K. M. Wilson, J. S. Bae, S. Maitra, S. Ganiatsas, C. A. Evans, M. M. DeAngelis, K. A. Ingalls, R. W. Nahf, L. T. Horton Jr., M. O. Anderson, A. J. Collymore, W. Ye, V. Kouyoumjian, I. S. Zemsteva, J. Tam, R. Devine, D. F. Courtney, M. T. Renaud, H. Nguyen, T. J. O'Connor, B. W. Birren, N. Goodman, T. L. Hawkins, and S. Foote, Whitehead–Massachusetts Institute of Technology (MIT) Center for Genome Research, Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA. L. D. Stein, Whitehead–MIT Center for Genome Research, Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA, and Department of Pathology, Brigham and Women's Hospital, Boston, MA 02115, USA. C. Fizames, S. Fauré, G. Gyapay, C. Dib, and J. Weissenbach, Généthon, CNRS URA 1922, 1 rue de l'Internationale, 91000 Evry, France. J. Morissette, Généthon, CNRS URA 1922, 1 rue de l'Internationale, 91000 Evry, France, and Centre de Recherche du Centre Hospitalier de l'Université Laval, 2705 Boulevard Laurier, Ste-Foy, Québec G1V 4G2, Canada. J. B. Orlin, Sloan School of Management, MIT, Cambridge, MA 02139, USA. D. C. Page, Whitehead–MIT Center for Genome Research, Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA; Howard Hughes Medical Institute, MIT, 9 Cambridge Center, Cambridge, MA 02139, USA; and Department of Biology, MIT, 9 Cambridge Center, Cambridge, MA 02139, USA. E. S. Lander, Whitehead–MIT Center for Genome Research, Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA, and Department of Biology, MIT, 9 Cambridge Center, Cambridge, MA 02139, USA.

*To whom correspondence should be addressed.

REFERENCES AND NOTES

1. M. Olson *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 7826 (1986); A. Coulson *et al.*, *ibid.*, p. 7821; Y. Kohara *et al.*, *Cell* **50**, 495 (1987).
2. M. Olson, L. Hood, C. Cantor, D. Botstein, *Science* **245**, 1434 (1989).
3. S. Foote, D. Vollrath, A. Hilton, D. C. Page, *ibid.* **258**, 60 (1992).
4. I. Chumakov *et al.*, *Nature* **359**, 380 (1992).
5. F. Collins and D. Galas, *Science* **262**, 43 (1993).
6. L. Selleri *et al.*, *Genomics* **14**, 536 (1992); M. Haldi *et al.*, *ibid.* **24**, 478 (1994).
7. I. Chumakov *et al.*, *Nature* **377**, S175 (1995).
8. H. Shizuya *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 8794 (1992); P. A. Ioannous *et al.*, *Nature Genet.* **6**, 84 (1994).
9. R. M. Gemmill *et al.*, *Nature* **377**, S299 (1995); N. A. Doggett *et al.*, *ibid.*, p. S335; J. E. Collins *et al.*, *ibid.*, p. S367.
10. K. Krauter *et al.*, *ibid.*, p. S321.
11. E. D. Green *et al.*, *Hum. Mol. Genet.* **3**, 489 (1994); J. Quackenbush *et al.*, *Genomics* **29**, 512 (1995); T. Alitalo *et al.*, *ibid.* **25**, 691 (1995).
12. E. D. Green and M. V. Olson, *Science* **250**, 94 (1990).
13. D. R. Cox, M. Burmeister, E. Roydon Price, S. Kim, R. M. Myers, *ibid.*, p. 245; M. A. Walter *et al.*, *Nature Genet.* **7**, 22 (1994).
14. D. Botstein *et al.*, *Am. J. Hum. Genet.* **32**, 314 (1980); J. L. Weber and P. E. May, *ibid.* **44**, 388 (1989).
15. J. Murray *et al.*, *Science* **265**, 2049 (1994).
16. J. Ott, *Analysis of Human Genetic Linkage* (Johns Hopkins Press, Baltimore, MD, 1991).
17. R. Arratia *et al.*, *Genomics* **11**, 806 (1991).
18. Loci were used only if they produced a single clear band visible by ethidium bromide staining, except for genetic markers, which were used even when they produced more than one band on an agarose gel (in view of their value in providing top-down orientation). Assays meeting this criterion are more likely, although not certain, to represent single unique loci in the genome.
19. Sequence data were analyzed with the Whitehead/MIT STS Pipeline software, which removes vector sequences, identifies duplicate sequences, and uses sequence similarity programs (FASTA and BLASTN) to eliminate known repeat sequences. Primers were chosen with PRIMER (M. J. Daly, S. Lincoln, E. S. Lander, Whitehead Institute) having the desired T_m (temperature at which 50% of double-stranded DNA is denatured) for primers set at 58°C.
20. M. Boguski and G. Schuler, *Nature Genet.* **10**, 369 (1995).
21. Nonredundant ESTs were part of the UniEST set prepared by M. Boguski and G. Schuler at the National Center for Biotechnology Information, derived from the dbEST database. These ESTs were taken from the Washington University–Merck EST project and the GenExpress project; R. Houlgatte *et al.*, *Genome Res.* **5**, 272 (1995).
22. R. Berry *et al.*, *Nature Genet.* **10**, 415 (1995); M. D. Adams *et al.*, *Nature* **377**, S3 (1995).
23. To select primers from ESTs, we modified the STS pipeline (*19*) to select shorter PCR products of 100 to 150 bp near the end of the 3′-UTRs [but 20 bp away from the polyadenylate [poly(A)] tract] in a region of high sequence quality.
24. J. Weissenbach *et al.*, *Nature* **359**, 794 (1992); G. Gyapay *et al.*, *Nature Genet.* **7**, 247 (1994); C. Dib *et al.*, *Nature*, in press.
25. C. J. Bell *et al.*, *Hum. Mol. Genet.* **4**, 59 (1995).
26. STSs were kindly shared by D. Cox and R. Myers, Stanford University, Stanford, CA, and J. Gastier, Harvard University, Cambridge, MA.
27. Some incomplete addresses could be resolved by a simple band-matching test with complete addresses by using CEPH fingerprint data, as described (*25*). Others could be resolved by virtue of comparison with complete addresses for nearby STSs.
28. Dot-blotted PCR products were initially detected by using ECL kits (Amersham), as described (*25*). We later switched to overnight hybridization with a biotinylated oligonucleotide probe to an internal sequence, followed by chemiluminescent detection with a peroxidase catalyzed luminol reaction, as described [R. P. M. Gijlswijk *et al.*, *Mol. Cell. Probes* **6**, 223 (1992)]. STSs known to contain an internal repeat sequence such as CA or AGAT were probed with an oligonucleotide for the repeat. Other STSs were probed with a specific internal oligonucleotide, having a T_m of 58°C. Computer images of each hybridization were obtained with a CCD camera. VIEW software (C. Rosenberg; Whitehead Institute) was used to locate and determine the intensity of positive dots. A small proportion of STSs were screened by standard agarose gel stained with ethidium bromide.
29. It is not possible to draw conclusions about library coverage from the overall number of STSs with no definite addresses, because many of these represented weak PCR assays that sometimes worked on human control DNAs but failed on YAC pools.
30. The probability that a unique sequence would occur more than 15 times in a random library with 8.4-fold coverage is about 1%. Some of these STSs may thus be unique loci, but they were excluded to guard against repeats.
31. G. Gyapay *et al.*, *Hum. Mol. Genet.*, in press.
32. Dot-blotted PCR products were detected as for the YAC screening, except that STSs containing CA repeats were screened with oligonucleotides containing unique internal sequence rather than $(CA)_n$, because the latter produced high background.
33. H. L. Drwinga *et al.*, *Genomics* **16**, 311 (1993); T. J. Hudson *et al.*, *ibid.* **13**, 622 (1992). In 300 cases, ambiguous or conflicting assignments were resolved by using the NIGMS Human/Rodent Somatic Cell Hybrid Panel #2, described in B. L. Dubois and S. Naylor *et al.*, *ibid.* **16**, 315 (1993).
34. In 151 cases, STSs were chromosomally assigned by virtue of having at least three single links to other markers on a chromosome and no links to any loci on any other chromosome.
35. About 200 such conflicts were resolved. Half were resolved by repeating the typing of the somatic cell hybrid panel. For the other half, an STS was demonstrated to amplify products from more than one chromosome. Such STSs were discarded.
36. RHMAPPER (L. Kruglyak, D. K. Slonim, L. D. Stein, E. S. Lander, unpublished data) uses a hidden Markov model to account for breaks in diploid DNA and for false positives and negatives, as in E. S. Lander and P. Green, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 2363 (1987); and E. S. Lander and S. Lincoln, *Genomics* **14**, 604 (1992). Framework maps were initiated and extended by a greedy algorithm and then subjected to local permutation tests, thereby allowing for efficient exploration of a vast space of possible orders.
37. In most cases, frameworks for the two chromosomal arms were constructed separately and then oriented and joined by using information from the genetic map. There is significant pairwise RH linkage (at lod > 5.0) between framework markers on opposite sides of the centromere on nine of the final framework maps, but not on the other 14.
38. M. R. James *et al.*, *Nature Genet.* **8**, 70 (1994).
39. For each pair of consecutive STSs, a positive score a_{++} was added for each YAC containing both, a negative score a_{+-} for each YAC containing one but not the other, a positive score b_{++} for each hybrid containing both, and a negative score b_{+-} for each hybrid containing one but not the other. Letting x_{11}, x_{10}, and x_{00} denote the probabilities that two STSs separated by about 500 kb would be observed to be both present, both absent, or one present and one absent in a randomly chosen YAC, we set $a_{++} = \log(x_{11}/x_{00})$ and $a_{+-} = \log(x_{10}/x_{00})$. The constants b_{++} and b_{+-} were defined similarly with respect to individual RH cell lines. The various probabilities were calculated on the basis of the distribution of fragment sizes and the inferred false positive and false negative rates. These four parameters, a_{++}, a_{+-}, b_{++}, and b_{+-}, were thus determined directly from the data and were not optimized. The two weighting parameters for conflicts with the genetic and framework RH maps were chosen by optimization in test cases.
40. S. Saccone *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 4913 (1992).
41. Three markers mapped into large centromeric intervals on the correct chromosome; they had high lod scores but were about 30 cR away from the closest marker. All were confirmed by double-linkage with YACs. For three other markers, chromosomal assignment could not be obtained from polychromosomal hybrid panels because of rodent background.
42. For one of these four loci, there was a (presumably chimeric) single YAC link to a marker on the same chromosome but located 70 cR from the correct location.
43. Three of the loci belonged to doubly linked contigs that were anchored by virtue of a CHLC genetic marker.
44. If gene promoters on chromosome X have the same average expression level as on autosomes, then the fact that only one X chromosome is active (due to hemizygosity in males and X inactivation in females) would cause transcripts from X-linked genes to be half as abundant. Because half of the cDNAs came from nonnormalized libraries and half from normalized libraries, the occurrence of ESTs in the relatively small set examined will partly reflect abundance. This issue will recede when enough ESTs have been isolated to overcome issues related to message levels. Underrepresentation of chromosome X could also conceivably represent some other systematic bias of which we are not aware.
45. T. J. Hudson *et al.*, data not shown.
46. J. S. Heslop-Harrison *et al.*, *Hum. Genet.* **84**, 27 (1989).
47. We thank A. Kaufman, O. Merport, and J. Spencer for technical assistance; L. Bennett for computer system administration; G. Rogers and M. Foley for assistance with media preparation and glass washing; S. Gordon, A. Christopher, P. Mansfield, and others at Intelligent Automation Systems for assistance in design, construction, and maintenance of automation equipment; D. Cox, R. Myers, J. Sikela, M. Adams, J. Murray, and K. Buetow for sharing data including sequences and markers; M. Boguski and G. Schuler for assistance in analysis of EST sequences; and D. Cohen and I. Chumakov for sharing the CEPH YAC library in 1992 and for public distribution of their STS-content, Alu-PCR, and fingerprint data. Supported by NIH award HG00098 to E.S.L. and by the Whitehead Institute for Biomedical Research. T.J.H. was a recipient of a Clinician Scientist Award from the Medical Research Council of Canada. L.K. is a recipient of a Special Emphasis Research Career Award (HG00017) from the National Center for Human Genome Research.

17 October 1995; accepted 8 November 1995

genetic maps. By measuring the frequency of such occurrences as a function of the distance between the loci in the STS map, we estimated that about 0.5% of the loci may be significantly misplaced in the maps.

In summary, the local order in the map must be regarded as uncertain. There will surely be many errors requiring attention and correction. The effective resolution of the map is certainly lower than the average spacing between loci and may be about 1 Mb. To improve the local accuracy of the maps, investigators interested in particular regions would be well advised to retest the STSs against an RH panel with higher resolution [such as the G3 panel developed by D. Cox and R. Myers, in which the fragments are about 1/10 those in the GeneBridge 4 panel] and against regional YAC panels, as described above. In this fashion, the map provides the tools for its own refinement.

Finally, we note that direct comparison of our STS-based map with the recently reported YAC-based map (*7*) is difficult, because of the very different natures of the maps. For example, it is not meaningful to compare the STS orders in the two maps: The YAC-based map almost exclusively involved genetic markers and provided no independent information about locus order, but instead simply adopted the genetic order. It is also problematic to compare the specific YACs identified, because the YAC-based map involved only partial screening of most STSs and did not fully distinguish clones actually belonging to paths through a region from those representing false positive hybridization. At the grossest level, it is possible to compare the coverage of the maps: The current map appears to cover about 95% of the genome (the precise amount depends on the type of mapping information used), whereas the other map was reported to cover about 75%. More detailed comparison would be worthwhile, as it would likely lead to improvements in both maps.

Table 4. STS-content mapping of YACs.

Chr.	STS-content mapped loci	STS spacing (kb/STS)	No. of YACs*	Contigs			YAC hits per STS‖	STSs per contig¶
				Before gap closure†	After gap closure‡	Avg. size (Mb)§		
1	1,048	237	1,393	49	34	7.3	6.7	30.8
2	933	258	1,469	56	20	12.0	7.3	46.7
3	791	255	1,192	46	30	6.7	7.5	26.4
4	718	267	1,272	42	11	17.4	8.2	65.3
5	651	281	1,163	35	19	9.6	7.9	34.3
6	641	269	1,091	40	24	7.2	7.8	26.7
7	559	288	942	39	13	12.4	7.8	43.0
8	552	265	945	23	11	13.3	8.0	50.2
9	394	347	675	28	12	11.4	8.0	32.8
10	519	262	750	36	26	5.2	6.8	20.0
11	490	277	696	23	14	9.7	7.2	35.0
12	509	265	842	29	16	8.4	7.4	31.8
13	300	308	556	12	5	18.5	8.0	60.0
14	352	249	593	9	6	14.6	8.0	58.7
15	301	279	439	16	10	8.4	7.0	30.1
16	255	362	308	26	16	5.8	6.0	15.9
17	267	325	330	27	17	5.1	5.8	15.7
18	315	254	478	16	8	10.0	7.6	39.4
19	79	800	76	17	15	4.2	4.7	5.3
20	266	255	328	15	10	6.8	6.6	26.6
21	113	325	182	4	2	18.4	8.0	56.5
22	182	223	134	11	11	3.7	5.4	16.5
X	408	379	406	53	46	3.4	4.7	8.9
Y#	207	128	234	1	1	26.4	4.1	207.0
Total	10,850	276	16,494	653	377	8.0	7.3	28.8

*Number of YACs hit by at least two STSs on the chromosome. YACs hit by only one STSs are omitted. †Number of contigs based only on STS-content data. ‡Number of contigs after gap closure by Alu-PCR hybridization and fingerprint data. §Average contig size estimated by assuming that 94% of the chromosome length is covered and dividing by the number of contigs. ‖Includes all YAC hits in the library screen (not limited to YACs having multiple STS hits on the chromosome) and thus reflects coverage by the library. ¶Reflects contigs after gap closure. #From our previously reported work (*3*).

Distribution of Genes

The map also sheds light on the organization of the human genome. By comparing the chromosomal distribution of the expressed sequences to the chromosomal distribution of the random single-copy sequences (both determined in the same manner), one can draw inferences about the density of genes on different chromosomes. We compared the observed number on each chromosome to the expected number, assuming that expressed sequences have the same distribution as random STSs (Table 2). Chromosomes 1, 11, 17, 19, and 22 showed a statistically significant excess of expressed sequences ($P = 0.001$ after correction for multiple testing). Chromosomes 17, 19, and 22, which showed the greatest excess, have been previously suggested to have a high density of genes on the basis of indirect evidence (*40*). Chromosome X was the only chromosome to show a statistically significant deficit of expressed sequences—only about half as many as expected. This would suggest that there is a low gene density on this sex chromosome, although alternative explanations are possible (*44*). We also analyzed the raw data from two recent papers reporting chromosomal assignment of expressed sequence tags (ESTs) (*21*, *22*) and found a similar deficit of X-linked loci.

A Scaffold for Sequencing the Genome

As genetic and physical maps approach their intended goals, attention is turning to the challenge of sequencing the entire human genome. A key issue is how to obtain the required sequence-ready clones. STS-based maps provide a general solution by making it possible to generate extensive physical coverage of a region by screening a single high-quality human genomic library.

One could, for example, proceed as follows: Screen the STSs in a region against a bacterial artificial chromosome (BAC) library having 150-kb inserts and 10-fold coverage, use a simple fingerprinting scheme to detect overlaps among adjacent clones, and select a minimally overlapping set for sequencing. Given a physical map containing 30,000 ordered STSs, one would screen about 100 STSs and fingerprint about 520 BACs to cover a 10-Mb region; this task could be readily accomplished in a few days with modest automation and would not contribute significantly to the cost of sequencing. The resulting BACs would be expected to cover about 95% of the region in ordered sequence contigs (*17*). The region could then be closed by straightforward walking—that is, serially screening the BAC library with STSs derived from sequences at the ends of each contig.

The current map falls short in terms of marker density and local order, but neither shortcoming poses a serious obstacle for initiating large-scale sequencing now. With the 15,000 STSs currently available, one should cover about 75% by direct screening, 90% by one round of walking, and more than 95% with two rounds (*17*). The desired map with 30,000 STSs will likely be available within the next 2 years, through current projects underway at several centers including our own. Uncertainties about locus order can be overcome simply by scoring the STSs from

tion breakage, are screened by PCR to identify those hybrids that have retained a given locus (*13*). Nearby loci tend to show similar retention patterns, allowing proximity to be inferred. RH linkage can be detected for distances of about 10 Mb, given the average fragment size of the RH panel used here.

3) Genetic mapping. A locus that is polymorphic in the human population can be screened by PCR to determine its inheritance patterns in families (*14*, *15*). Nearby loci tend to show similar inheritance patterns, allowing proximity to be inferred. Genetic linkage can be reliably detected over distances of about 30 Mb, given the recombination rate of human chromosomes (*16*).

These three methods were used to produce independent maps and then combined to produce an integrated map. Because RH mapping and genetic mapping can detect linkage over large regions (0.3 to 1% of the genome), comprehensive RH and genetic maps spanning all chromosomes can be assembled with a few thousand loci. The order of loci can be inferred from the extent of correlation in the retention or inheritance patterns, although estimates of fine-structure order are not precise. These methods can thus provide "top-down" information about global position in the genome.

In contrast, STS-content mapping provides "bottom-up" information. It reveals tight physical linkage among loci but is useful only over short distances and does not provide extensive long-range connectivity across chromosomes (*17*). Two STSs are said to be singly linked if they share at least one YAC in common and doubly linked if they share at least two YACs (*17*). Single linkage is an inadequate criterion for declaring adjacency of STSs, because of the high rate of YAC chimerism (about 50%) and the possibility of laboratory error. Double linkage, however, turns out to be a reliable indication, because two genomic regions are unlikely to be juxtaposed in multiple independent YACs. Accordingly, a three-step procedure was used. (i) STSs were assembled into doubly linked contigs (groups of STSs connected by double linkage). (ii) The doubly linked contigs were localized within the genome on the basis of RH and genetic map information about loci in the contig. (iii) Single linkage was then used to join contigs localized to the same small genomic region. The overall strategy is illustrated in Fig. 1. We now describe the data generation, map construction, and map analysis in greater detail.

Data Generation

Marker development. Over the course of the project, we tested 20,795 distinct PCR assays. These candidate STSs were initially characterized to see whether they were likely to detect a unique genomic locus (*18*) and whether they consistently yielded correct results on control samples under uniform production conditions. A total of 16,239 STSs met these stringent criteria and were used for mapping. The STSs fell into one of the following four categories.

1) Random loci. We generated 3027 working STSs by sequencing random human genomic clones and discarding those that appeared to contain repetitive sequences (*19*).

2) Expressed sequences. We developed 921 STSs from complete complementary DNA (cDNA) sequences in GenBank, taken from the Unigene collection (*20*). Another 3349 STSs were developed from expressed sequence tags (ESTs). Of these, 71% came from the dbEST database (*21*), 13% from the laboratory of Jim Sikela, 9% from the Institute for Genomic Research, and 7% from various other sources (*22*). We found that the success rate for STSs derived from the last 200 base pairs (bp) of 3′-untranslated regions (UTRs) of cDNAs was similar to that for STSs derived from random genomic DNA, consistent with the idea that introns rarely occur near the ends of 3′-UTRs (*23*). The results indicate that PCR assays can be readily derived for the vast majority of cDNAs.

3) Genetic markers. A total of 6986 loci were used, consisting of 5264 polymorphic loci developed at Généthon (primarily CA repeats) (*24*) and 1722 loci developed by the Cooperative Human Linkage Center (CHLC) (primarily tri- and tetranucleotide repeats) (*15*).

4) Other loci. A total of 1956 STSs were developed from various sources. These included 1091 CA-repeat loci developed at Généthon that were not sufficiently polymorphic to be useful for genetic mapping, as well as 865 loci from chromosome 22–specific and chromosome Y–specific libraries and gifts from other laboratories (*3*, *25*, *26*).

A total of 15,086 STSs appear in the final maps. The number of markers of each type appearing in the final STS-content, RH, and genetic maps is shown in Table 1.

STS-content mapping: Methodology. STSs were screened against 25,344 clones from plates 709 to 972 of the CEPH mega-YAC library (*7*), estimated to have an average insert size of 1001 kb and to provide roughly 8.4-fold coverage of the genome. To facilitate screening, we used a hierarchical pooling system. The library was divided into 33 "blocks," each corresponding to eight microtiter plates or roughly 0.25 genome equivalent. For each block, we prepared one "superpool" containing DNA from all the clones and 28 "subpools" by using a three-dimensional pooling system based on the row, plate, and column address of each clone. Specifically, there were 8, 8, and 12 subpools consisting of YACs in the same plate, row, and column, respectively. There was thus a total of 957 super- and subpools.

For blocks with a single positive YAC, the row, column, and plate subpools should specify the precise address of the YAC ("definite addresses"). If a block contained two or more positive YACs or if one of the three subpool dimensions did not yield a positive, partial information was obtained ("incomplete addresses") (*27*). Such incomplete addresses could consist of up to 12 possible addresses (for example, in the case that a column address was missing). Incomplete addresses were not used in initial map assembly but were used at the final stages to detect connections between nearby loci. Definite addresses composed 88% of the total hits.

Half of the markers were screened by a two-level procedure, in which we first identified the positive superpools and then tested only the corresponding subpools. The other half were screened by a one-level procedure, bypassing the superpools and directly screening all subpools. Although the latter procedure involves more reactions,

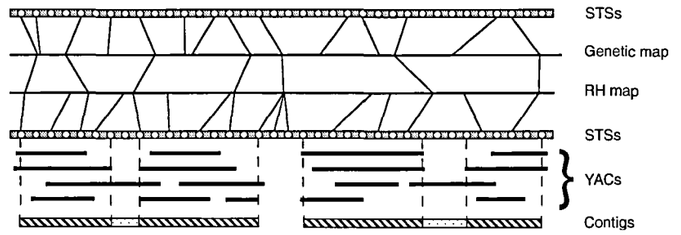


Fig. 1. Schematic diagram of the STS-based map. STSs are shown as circles on the first and fourth line. Loci that are genetically mapped or RH mapped are connected to the appropriate position on these maps, with connections between these maps in the cases of loci present in both maps. YACs containing STSs are shown below. The STSs fall into two singly linked contigs (stippled rectangles) and four doubly linked contigs (striped rectangles). Single linkage is not reliable for connecting arbitrary doubly linked contigs, but it is reliable in the case of anchored doubly linked contigs known to be adjacent on the genetic or RH map, as in the figure.

each locus is treated in an identical manner, which offers advantages for automation. In both procedures, we identified the positive pools by spotting the PCR reactions on membranes, hybridizing them to a chemiluminescent probe specific for each STS, capturing the resulting signal directly by a charge-coupled device (CCD) camera, and up-loading the results into our database (*28*); this approach proved to be much more efficient than the traditional detection procedure of gel electrophoresis.

Because the project involved processing more than 15 million reactions, laboratory automation was essential. We collaborated with an engineering firm, Intelligent Automation Systems, Incorporated, (IAS) of Cambridge, Massachusetts, to design and build various special-purpose machines to accelerate STS-based mapping.

The two-level screening procedure was carried out with a large robotic liquid-pipetting workstation and two custom-designed thermocyclers (Fig. 2). A laboratory information management system used the superpool results to automatically program the robotic workstation to set up the appropriate subpool screens. The system has a maximal throughput of 6144 PCR reactions per run.

The one-level screening procedure was made feasible by the development of a massively parallel factory-style automation system nicknamed the Genomatron (Fig. 2). The Genomatron was also developed in collaboration with IAS and consists of three stations. The first station assembles PCR reactions in custom-fabricated 1536-well microtiter "cards" and seals the wells by welding a thin plastic film across the card. The second station thermocycles the reactions by transporting the cards over three chambers that force temperature-controlled water to flow uniformly between the cards. The third station transfers the reactions from one microtiter card onto a hybridization membrane affixed to the bottom of a second microtiter card by piercing the first card with a bed of 1536 hypodermic needles and sucking the reactions downward with a vacuum plenum. These "filter cards" were then manually hybridized with a chemiluminescent probe and read by the CCD camera. The stations were computer controlled, and the microtiter cards were assigned a bar code to facilitate sample tracking. Each station was designed to process 96 microtiter cards, providing a throughput of nearly 150,000 reactions per run.

STS-content mapping: Results. A total of 11,750 STSs yielded from 1 to 15 definite YAC addresses and were considered successfully screened (*29*); typical loci yielded approximately one additional incomplete address. STSs having more than 15 definite hits were excluded as likely to detect multiple genomic loci (*30*).

The successfully screened loci produced an average of 6.4 YACs per STS, considering only definite addresses. A total of 18,879 YACs were hit by at least one STS. For these YACs, the average hit rate was 3.8 STSs per YAC. The average size of the YACs hit by the STSs was about 1.1 Mb (~10% greater than for the library as a

Table 1. Overview of mapped STSs.

STSs on final map	No. of loci
STS-content map	10,850
RH map	6,193
Genetic map	5,264
Intersection of	
STS-content and RH maps	4,036
STS-content and genetic maps	3,106
RH and genetic maps	887
All three maps	807
Total loci	15,086

Fig. 2. The first automated system developed for the project was (**A**) a robotic station to set up PCR reactions and (**B**) custom-built "waffle iron" thermocyclers accomodating 16 192-well microtiter plates; the system has a capacity of 6144 PCR reactions per run. The second automated system was the Genomatron, which consists of three robotic stations. PCR reactions are set up in 1536-well microtiter cards (consisting of 15 cm by 24 cm injection molded plastic cards with 1536 holes, to the bottom of which a plastic film is heat-sealed to create wells). The first station (**C**) assembles the PCR reactions. Each run can process up to 96 cards per run, providing a capacity of nearly 150,000 wells. Cards are dispensed by a coining mechanism and travel along a conveyor belt to substations containing a bar code reader; a 1536-head pipettor (**D**) that dispenses template DNAs to be screened; a 48-head pipettor that dispenses PCR primer mixes, including polymerase; a plate sealer that heat-seals a plastic film on the top of the card to create separate reaction chambers; and a refrigerated storage station. The second station is a thermocycler (**E**) that uses three large waterbaths. Up to 96 sealed cards containing PCR reactions are placed in a chamber that travels over the water baths, which pump water at the appropriate denaturing, annealing, and extension temperature. The third station is a parallel "spotting" device that transfers PCR reactions from a card to a nylon filter affixed to the bottom of a second card. After the two cards are aligned, a bed of 1536 hypodermic needles (**F**) pierces a sealed card containing the reactions while a vacuum manifold draws the reaction mixtures down onto the membrane on the second. The filter cards are manually hybridized and subjected to a chemiluminescent detection protocol. Light signals are recorded with a cooled CCD camera.

could thus not be localized on the STS-content map (*42*). These four STSs appear to be in regions of low YAC coverage, inasmuch as they hit one, one, one, and two YACs, respectively. The remaining 94 STSs could all be localized on the STS content map [with 91 being doubly linked and three being singly linked to existing contigs anchored in the correct chromosomal region in the top-down map (*43*)]. The 100 loci detected an average of 6.5 YACs.

The map covers the vast majority of the human genome. We estimate that 99% of random STSs can be readily positioned on the RH map, and 94% can be positioned on the STS-content map relative to YAC clones.

The physical map thus fills a major need in human genetics, providing a general method by which an investigator can map a locus in the human genome by screening readily available RH or YAC pools and comparing the resulting pattern with the map. To make this information easily accessible to the scientific community, we have written a "map server." The server reports the likely position of an STS given information about its YACs, RH pattern, and chromosomal assignment. The server is freely available via our World Wide Web site.

Fig. 3 (previous pages). Integrated map of human chromosome 14q. Long vertical lines represent the STS-content map (first and fourth lines, in black), genetic map (second line, in blue), and RH map (third line, in orange), in the same fashion as the diagram in Fig. 1. All three maps are drawn to equal length. The four columns of STS names correspond to the four lines. For the STS-content map, intermarker distance is not known and loci are displayed as equally spaced. For genetic and RH maps, loci are indicated at positions spaced proportionally along the map according to the respective metrics. Loci in common between two maps are connected by black lines. Loci belonging to the RH framework map (in which the relative ordering is supported by lod > 2.5) are shown in bold type and with thicker connecting lines. Loci derived from expressed sequences are shown in purple. YACs are displayed as black rectangles, to the right of the STSs that were found to be contained in the clone. YAC names are shown to the top right. Unfilled portions of YACs represent assays that were negative. Thin red lines in some YACs represent incomplete addresses that were resolved by virtue of overlap with addresses from a nearby locus. Gaps between contigs are shown as horizontal lines separating groups of YACs. Gaps that were likely to be undetected overlaps based on Alu-PCR hybridization or fingerprint information (see text) are shown in yellow; gaps for which there is no evidence of overlap are shown in gray. Vertical dotted gray bars indicate STSs with identical data for given mapping method. YACs detected by only a single STS were omitted from this display. These YAC addresses can be obtained from the Whitehead Institute–MIT Center for Genome Research World Wide Web server at URL http://www-genome.wi.mit.edu/. Figure represents slightly earlier version of the map, from the 14,000-marker stage.

Accuracy

Although the long-range order of the map is reliable because of top-down anchoring, precise local orders must be regarded as only approximate. Local ordering depends on the position of loci with respect to individual breakpoints, that is, the ends of YAC or RH fragments. The accuracy of such inference is limited by the presence of false positives and false negatives in our data, as well as by the presence of internal deletions in YACs. Whereas the long-range order tends to be over-determined in genomic maps, several alternative local orders may be reasonably compatible with the data. The "best" order may change with the alteration of a few data points.

We used three approaches to evaluate the accuracy of the data and the map.

1) Rescreening of loci on chromosome 14. Chromosome 14 was divided into 16 regions and regional YAC panels were defined, consisting of all clones hit by one or more loci in the region. For each regional YAC panel, individual DNAs were prepared from each clone. We tested 112 STSs against their corresponding panels to directly compare the results from high-throughput screening of pools with the screening of individual clones. We found a false positive rate of 5.5% and a false negative rate of 19.5% in our high-throughput screening data, both of which were consistent with our earlier indirect estimates. We constructed a new STS map of the chromosome using these more complete data; the new map showed about six instances of local reorderings involving two to five loci.

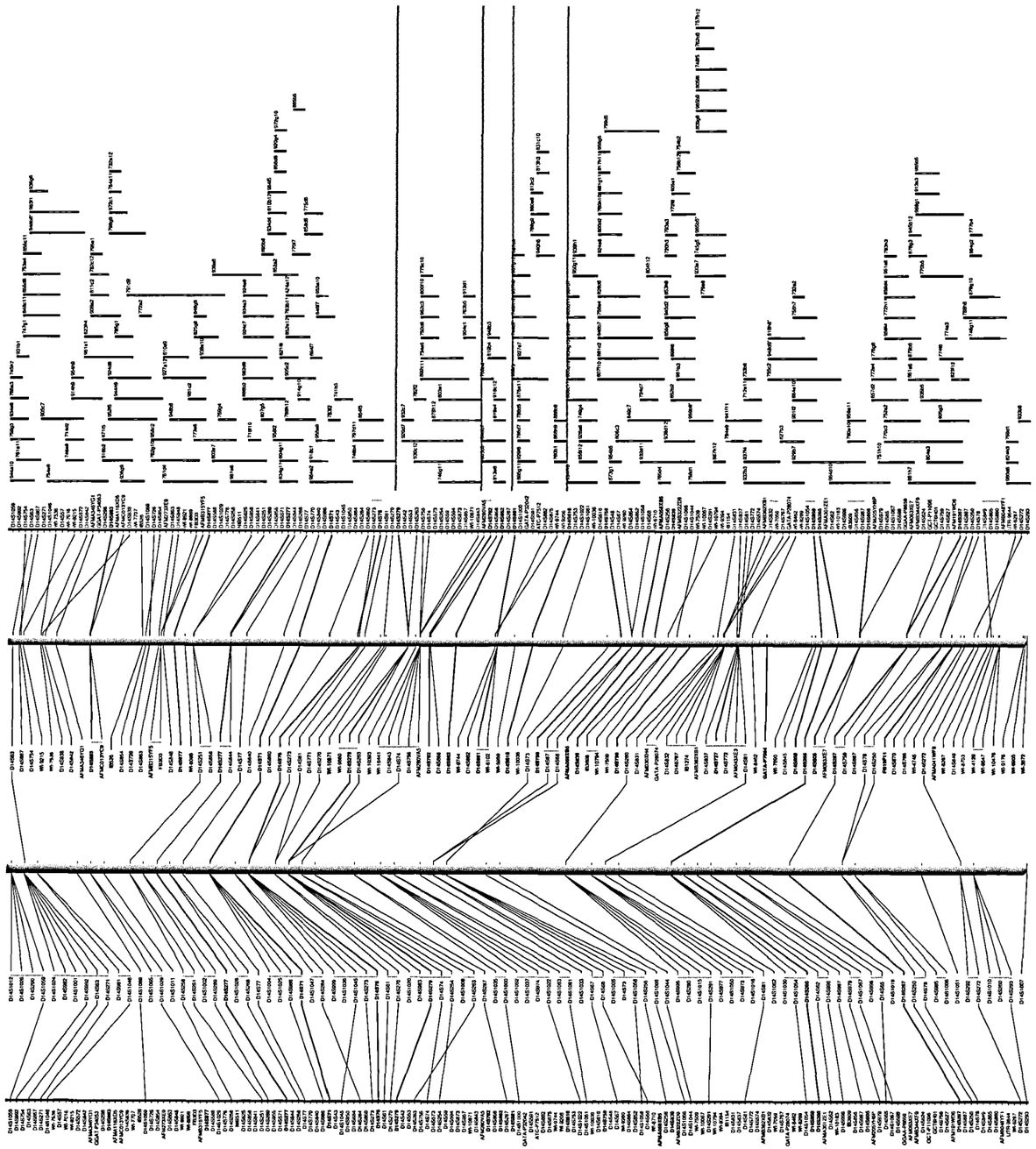
2) Comparison with an independently constructed map of chromosome 12. We compared our map with a recently reported map of this chromosome (*10*) containing enough loci in common to provide a meaningful test. Of 171 loci in common, there were about a dozen instances of small local inversions involving two to three adjacent markers. A substantial difference in position was seen for only a single marker, AFM263WH1. Our map shows tight STS-content linkage of this locus to genetic markers at 91 cM on the Généthon map, whereas the other map places it near genetic markers at 105 cM. In fact, the position on our map agrees well with the reported genetic map location for this locus (at 93 cM), so we believe it to be correct. In any case, the two maps showed relatively few conflicts.

3) Internal consistency checking. We looked for instances in which pairs of loci occurred in an order on the final STS map that was strongly disfavored by the RH or

Table 3. Genetic and RH maps. Dashes indicate not applicable.

Chr.	Physical length (Mb)*	Genetic map			RH map			
		No. of loci	Length (cM)	Genetic vs. physical (cM/Mb)	Framework map No. of loci	Total RH map	RH length (cR)†	RH vs. Physical (cR/Mb)
1	248	461	293	1.2	107	559	743	3.0
2	240	452	277	1.2	119	532	977	4.1
3	202	353	233	1.2	95	475	801	4.0
4	191	280	212	1.1	80	370	552	2.9
5	183	312	198	1.1	60	339	508	2.8
6	173	311	201	1.2	97	374	739	4.3
7	161	272	184	1.1	63	360	591	3.7
8	146	249	166	1.1	77	264	711	4.9
9	137	189	166	1.2	75	260	440	3.2
10	136	281	182	1.3	71	297	599	4.4
11	136	273	156	1.1	66	302	515	3.8
12	135	249	169	1.3	58	294	565	4.2
13	92	164	117	1.3	46	169	309	3.3
14	88	162	129	1.5	38	210	319	3.6
15	84	145	110	1.3	41	185	342	4.1
16	92	180	131	1.4	33	186	235	2.5
17	87	186	129	1.5	34	156	347	4.0
18	80	136	124	1.5	52	175	450	5.6
19	63	121	110	1.7	21	107	221	3.5
20	68	144	96	1.4	30	157	265	3.9
21	37	61	60	1.6	15	61	151	4.1
22	41	67	58	1.4	15	89	141	3.5
X	155	216	198	1.3	46	272	521	3.4
Y	26	–	–	–	–	–	–	–
Total	3,000	5,264	3,699	1.2	1,339	6,193	11,042	3.7

*Physical lengths were calculated on the basis of a genome of 3000 Mb, with proportional lengths of chromosomes as reported (*46*). †Total length of the RH framework map, omitting the large interval at the centromere.



whole), corresponding to 6.9-fold coverage of the genome. Some 78% of the STSs showed double-linkage to at least one other STS.

The false positive rate was investigated by regrowing and testing individual YACs. Several thousand addresses were tested, and 95% could be directly confirmed, with the remainder constituting actual false positives, deletions during regrowth, or technical failures during retesting. The false positive rate is thus at most 5% of definite addresses, and the chance of any particular YAC occurring as a false positive in a given screen is about 1.5×10^{-5}. False positive addresses thus will rarely create false links among STSs known to lie in the same genomic region. The false negative rate cannot be computed directly, but the fact that an average of 6.4 hits was seen in 8.4 genome equivalents suggests a rate of about 20%. False negatives pose a less serious problem than false positives (which join incorrect genomic regions), but they can lead to incorrect local ordering of STSs. The false positive and negative rates were reinvestigated once the maps were constructed, as discussed below.

Radiation hybrid mapping. STSs were screened against the GeneBridge 4 whole-genome radiation hybrid panel, consisting of 91 human-on-hamster somatic hybrid cell lines. Each line retains about one-third of the human genome in fragments of about 10 Mb in size. The GeneBridge 4 panel (Research Genetics, Huntsville, Alabama) was developed in the laboratory of P. Goodfellow and distributed to the scientific community as a resource for the mapping of expressed sequences. As part of a separate project, the panel has been characterized for more than 500 well-spaced genetic markers to confirm that substantial linkage can be obtained across the genome (*31*).

RH mapping was performed with essentially the same protocol as for the YAC screening: PCR reactions were set up either by the Genomatron (with each 1536-well microtiter card containing reactions for eight loci) or by the robotic workstation (by using 192-well microtiter plates), spotted on membranes, hybridized to a chemiluminescent probe, and detected by a CCD camera (*32*).

Scoring results from RH panels requires considerable caution. Human chromosomal fragments are present at various molarities among the hybrid cell lines; thus, the ability to detect their presence may vary with the sensitivity of each PCR assay. As a result, STSs that are immediately adjacent in the genome could conceivably give somewhat different retention patterns, which would limit the ability to determine fine-structure order. To minimize discrepancies due to assays near the limit of detection, we performed all assays in duplicate. Hybrids were scored if the two duplicates gave concordant positive or negative results but were recorded as "discrepant" if the duplicates were discordant. The mean discrepancy rate was 1.2%; loci with a discrepancy rate exceeding 4.5% were eliminated as unreliable.

A total of 6469 STSs were successfully screened on the GeneBridge 4 RH panel. The overall retention rate of the panel was 32% (or about 18% per haploid genome from the diploid donor cell).

Genetic mapping. Genetic linkage information was used from the recent Généthon linkage map of the human genome, containing 5264 polymorphic markers (*24*). Genetic linkage information was not incorporated for the 1722 CHLC genetic markers studied.

Chromosomal assignment. Before undertaking map construction, we attempted to assign all loci to specific chromosomes by multiple, independent methods. Most STSs were screened against the NIGMS 1 polychromosomal hybrid panel (*33*), resulting in unambiguous chromosomal assignment in about 75% of the cases (with the remainder having high background from the host genome or poor signal). STSs defining genetic markers typically had chromosomal assignments on the basis of linkage analysis. STSs were also assigned to chromosomes if they were tightly linked by RH screening or doubly linked by YAC screening to chromosomally assigned loci (*34*).

Some 96% of the loci could be chromosomally assigned, with the majority of these being assigned by at least two independent methods. Conflicting assignments were noted in a small proportion of cases (2%); these were subjected to intense scrutiny and resolved in the majority of cases (35). Loci that could not be reliably assigned to a chromosome were omitted from map construction, to avoid problems associated with chimeric linkages.

Personnel. The project was carried out during a period of 2.5 years by a team at Whitehead having an average of 16 people involved in mapping, three people involved in sequencing, and five people involved in data management and computational analysis.

Map Construction

Top-down maps. The genetic and RH maps are top-down maps, which provide a global framework and offer many tests of internal consistency. The first step in constructing an RH linkage map was to make high-quality "framework" maps across each chromosome. For this purpose, we included only loci with independent chromosomal assignments and with retention rates in the range of 10 to 60% (unusually high or low retention rates can produce spurious linkage). We wrote a computer package, RHMAPPER, that implements RH mapping for hybrids constructed from diploid sources and incorporates probabilistic error detection and error correction (36). Using this program, we generated a framework map—that is, an ordered set of markers such that each consecutive pair was linked with a lod score > 10 (lod score is the logarithm of the likelihood ratio for linkage), and the order was better than all local alternatives by a lod score > 2.5. The framework map included 1339 loci and provided complete connectivity across each chromosome arm with no gaps over 30 centiRays (cR) (cR is a measure of distance that is analogous to centimorgans but depends on the radiation dose). There were, however, large intervals across most centromeres (*37*), a phenomenon that has been previously seen for chromosome 11 (38). The total length of the map is 11,042 cR (omitting the centromeric intervals), corresponding to a fairly uniform average of about 300 kb/cR across most chromosomes.

We then localized the remaining markers relative to the framework map. These loci could not be uniquely ordered, either because of close proximity to a framework marker (loci with identical retention patterns cannot be ordered with respect to one another) or because of potentially erroneous typing results (that cause apparent "double-breaks" regardless of the interval in which the marker is placed). RHMAPPER allowed for the possibility of false positive and false negative typings and flagged probable errors (about two-thirds of which were found to be real errors in cases that were subsequently retested). The nonframework markers were estimated by the computer analysis to have an average residual error rate of just less than 1%. To reflect the uncertainty in order, each locus was assigned to the collection of intervals for which the lod score was within three of the optimal position. Loci were not included if they mapped more than 15 cR from a framework marker (that is, past the end of the map or in a large centromeric gap), because such positions could result from a high proportion of errors. In all, 6193 of 6469 loci tested were placed in the RH map.

Together, the two top-down maps contained a total of 10,572 loci. The reliability of the maps can be assessed by studying the loci in common. For loci present in both the genetic map and the framework RH map, there were only four conflicts in order; the loci involved were separated by 1 centimorgan (cM) in three cases and 3 cM in one case. The close agreement between the maps suggests that they correctly reflect the global order of loci in the genome.

Bottom-up map. Using the STS-content data, we assembled doubly linked contigs and checked that they did not connect loci known to map in different chromosomal regions. We then noted information about

single linkages among loci, which could provide connections between nearby doubly linked contigs in the course of integrating the top-down and bottom-up maps. Of the 11,750 STSs successfully screened against the YAC library, 10,850 (92%) showed single linkage to other STSs on the same chromosome. The remaining 8% were not included in the STS-content map.

Integrated map. We next sought to construct an integrated map by combining the STS-content, RH, and genetic linkage information. Each chromosome was treated separately: Only loci that had been assigned to the chromosome were used. Possible orders for the loci were compared by means of a linear scoring function, with the following three components: (i) continuity of STS content, reflecting whether the loci were present in the same YACs; (ii) continuity of RH linkage, reflecting whether the loci were present in the same RH hybrids; and (iii) consistency with top-down maps, incorporating a modest penalty for each violation of the genetic order or RH framework order. The specific parameters were chosen on the basis of the expected chance of concordance and discordance for nearby loci, so that the overall scoring function approximated a logarithm-likelihood for the order (39). The "optimal" order for the loci was found by combinatorial search through simulated annealing. Once the basic orders were established, incomplete addresses were used to identify additional links between nearby loci. The orders were then subjected to local optimization, manual inspection, and refinements where appropriate.

Gap closure. Loci fell into contigs of consecutive STSs connected by YACs and separated by gaps with no apparent YAC connection. Many of these apparent gaps are likely to be undetected overlaps; theoretical considerations would suggest that most gaps should actually be closed (*17*). We attempted to close these gaps by using non–STS-based information from the recent CEPH physical mapping project (*7*), inferring YAC overlaps on the basis of fingerprint analysis and Alu-PCR hybridization. Because the Alu-PCR hybridization data have a high false positive rate, gaps were closed only when there were at least seven hybridization links between adjacent contigs. Such closures should usually be correct, because only 3% of pairs of distant contigs meet this criterion. The data indicate overlap for about 50% of adjacent contigs. These gaps were declared tentatively closed, pending direct evaluation.

Description of the Map

The final map contains 15,086 loci, distributed across the 22 autosomes and two sex chromosomes (Tables 2, 3, and 4). The 10,850 loci mapped on YACs fall into 653 contigs connecting an average of 17 STSs each before gap closure and 377 contigs with an average of 29 STSs after gap closure. We examined the local density of YAC hits and contigs across the length of each chromosome. The results were relatively similar across the genome, with the notable exception of the chromosomes 1p36, 19, 22, and X. The map has less continuity in these regions, apparently because of systematic underrepresentation in the CEPH Mega-YAC library (see YAC density in Table 4), a problem that has been previously noted (*7*). Chromosome X is underrepresented because the library was made from a male cell line. The autosomal deficits could reflect cloning biases of the yeast host, inasmuch as these are all regions of high GC content (*40*).

The physical map contains a wealth of information, which is ill-suited for presentation in traditional printed form. The complete physical map—including the STS sequences, RH retention patterns, YAC addresses, and order of loci—would require more than 900 journal pages to display. A compressed view of chromosome 14 is shown in Fig. 3, to illustrate the general nature of the map. The complete data for the map can be freely accessed through a World Wide Web server at the Whitehead Institute (http://www-genome.wi.mit.edu/), which includes various tools for analysis.

Table 2. Types of STSs. Chr., chromosome.

Chr.	Total STSs	Random STS*	Genes		Genetic markers		Other loci	ESTs (obs/exp)†
			ESTs	GenBank	Généthon	CHLC		
1	1,374	252	275	106	460	153	128	1.4
2	1,275	307	181	67	452	146	122	0.8
3	1,097	269	181	64	353	134	96	0.9
4	919	210	112	45	281	121	150	0.7
5	858	196	125	30	312	97	98	0.8
6	858	181	114	39	312	108	104	0.8
7	781	168	141	39	272	83	78	1.1
8	739	183	104	35	248	104	65	0.7
9	577	132	106	30	188	68	53	1.1
10	719	154	131	26	281	60	67	1.1
11	706	122	140	42	272	64	66	1.5
12	707	132	104	64	250	91	66	1.0
13	418	102	48	13	164	54	37	0.6
14	489	106	95	27	163	53	45	1.2
15	428	97	97	22	145	30	37	1.3
16	435	87	79	18	180	32	39	1.2
17	447	66	97	39	186	34	25	1.9
18	403	91	46	18	136	64	48	0.7
19	246	23	45	20	121	15	22	2.6
20	386	84	68	26	144	32	32	1.1
21	156	28	18	12	61	13	24	0.8
22	274	19	38	17	67	12	122	2.6
X	587	145	63	28	216	28	107	0.6
Y‡	207	0	0	0	0	0	207	
Total	15,086	3,154	2,408	827	5,264	1,595	1,838	1.0

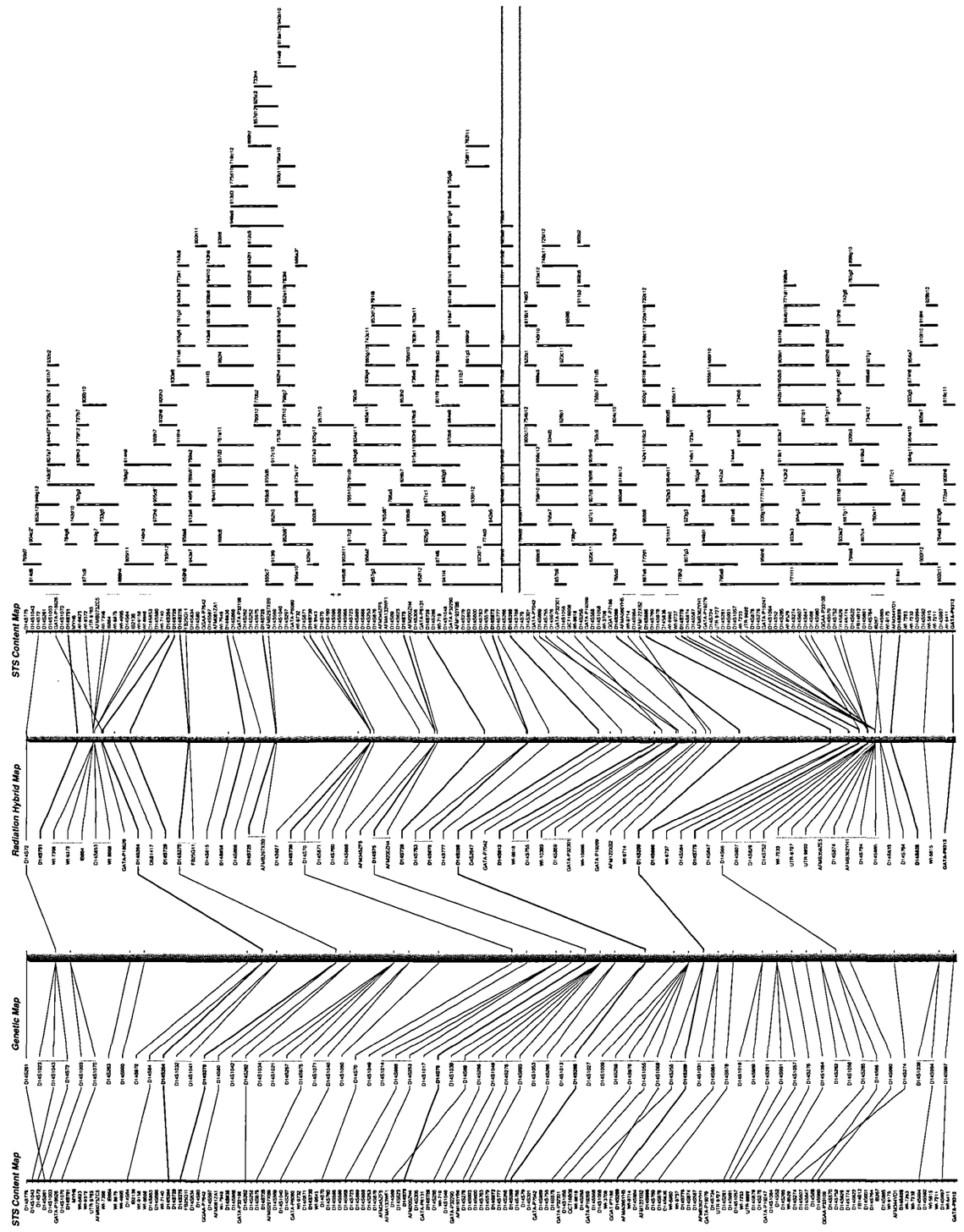
*Unbiased STSs, generated by sequencing from a random genomic library. †Ratio of observed (obs) number of ESTs divided by expected (exp) number, assuming that ESTs follow the same distribution as random STSs. ‡From our previously reported work (*3*).

Coverage

We sought to determine how much of the human genome is covered by the physical map. For this purpose, we derived a new collection of random STSs—by sequencing random clones from an M13 library, selecting PCR primers, and retaining those loci that gave consistent amplification of a single fragment in control experiments. The first 100 STSs produced in this fashion were then screened against the NIGMS 1 hybrid panel, the RH panel, and the YAC library. Because the goal was to obtain an unbiased assessment of coverage, special efforts were made to obtain complete data for each locus.

RH data was obtained for all 100 STSs. (In six cases, it was necessary to resort to acrylamide gel electrophoresis of radioactively labeled products to circumvent problems posed by rodent background.) All 100 loci could be positioned on the RH map with a lod ≥8, on the correct chromosome as determined by the polychromosomal hybrid panel (*41*). The RH map thus appears to cover the vast majority of the human genome.

YAC screening data was also obtained for all 100 STSs. Two STSs detected no YACs in the library, consistent with previous observations that about 2% of DNA sequences appear to be absent from the CEPH Mega-YAC library (*7*). Four STSs detected YAC hits, but none with links to another STS in the correct chromosomal region; these loci



A comprehensive large-insert yeast artificial chromosome library for physical mapping of the mouse genome

Maryann L. Haldi[1], Corinne Strickland[1], Prudence Lim[1], Victor VanBerkel[1], Xiao-Ning Chen[2], David Noya[2], Julie R. Korenberg[2], Zeeshan Husain[1], Joyce Miller[1], Eric S. Lander[1,3]

[1]Whitehead/MIT Center for Genome Research, Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge MA 02142 USA
[2]Ahmanson Department of Pediatrics, Genetics, Cedars-Sinai Research Institute, UCLA, Los Angeles, CA 90048 USA
[3]Department of Biology, Massachusetts Institute of Technology, Cambridge MA 02139 USA

Running head: Construction of YAC library of mouse genome

Correspondence to: Eric S. Lander
Phone: (617) 252-1906
FAX: (617) 252-1933

A yeast artificial chromosome (YAC) library with large insert size and deep coverage is an essential resource for the construction of physical maps of mammalian genomes. Two large-insert YAC libraries of the mouse genome have previously been reported. Larin et al. (1991) constructed a 3-fold coverage library with average insert size of 700 kb, using the mouse strain C3H. Kusumi et al. (1993) constructed a library with the larger-insert portion providing 3.6-fold with an average insert size of 680 kb, using the strain C57Bl/6J. These libraries are excellent resources for positional cloning, but neither is ideal for construction of a physical map of the entire mouse genome.

Here, we report the construction and availability of a mouse YAC library providing roughly 10-fold coverage with an average insert size of 820 kb. The library will be the basis for our current effort to construct a complete physical map, using the mouse genetic map as a scaffold.

The library was constructed with a different YAC vector than the traditionally used vector pYAC4. The vector (Spencer et al. 1993) consists of two arms carried on different plasmids. The pRML1 vector arm carries a *TRP1* selectable marker with a complete promoter element, and the pRML2 arm carries the *URA3* marker. Among its advantages, the vector allows simultaneous selection for both Trp^+ and Ura^+ transformants. By contrast, use of pYAC4 vector requires single selection for Ura^+ followed by screening for Trp^+, as its *TRP1* promoter is weak. The vector also contains T3 and T7 promoters flanking the cloning site to facilitate production of probes from the insert DNA. Additionally, pRML1 carries a yeast centromere with an adjacent *GAL1* promoter and a heterologous thymidine kinase gene. Growth on galactose to inactivate the centromere, plus selection for thymidine kinase expression, increases the copy number of the YAC.

The YAC library was prepared using genomic DNA from C57BL/6J female mice according to Foote (1994) with several modifications. DNA was isolated from kidney nuclei as described by Strauss et al. (1992) and partially digested by *Eco*RI-*Eco*RI methylase competition. The products of this digestion were size-selected by pulsed field gel electrophoresis to be larger than 800 kb, by using conditions under which DNA of this size migrates in the zone of limiting mobility. This DNA was ligated to pRML1 and pRML2 vector arms prepared by digestion with *Not*I and *Eco*RI. The ligation product was again size-selected to be greater than 800 kb. The 1% Seaplaque GTG

(FMC) agarose sizing gels were run in 0.5X TBE buffer on a Bio Rad CHEF apparatus at 14°C, 55-sec pulse time and 6V/cm for 24 h (partial digestion product) or 30 h (ligation product). The size-selected ligation mixture was transformed into the yeast host strain J57D (*ura3-52, trp1 ade2-101 can1-100 leu2-3, 112 his3-6*, a gift from Vladimir Larionov). YACs possessing both vector arms were doubly selected as Trp^+ Ura^+ transformants. Agarase treatment of the ligation product prior to yeast transformation was at 40°-42°C for 45 min and no calf thymus or other carrier DNA was included in the transformation. We have observed that complete digestion by agarase is critical, and the inclusion of carrier DNA reduces transformation efficiency (data not shown).

The quality of the library was monitored during construction in two ways. Initially, the presence of mouse DNA in the transformants was assessed by a rapid PCR assay to detect the B2-repeat element as described in Kusumi et al. (1993). Some clones containing mouse DNA will fail to amplify in this assay, as intact yeast cells rather than purified DNA were used as the test material. Four percent of the transformants from each ligation were screened and if greater than 90% of the clones yielded a PCR product the ligation was judged to be successful. The insert sizes of the YACs were determined by pulsed field gel electrophoresis of yeast DNA prepared in agarose blocks (Gemmill et al. 1994) in 1% Fast Lane Agarose (FMC) gels. The PFGE conditions used were: 14°C, 12 to 160 sec pulse time and 6V/cm for 24hr. The DNA was transfered to a nylon membrane and probed with pUC19 DNA to detect vector sequences. Approximately one percent of the YACs from each ligation were examined. About 20% appeared to be unstable as evidenced by a ladder of smaller bands. Apparent double transformants (only two bands of similar intensity) were present in less than 5% of the clones.

Figure 1 shows the distribution of YAC sizes in the 550 clones tested. The mean size is 820 kb and the median size is 780 kb. Based on the estimated length of the mouse genome as 3 billion bp, the library provides 10-fold coverage. This would correspond to 99.995% coverage assuming no cloning bias. Unfortunately, systematic cloning biases are known to occur in YAC libraries and so the actual coverage will be lower. Nonetheless, we would expect, based on experience with human YAC libraries (Hudson et al., 1995), that the library covers perhaps 98% of the mouse genome.

We tested the library for representation by screening a randomly chosen marker from each chromosome against a subset of the library. Total yeast DNA was prepared according to Gemmill et al. (1994), from 30 pools of YACs, each pool containing 960 clones. In principle, these pools should represent 7.8 fold coverage of the genome. Each pool was screened by a polymerase chain reaction (PCR) assay for a unique STS marker (Green and Olson, 1990). As shown in Table 1, all markers were detected in at least one pool, with an average predicted coverage of 7.2 fold. This result is not statistically different than the expectation of 7.8-fold coverage.

We estimated the chimerism rate of the library, by testing 42 of the YACs by fluorescence in situ hybridization (FISH). Total DNA was prepared from the yeast clones according to Rose et al. (1990). FISH was performed essentially as described by Korenberg and Chen (1995). Mouse chromosomes were prepared from female mouse spleen cells using a modification of the method described by Boyle et al. (1990) and Zhu et al. (1995). Forty metaphase cells were evaluated for each test. A test was scored as positive if there were signals on both chromatids of at least one chromosome in at least 50% of the cells examined. Those scored as chimeric also showed hybridization signals on a second pair of chromosomes in at least 50% of positive cells.

There are two unavoidable sources of error in the estimation of the rate of chimerism by FISH. A small proportion of YACs will cross-hybridize to truly homologous sequences elsewhere in the genome and will thereby artificially elevate the estimation of chimerism. Conversely, chimerism involving a small segment of DNA from a second region may be missed due to the weakness of a signal as well as due to the size limit of detection of the YAC FISH assay. This sensitivity has been estimated as approximately 10% of the total YAC size (Korenberg et al., 1996).

The results of the FISH analysis are shown in Table 2. Positive results by the above criteria were obtained in 88% of the tests (37/42). Of these, 35% (12/37) indicated that the YAC was chimeric. Since some small segments will not be detected by FISH, this estimate of chimerism is minimal. One YAC hybridized to three pairs of chromosomes. This clone appeared to contain a single unstable molecule on pulsed field gel analysis (data not shown). Although the instability of this insert may reflect a region containing localized repetitive sequences present at three different genomic sites, the insert may also contain more than one genomic fragment.

The entire library, containing approximately 38,400 clones, has been distributed to: Research Genetics, Inc. (Huntsville AL), Genome Systems, Inc. (St. Louis MO), Philip Avner (Institute Pasteur, Paris) and Steven Brown (MRC Mouse Genome Center, Harwell, England), to make it widely accessible to the mammalian genetics community.

Acknowledgments. We thank Ndubuisi Azubuine, Gail Farino, Steven Yoo and Jean Whang for excellent technical assistance. This work was supported by a grant from the National Center for Human Genome Research (NCHGR) to E. S. L., and grants to J. R. K. from the Department of Energy and the NCHGR. J. R. K. holds the Geri and Richard Brawerman chair in Molecular Genetics.

References

Boyle, A., Ballard, S., Ward, D. C. (1990). Differential distribution of long and short interspersed element sequences in the mouse genome: Chromosome karyotyping by fluorescence in situ hybridization. Proc. Natl. Acad. Sci. USA 87: 7757-7761.

Foote, S. and Denny, C. (1994). Construction of YAC libraries with large inserts. In *Current Protocols in Human Genetics*, S. Bonitz, ed. (New York: John Wiley & Sons), pp. 5.2.1-5.2.20.

Gemmill, R., Bolin, R., Albertsen, H. (1994). Pulsed-Field Gel Electrophoresis for long-range Restriction mapping. In *Current Protocols in Human Genetics*, S. Bonitz, ed. (New York: John Wiley & Sons), pp. 5.1.1-5.1.24.

Green, E. and Olson, M. (1990). Systematic screening of yeast artificial chromosome libraries by use of the polymerase chain reaction. Proc. Natl. Acad. Sci. USA 87, 1213-1217.

Hudson, T., Stein, L., Gerety, S., Ma, J., Castle, A., Silva, J., Slonim, D., Baptista, R., Kruglyak, L., Xu, S., Hu, X., Colbert, A., Rosenberg, C., Reeve-Daly, M., Rozen, S., Hui, L., Wu, X., Vestergaard, C., Wilson, K., Bae, J., Maitra, S., Ganiatsas, S., Evans, C., DeAngelis, M., Ingalls, K., Nahf, R., Horton, L., Anderson, M., Collymore, A., Ye, W., Kouyoumjian, V., Zemsteva, I., Tam, J., Devine, R., Courtney, D., Renaud, M., Nguyen, H., O'Connor, T., Fizames, C., Faure, S., Gyapay, G., Dib, C., Morissette, J., Orlin, J., Birren, B., Goodman, N., Weissenbach, J., Hawkins, T., Foote, S., Page, D., Lander, E. (1995). An STS-based map of the human genome. Science 270, 1945-1954.

Korenberg, J.R., Chen, X-N. (1995). Human cDNA mapping using a high resolution R-banding technique and fluorescence in situ hybridization. Cytogenetic and Cell Genetics 69, 196-200.

Korenberg, J.R., Chen, X.-N., Mitchell, S., Fannin, S., Gerwehr, S., Cohen, D., Chumakov, I. (1996). A high-fidelity physical map of human chromosome 21q in yeast artificial chromosomes. Genome Research 5, 427-443.

Kusumi, K., Smith, J., Segre, J., Koos, D., Lander, E. (1993) Construction of a large-insert yeast artificial chromosome library of the mouse genome. Mammalian Genome 4, 391-392.

Larin, Z., Monaco, A., Lehrach, H. (1991). Yeast artificial chromosome libraries containing large inserts from mouse and human DNA. Proc. Natl. Acad. Sci. USA 88, 4123-4127.

Rose, M., Winston, F., Hieter, P. (1990). Methods in Yeast Genetics. A Laboratory Course Manual. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press), pp. 128-129.

Spencer, F., Ketner, G., Connelly, C., Hieter, P. (1993). Targeted recombination-based cloning and manipulation of large DNA segments in yeast. Methods: A companion to Methods in Enzymology 5, 161-175.

Strauss, W., Jaenisch, E., Jaenisch, R. (1992). A strategy for rapid production and screening of yeast artificial chromosome libraries. Mammalian Genome 150-157.

Zhu, Y., Qi, C., Korenberg, J.R., Chen X-N., Noya, D., Rao, M., Reddy, J. (1995). Structural organization of mouse peroxisome proliferator-activated receptor g (mPPARg) gene: Alternative promoter use and different splicing yield two mPPARg isoforms. Proc. Natl. Acad. Sci. USA 92: 7921-7925.

Figure Legend

Fig. 1. Histogram showing the sizes of 550 randomly selected YAC clones.

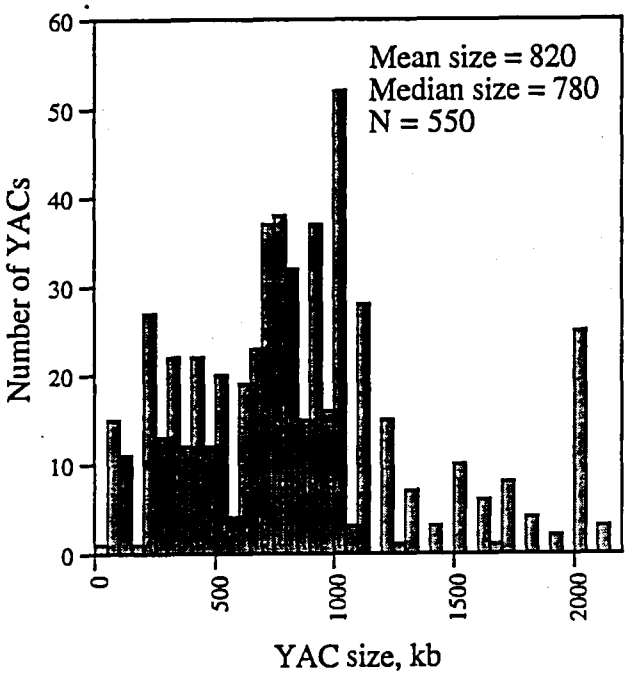

Mean size = 820
Median size = 780
N = 550
Number of YACs
YAC size, kb
0
10
20
30
40
50
60
0
500
1000
1500
2000

Table 1. Detection of chromosomal markers in YAC pools for subset of library predicted to provide 7.8-fold coverage of genome.

Marker	Chromosome	Number of positive pools
D1Mit464	1	7
D2Mit104	2	4
D3Mit60	3	11
D4Mit182	4	11
MPC1896	5	9
D6Mit133	6	11
D7Mit270	7	7
D8Mit64	8	7
D9Mit227	9	1
D10Mit152	10	9
D11Mit173	11	8
D12Mit37	12	7
D13Mit78	13	7
D14Mit80	14	5
D15Mit56	15	7
D16Mit138	16	3
D17Mit177	17	3
D18Mit177	18	5
D19Mit36	19	12
DXMit166	X	10
Average		7.2

Table 2. Analysis of chimerism in YACs by FISH.

YAC clone	size (kb)	chimeric
02-1	945kb	no
02-9	1250	no
02-10	610	no
02-20	1100	no
02-23	1100	no
02-24	1050	yes
02-26	1200	no
02-27	610	yes
02-28	1100	yes
02-29	1050	yes
04-17	1800	yes
04-19	1100	yes
04-22	750	yes
04-24	750	yes
04-25	750	no
05-4	750	no
05-5	2000 (unstable)	yes (three signals)
05-7	750	no
360A1	1200	no
360A9	915	no
360A10	945	no
361A8	1300	yes
362A1	920	no
362A2	920	no
362A8	1100	no
362A9	980	no
362A10	870	no
387A1	920	no
387A2	680	no
387A9	1200	no
387A11	600	yes
388A7	1000	no
388A8	730	yes

(continuation of Table 2)

388A10	920	no
397A8	1000	no
405A12	900	no
407A1	915	no

A comprehensive large-insert yeast artificial chromosome library for physical mapping of the mouse genome

Maryann L. Haldi[1], Corinne Strickland[1], Prudence Lim[1], Victor VanBerkel[1], Xiao-Ning Chen[2], David Noya[2], Julie R. Korenberg[2], Zeeshan Husain[1], Joyce Miller[1], Eric S. Lander[1,3]

[1]Whitehead/MIT Center for Genome Research, Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge MA 02142 USA
[2]Ahmanson Department of Pediatrics, Genetics, Cedars-Sinai Research Institute, UCLA, Los Angeles, CA 90048 USA
[3]Department of Biology, Massachusetts Institute of Technology, Cambridge MA 02139 USA

Running head: Construction of YAC library of mouse genome

Correspondence to: Eric S. Lander
Phone: (617) 252-1906
FAX: (617) 252-1933

A yeast artificial chromosome (YAC) library with large insert size and deep coverage is an essential resource for the construction of physical maps of mammalian genomes. Two large-insert YAC libraries of the mouse genome have previously been reported. Larin et al. (1991) constructed a 3-fold coverage library with average insert size of 700 kb, using the mouse strain C3H. Kusumi et al. (1993) constructed a library with the larger-insert portion providing 3.6-fold with an average insert size of 680 kb, using the strain C57Bl/6J. These libraries are excellent resources for positional cloning, but neither is ideal for construction of a physical map of the entire mouse genome.

Here, we report the construction and availability of a mouse YAC library providing roughly 10-fold coverage with an average insert size of 820 kb. The library will be the basis for our current effort to construct a complete physical map, using the mouse genetic map as a scaffold.

The library was constructed with a different YAC vector than the traditionally used vector pYAC4. The vector (Spencer et al. 1993) consists of two arms carried on different plasmids. The pRML1 vector arm carries a *TRP1* selectable marker with a complete promoter element, and the pRML2 arm carries the *URA3* marker. Among its advantages, the vector allows simultaneous selection for both Trp^+ and Ura^+ transformants. By contrast, use of pYAC4 vector requires single selection for Ura^+ followed by screening for Trp^+, as its *TRP1* promoter is weak. The vector also contains T3 and T7 promoters flanking the cloning site to facilitate production of probes from the insert DNA. Additionally, pRML1 carries a yeast centromere with an adjacent *GAL1* promoter and a heterologous thymidine kinase gene. Growth on galactose to inactivate the centromere, plus selection for thymidine kinase expression, increases the copy number of the YAC.

The YAC library was prepared using genomic DNA from C57BL/6J female mice according to Foote (1994) with several modifications. DNA was isolated from kidney nuclei as described by Strauss et al. (1992) and partially digested by *Eco*RI-*Eco*RI methylase competition. The products of this digestion were size-selected by pulsed field gel electrophoresis to be larger than 800 kb, by using conditions under which DNA of this size migrates in the zone of limiting mobility. This DNA was ligated to pRML1 and pRML2 vector arms prepared by digestion with *Not*I and *Eco*RI. The ligation product was again size-selected to be greater than 800 kb. The 1% Seaplaque GTG

(FMC) agarose sizing gels were run in 0.5X TBE buffer on a Bio Rad CHEF apparatus at 14°C, 55-sec pulse time and 6V/cm for 24 h (partial digestion product) or 30 h (ligation product). The size-selected ligation mixture was transformed into the yeast host strain J57D (*ura3-52, trp1 ade2-101 can1-100 leu2-3, 112 his3-6*, a gift from Vladimir Larionov). YACs possessing both vector arms were doubly selected as Trp^+ Ura^+ transformants. Agarase treatment of the ligation product prior to yeast transformation was at 40°-42°C for 45 min and no calf thymus or other carrier DNA was included in the transformation. We have observed that complete digestion by agarase is critical, and the inclusion of carrier DNA reduces transformation efficiency (data not shown).

The quality of the library was monitored during construction in two ways. Initially, the presence of mouse DNA in the transformants was assessed by a rapid PCR assay to detect the B2-repeat element as described in Kusumi et al. (1993). Some clones containing mouse DNA will fail to amplify in this assay, as intact yeast cells rather than purified DNA were used as the test material. Four percent of the transformants from each ligation were screened and if greater than 90% of the clones yielded a PCR product the ligation was judged to be successful. The insert sizes of the YACs were determined by pulsed field gel electrophoresis of yeast DNA prepared in agarose blocks (Gemmill et al. 1994) in 1% Fast Lane Agarose (FMC) gels. The PFGE conditions used were: 14°C, 12 to 160 sec pulse time and 6V/cm for 24hr. The DNA was transfered to a nylon membrane and probed with pUC19 DNA to detect vector sequences. Approximately one percent of the YACs from each ligation were examined. About 20% appeared to be unstable as evidenced by a ladder of smaller bands. Apparent double transformants (only two bands of similar intensity) were present in less than 5% of the clones.

Figure 1 shows the distribution of YAC sizes in the 550 clones tested. The mean size is 820 kb and the median size is 780 kb. Based on the estimated length of the mouse genome as 3 billion bp, the library provides 10-fold coverage. This would correspond to 99.995% coverage assuming no cloning bias. Unfortunately, systematic cloning biases are known to occur in YAC libraries and so the actual coverage will be lower. Nonetheless, we would expect, based on experience with human YAC libraries (Hudson et al., 1995), that the library covers perhaps 98% of the mouse genome.

We tested the library for representation by screening a randomly chosen marker from each chromosome against a subset of the library. Total yeast DNA was prepared according to Gemmill et al. (1994), from 30 pools of YACs, each pool containing 960 clones. In principle, these pools should represent 7.8 fold coverage of the genome. Each pool was screened by a polymerase chain reaction (PCR) assay for a unique STS marker (Green and Olson, 1990). As shown in Table 1, all markers were detected in at least one pool, with an average predicted coverage of 7.2 fold. This result is not statistically different than the expectation of 7.8-fold coverage.

We estimated the chimerism rate of the library, by testing 42 of the YACs by fluorescence in situ hybridization (FISH). Total DNA was prepared from the yeast clones according to Rose et al. (1990). FISH was performed essentially as described by Korenberg and Chen (1995). Mouse chromosomes were prepared from female mouse spleen cells using a modification of the method described by Boyle et al. (1990) and Zhu et al. (1995). Forty metaphase cells were evaluated for each test. A test was scored as positive if there were signals on both chromatids of at least one chromosome in at least 50% of the cells examined. Those scored as chimeric also showed hybridization signals on a second pair of chromosomes in at least 50% of positive cells.

There are two unavoidable sources of error in the estimation of the rate of chimerism by FISH. A small proportion of YACs will cross-hybridize to truly homologous sequences elsewhere in the genome and will thereby artificially elevate the estimation of chimerism. Conversely, chimerism involving a small segment of DNA from a second region may be missed due to the weakness of a signal as well as due to the size limit of detection of the YAC FISH assay. This sensitivity has been estimated as approximately 10% of the total YAC size (Korenberg et al., 1996).

The results of the FISH analysis are shown in Table 2. Positive results by the above criteria were obtained in 88% of the tests (37/42). Of these, 35% (12/37) indicated that the YAC was chimeric. Since some small segments will not be detected by FISH, this estimate of chimerism is minimal. One YAC hybridized to three pairs of chromosomes. This clone appeared to contain a single unstable molecule on pulsed field gel analysis (data not shown). Although the instability of this insert may reflect a region containing localized repetitive sequences present at three different genomic sites, the insert may also contain more than one genomic fragment.

The entire library, containing approximately 38,400 clones, has been distributed to: Research Genetics, Inc. (Huntsville AL), Genome Systems, Inc. (St. Louis MO), Philip Avner (Institute Pasteur, Paris) and Steven Brown (MRC Mouse Genome Center, Harwell, England), to make it widely accessible to the mammalian genetics community.

Acknowledgments. We thank Ndubuisi Azubuine, Gail Farino, Steven Yoo and Jean Whang for excellent technical assistance. This work was supported by a grant from the National Center for Human Genome Research (NCHGR) to E. S. L., and grants to J. R. K. from the Department of Energy and the NCHGR. J. R. K. holds the Geri and Richard Brawerman chair in Molecular Genetics.

References

Boyle, A., Ballard, S., Ward, D. C. (1990). Differential distribution of long and short interspersed element sequences in the mouse genome: Chromosome karyotyping by fluorescence in situ hybridization. Proc. Natl. Acad. Sci. USA 87: 7757-7761.

Foote, S. and Denny, C. (1994). Construction of YAC libraries with large inserts. In *Current Protocols in Human Genetics*, S. Bonitz, ed. (New York: John Wiley & Sons), pp. 5.2.1-5.2.20.

Gemmill, R., Bolin, R., Albertsen, H. (1994). Pulsed-Field Gel Electrophoresis for long-range Restriction mapping. In *Current Protocols in Human Genetics*, S. Bonitz, ed. (New York: John Wiley & Sons), pp. 5.1.1-5.1.24.

Green, E. and Olson, M. (1990). Systematic screening of yeast artificial chromosome libraries by use of the polymerase chain reaction. Proc. Natl. Acad. Sci. USA 87, 1213-1217.

Hudson, T., Stein, L., Gerety, S., Ma, J., Castle, A., Silva, J., Slonim, D., Baptista, R., Kruglyak, L., Xu, S., Hu, X., Colbert, A., Rosenberg, C., Reeve-Daly, M., Rozen, S., Hui, L., Wu, X., Vestergaard, C., Wilson, K., Bae, J., Maitra, S., Ganiatsas, S., Evans, C., DeAngelis, M., Ingalls, K., Nahf, R., Horton, L., Anderson, M., Collymore, A., Ye, W., Kouyoumjian, V., Zemsteva, I., Tam, J., Devine, R., Courtney, D., Renaud, M., Nguyen, H., O'Connor, T., Fizames, C., Faure, S., Gyapay, G., Dib, C., Morissette, J., Orlin, J., Birren, B., Goodman, N., Weissenbach, J., Hawkins, T., Foote, S., Page, D., Lander, E. (1995). An STS-based map of the human genome. Science 270, 1945-1954.

Korenberg, J.R., Chen, X-N. (1995). Human cDNA mapping using a high resolution R-banding technique and fluorescence in situ hybridization. Cytogenetic and Cell Genetics 69, 196-200.

Korenberg, J.R., Chen, X.-N., Mitchell, S., Fannin, S., Gerwehr, S., Cohen, D., Chumakov, I. (1996). A high-fidelity physical map of human chromosome 21q in yeast artificial chromosomes. Genome Research 5, 427-443.

Kusumi, K., Smith, J., Segre, J., Koos, D., Lander, E. (1993) Construction of a large-insert yeast artificial chromosome library of the mouse genome. Mammalian Genome 4, 391-392.

Larin, Z., Monaco, A., Lehrach, H. (1991). Yeast artificial chromosome libraries containing large inserts from mouse and human DNA. Proc. Natl. Acad. Sci. USA 88, 4123-4127.

Rose, M., Winston, F., Hieter, P. (1990). Methods in Yeast Genetics. A Laboratory Course Manual. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press), pp. 128-129.

Spencer, F., Ketner, G., Connelly, C., Hieter, P. (1993). Targeted recombination-based cloning and manipulation of large DNA segments in yeast. Methods: A companion to Methods in Enzymology 5, 161-175.

Strauss, W., Jaenisch, E., Jaenisch, R. (1992). A strategy for rapid production and screening of yeast artificial chromosome libraries. Mammalian Genome 150-157.

Zhu, Y., Qi, C., Korenberg, J.R., Chen X-N., Noya, D., Rao, M., Reddy, J. (1995). Structural organization of mouse peroxisome proliferator-activated receptor g (mPPARg) gene: Alternative promoter use and different splicing yield two mPPARg isoforms. Proc. Natl. Acad. Sci. USA 92: 7921-7925.

Figure Legend

Fig. 1. Histogram showing the sizes of 550 randomly selected YAC clones.

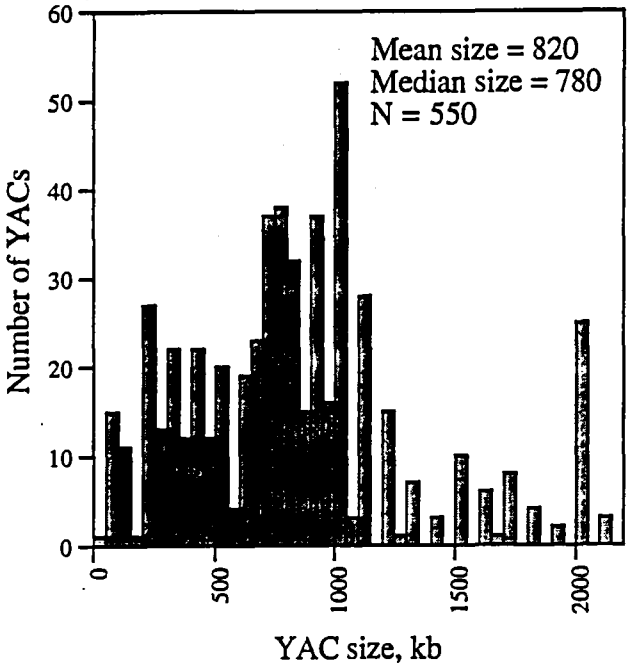

Mean size = 820
Median size = 780
N = 550
Number of YACs
YAC size, kb
0
10
20
30
40
50
60
0
500
1000
1500
2000

Table 1. Detection of chromosomal markers in YAC pools for subset of library predicted to provide 7.8-fold coverage of genome.

Marker	Chromosome	Number of positive pools
D1Mit464	1	7
D2Mit104	2	4
D3Mit60	3	11
D4Mit182	4	11
MPC1896	5	9
D6Mit133	6	11
D7Mit270	7	7
D8Mit64	8	7
D9Mit227	9	1
D10Mit152	10	9
D11Mit173	11	8
D12Mit37	12	7
D13Mit78	13	7
D14Mit80	14	5
D15Mit56	15	7
D16Mit138	16	3
D17Mit177	17	3
D18Mit177	18	5
D19Mit36	19	12
DXMit166	X	10
Average		7.2

Table 2. Analysis of chimerism in YACs by FISH.

YAC clone	size (kb)	chimeric
02-1	945kb	no
02-9	1250	no
02-10	610	no
02-20	1100	no
02-23	1100	no
02-24	1050	yes
02-26	1200	no
02-27	610	yes
02-28	1100	yes
02-29	1050	yes
04-17	1800	yes
04-19	1100	yes
04-22	750	yes
04-24	750	yes
04-25	750	no
05-4	750	no
05-5	2000 (unstable)	yes (three signals)
05-7	750	no
360A1	1200	no
360A9	915	no
360A10	945	no
361A8	1300	yes
362A1	920	no
362A2	920	no
362A8	1100	no
362A9	980	no
362A10	870	no
387A1	920	no
387A2	680	no
387A9	1200	no
387A11	600	yes
388A7	1000	no
388A8	730	yes

A comprehensive genetic map of the mouse genome

William F. Dietrich*, Joyce Miller*, Robert Steen*, Mark A. Merchant*, Deborah Damron-Boles*, Zeeshan Husain*, Robert Dredge*, Mark J. Daly*, Kimberly A. Ingalls*, Tara J. O'Connor*, Cheryl A. Evans*, Margaret M. DeAngelis*, David M. Levinson*, Leonid Kruglyak*, Nathan Goodman*, Neal G. Copeland†, Nancy A. Jenkins†, Trevor L. Hawkins*, Lincoln Stein*, David C. Page*‡§ & Eric S. Lander*‡||

* Whitehead/MIT Center for Genome Research, Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA
† Mammalian Genetics Laboratory, ABL-Basic Research Program, NCI-Frederick Cancer Research and Development Center, Frederick, Maryland 21702, USA
‡ Department of Biology, and § Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA
|| To whom correspondence should be addressed.

The availability of dense genetic linkage maps of mammalian genomes makes feasible a wide range of studies, including positional cloning of monogenic traits, genetic dissection of polygenic traits, construction of genome-wide physical maps, rapid marker-assisted construction of congenic strains, and evolutionary comparisons[1,2]. We have been engaged for the past five years in a concerted effort to produce a dense genetic map of the laboratory mouse[3-5]. Here we present the final report of this project. The map contains 7,377 genetic markers, consisting of 6,580 highly informative simple sequence length polymorphisms integrated with 797 restriction fragment length polymorphisms in mouse genes. The average spacing between markers is about 0.2 centimorgans or 400 kilobases.

To construct a simple sequence length polymorphism (SSLP) map, we identified more than 9,000 sequences from random genomic clones and public databases containing simple sequence repeats (mostly, $(CA)_n$-repeats), designed polymerase chain reaction (PCR) primers flanking the repeat, and tested each for polymorphism by measuring the allele sizes in 12 inbred mouse strains. Of the successful PCR assays, we genotyped the 90% of loci that revealed different alleles between the OB and CAST strains in an (OB × CAST) F_2 intercross with 46 progeny. These data were assembled into a map by performing genetic linkage analysis with the MAPMAKER computer package[7,8].

A total of 6,336 SSLP loci were scored in the F_2 intercross, with 6,111 derived from anonymous sequence and 225 from known genes (Table 1). Of these, 5,905 were scored as codominant markers and 431 as dominant markers (because the pattern of one allele obscured the other). The map provides dense coverage of all 20 mouse chromosomes, with a total genetic length of 1,361 centimorgans (cM). Because the cross involves 92 meioses, the mean spacing between crossovers is 1.1 cM and thus loci can be mapped to 'bins' of this average size. The map has 1,001 occupied bins (Table 3(*a*)), with an average of 6.3 markers per bin and an average spacing of 1.36 cM between consecutive bins.

We next sought to integrate the map of largely anonymous SSLPs with the locations of known genes, because this information can suggest candidates for the genes underlying mouse mutations. We analysed a (B6 × SPRET) backcross that has been extensively used for restriction fragment length polymorphism (RFLP) mapping[9-11]. The backcross has been genotyped for 797 RFLPs. To integrate the maps, we genotyped 1,245 SSLPs from our map in 46 progeny from the SPRET backcross, providing a common reference point approximately every 1.1 cM. We also genotyped 244 additional SSLPs that were not polymorphic—and thus could not be mapped—in the (OB × CAST) intercross, but were polymorphic in the (B6 × SPRET) backcross. The SPRET cross was thus scored for a total of 1,543 SSLPs and 797 RFLPs.

The final map with 7,377 loci is shown in Fig. 1, with the SSLP map on the right and the integration with the RFLP map on the left. A full description of the markers—including primer sequences, locus sequence, genotypes in each cross, and allele

TABLE 1 Genetic markers, genetic length and polymorphism* by chromosome

Chromosome	No. of markers	No. of random markers	No. from GENBANK	'Consensus' genetic length†	Observed genetic length‡	Polymorphism among lab strains (%)§‖	Lab strains versus SPR or CAST (%)‖
1	511	494	17	98	109.9	57	92
2	507	491	16	107	95.7	49	94
3	343	332	11	100	67.5	51	95
4	350	342	8	81	74.2	51	93
5	402	391	11	93	82.9	48	95
6	368	349	19	74	59.1	46	94
7	357	341	16	89	59.8	48	94
8	350	345	5	81	72.0	44	94
9	336	318	18	70	62.9	52	95
10	293	286	7	78	73.0	35	96
11	350	326	24	78	82.0	53	94
12	278	268	10	68	61.5	50	94
13	303	296	7	72	60.2	48	95
14	259	246	13	53	65.6	49	94
15	264	257	7	62	62.2	51	94
16	215	214	1	59	51.0	43	94
17	255	239	16	53	51.0	56	93
18	231	226	5	57	39.7	53	95
19	134	131	3	42	57.2	52	93
X	230	219	11	88	73.5	33	95
Total	6,336	6,111	225	1,503	1,360.9¶	48	94

* Polymorphism survey was based on visual comparisons of fragments across large acrylamide gels and was thus subject to mobility differences among lanes. To assess the accuracy of data in our database, 3,000 individual pairwise comparisons were repeated. Some 6% of reported polymorphic pairs turn out to be monomorphic upon careful comparison, while 4% of reported monomorphic pairs turn out to be polymorphic. The data are thus accurate enough to allow selection of markers for crosses, but geneticists wishing to know every polymorphic marker in a narrow region (for fine-structure genetic mapping and positional cloning, for example) are advised to recheck each locus.
† Based on 'consensus' genetic map in Encyclopedia of the Mouse Genome, http://www.informatics.jax.org.encyclo.html (1993).
‡ Distance between most proximal and most distal markers in the map reported here.
§ Pairwise comparisons of OB, B6, DBA, A, C3H, BALB, AKR, NON, NOD and LP.
‖ Standard error of the mean for each chromosome depends on number of markers studied, but is < 1% in all cases.
¶ Distance is shorter than in previously published versions of this map (ref. 6) because final error checking reduced the number of apparent crossovers.

(Chromosome 19 shows a slight deficit, which is not statistically significant after correction for multiple testing; it may reflect the unusually large proportion of heterochromatin on this chromosome.) In contrast, chromosome X shows a clear deficit, with only about 57% as many as expected (Table 2). This phenomenon appears to be general in mammalian genomes, as we have also found a similar deficit in an SSLP map of the rat[12] (62% of expectation), and Weissenbach and colleagues report a slightly less pronounced deficit in the human genome[13] (75% of expectation). In principle, the deficit of SSLPs on chromosome X could occur if $(CA)_n$-repeats were either less frequent on chromosome X, or were equally frequent but less polymorphic. The latter hypothesis would predict that the deficit of polymorphic loci on chromosome X would be offset by a great excess of non-polymorphic repeats. Of the SSLPs monomorphic between OB and CAST, 37% would have to lie to chromosome X to explain the observed data. We determined the chromosomal location of > 100 monomorphic loci (by genetic mapping for those that were polymorphic between B6 and SPRET and by somatic cell hybrid mapping for those that were not), but we found no significant excess on chromosome X. Accordingly, the deficit appears to be primarily due to an actual shortage of $(CA)_n$-repeats on chromosome X.

The SSLPs show a polymorphism rate of about 50% among inbred laboratory strains surveyed and about 95% between laboratory strains and CAST or SPR (Table 1). The pairwise polymorphism rates among the 12 strains surveyed have not changed significantly from our previous report[6] and are not presented here. Interestingly, the distribution of polymorphism across the genome is not uniform[11]. The average polymorphism rate among the *Mus musculus* strains surveyed was just under 50%, but two chromosomes showed substantially lower polymorphism rates: chromosome X at 33%, and chromosome 10 at 35% (Table 1). Decreased polymorphism could reflect recent selection for specific ancestral chromosomes. For the X chromosome, it could also reflect a different mutation rate (inasmuch as each chromosome X resides in males only two-thirds as often each autosome, and most mutations are thought to occur in male germline) or different population genetic forces (with hemizygosity affecting selection and effective population size).

Our mouse genetic-mapping project is now at its conclusion. Although more SSLPs remain to be found (newly

TABLE 3 Clusters of consecutive crossovers and markers

(a) Number of crossovers between consecutive random markers*

No. of crossovers per interval	Observed No.	Observed (percentage)	Expected† No.	Expected† (percentage)	P(longest run ≥ n) (%)†
0	5,095	(83.85)	5,035.5 ± 29.6	(82.59)	
1	784	(12.90)	876.7 ± 27.4	(14.38)	100.0
2	151	(2.49)	152.6 ± 12.2	(2.50)	100.0
3	27	(0.44)	26.6 ± 5.1	(0.44)	100.0
4	14	(0.23)	4.6 ± 2.2	(0.08)	99.6
5	4	(0.07)	0.8 ± 0.9	(0.01)	62.2
6	0	(0.00)	0.1 ± 0.4	(< 0.01)	15.6
7	0	(0.00)	0.0 ± 0.2	(< 0.01)	2.9
8	1	(0.02)	0.0 ± 0.1	(< 0.01)	0.5
Total	6,076				

(b) Random markers occuring between consecutive crossovers‡

No. of markers per block	Observed No.	Observed (percentage)	Expected§ Number	Expected§ (percentage)	P(longest run ≥ n) (%)§
0	288	(22.3)	227.9 ± 13.7	(17.4)	100.0
1	208	(16.1)	188.2 ± 12.7	(14.4)	100.0
2	126	(9.8)	155.5 ± 11.7	(11.9)	100.0
3	111	(8.6)	128.4 ± 10.8	(9.8)	100.0
4	84	(6.5)	106.0 ± 9.9	(8.1)	100.0
5	73	(5.7)	87.6 ± 9.0	(6.7)	100.0
6	62	(4.8)	72.3 ± 8.3	(5.5)	100.0
7	51	(4.0)	59.7 ± 7.6	(4.6)	100.0
8	36	(2.8)	49.3 ± 6.9	(3.8)	100.0
9	38	(2.9)	40.7 ± 6.3	(3.1)	100.0
10	32	(2.5)	33.7 ± 5.7	(2.6)	100.0
11	37	(2.9)	27.8 ± 5.2	(2.1)	100.0
12	19	(1.5)	23.0 ± 4.7	(1.8)	100.0
13	28	(2.2)	19.0 ± 4.3	(1.4)	100.0
14	18	(1.4)	15.7 ± 3.9	(1.2)	100.0
15	7	(0.5)	12.9 ± 3.6	(1.0)	100.0
16	12	(0.9)	10.7 ± 3.3	(0.8)	100.0
17	5	(0.4)	8.8 ± 3.0	(0.7)	100.0
18	5	(0.4)	7.3 ± 2.7	(0.6)	100.0
19	6	(0.5)	6.0 ± 2.4	(0.5)	100.0
20	10	(0.8)	5.0 ± 2.2	(0.4)	100.0
21	3	(0.2)	4.1 ± 2.0	(0.3)	100.0
22	5	(0.4)	3.4 ± 1.8	(0.3)	100.0
23	7	(0.5)	2.8 ± 1.7	(0.2)	100.0
24	4	(0.3)	2.3 ± 1.5	(0.2)	100.0
25	0	(0.0)	1.9 ± 1.4	(0.1)	100.0
26	5	(0.4)	1.6 ± 1.3	(0.1)	99.9
27	1	(0.1)	1.3 ± 1.1	(0.1)	99.8
28	1	(0.1)	1.1 ± 1.0	(0.1)	99.3
29	0	(0.0)	0.9 ± 0.9	(0.1)	98.4
30	1	(0.1)	0.7 ± 0.9	(0.1)	96.7
31	1	(0.1)	0.6 ± 0.8	(< 0.1)	94.0
32	0	(0.0)	0.5 ± 0.7	(< 0.1)	90.3
33	0	(0.0)	0.4 ± 0.6	(< 0.1)	85.4
34	1	(0.1)	0.3 ± 0.6	(< 0.1)	79.6
35	1	(0.1)	0.3 ± 0.5	(< 0.1)	73.1
38	1	(0.1)	0.2 ± 0.4	(< 0.1)	52.2
40	1	(0.1)	0.1 ± 0.3	(< 0.1)	39.6
54	1	(0.1)	< 0.1 ± 0.1	(< 0.1)	3.4
Total	1,289				

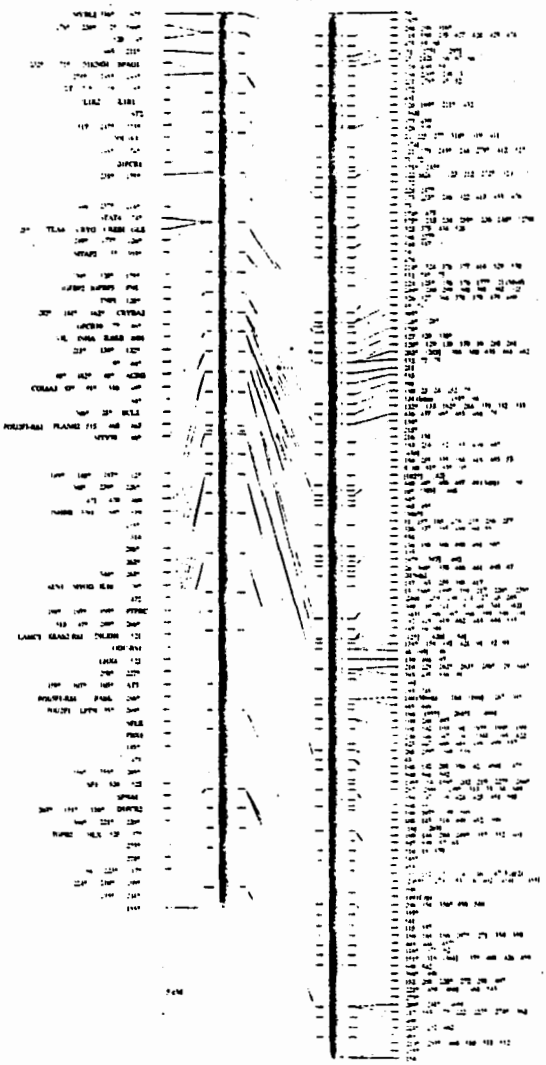
*The intervals with ≥ 1 crossover represent the 981 gaps between consecutive bins of recombinationally inseparable markers. Only random markers are considered to avoid biases in distribution of known genes.

†The probability of the longest run is calculated in ref. 6. Briefly, if a coin with heads probability P is tossed n times, the length R_n of the longest head run has expected value $\mu = \log_{1/p}[(n-1)(1-p)+1]$ and the distribution of R_n is given approximately by $\text{Prob}(R_n - \mu > t) = 1 - \exp(-p^t)$. In this case, $p = 0.17$.

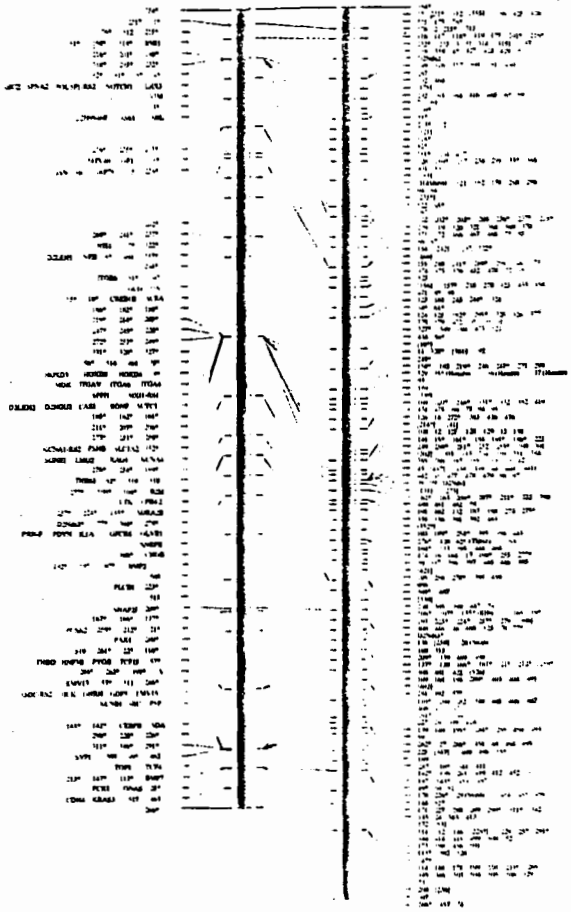
‡The blocks with ≥ 1 marker represent the 1,001 bins of recombinationally separable markers. Only random markers are considered to avoid biases in distribution of known genes.

§The probability of the longest head run is calculated with $p = 0.83$.

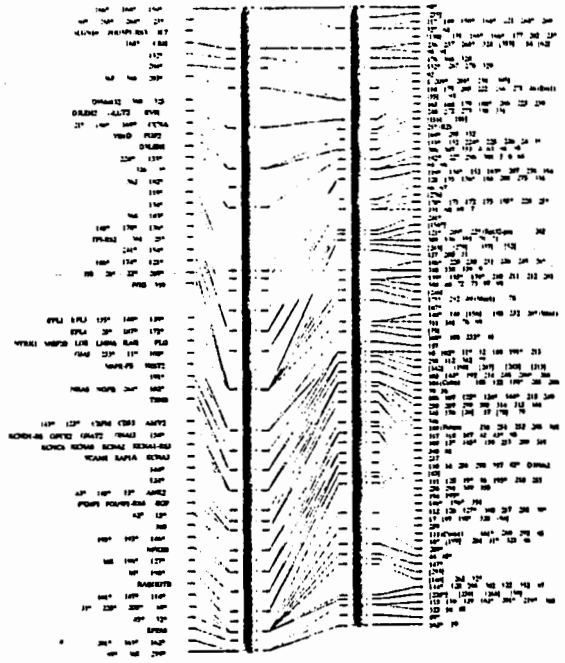
Chromosome 1



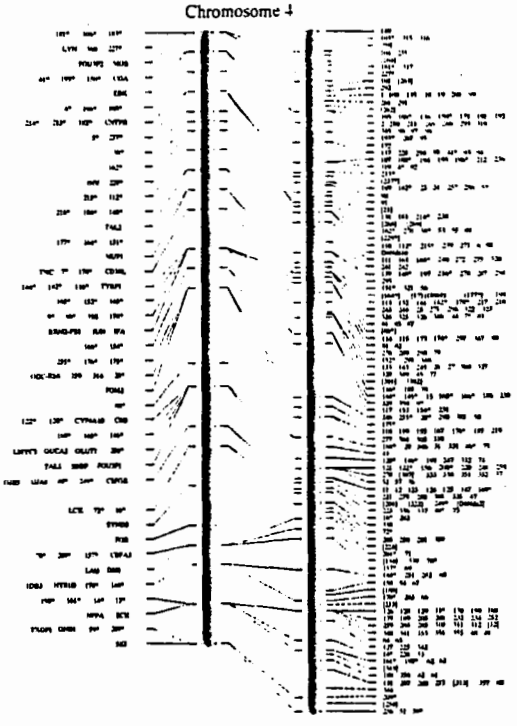
Chromosome 2



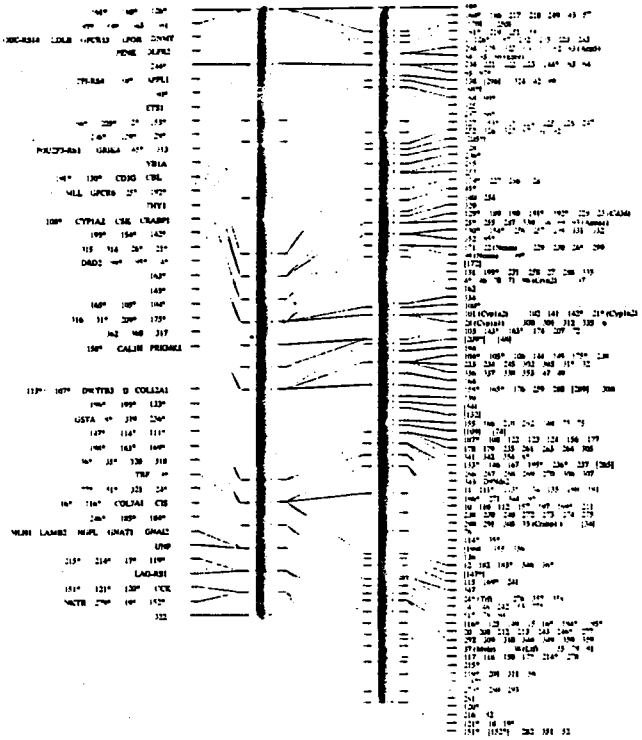
Chromosome 3



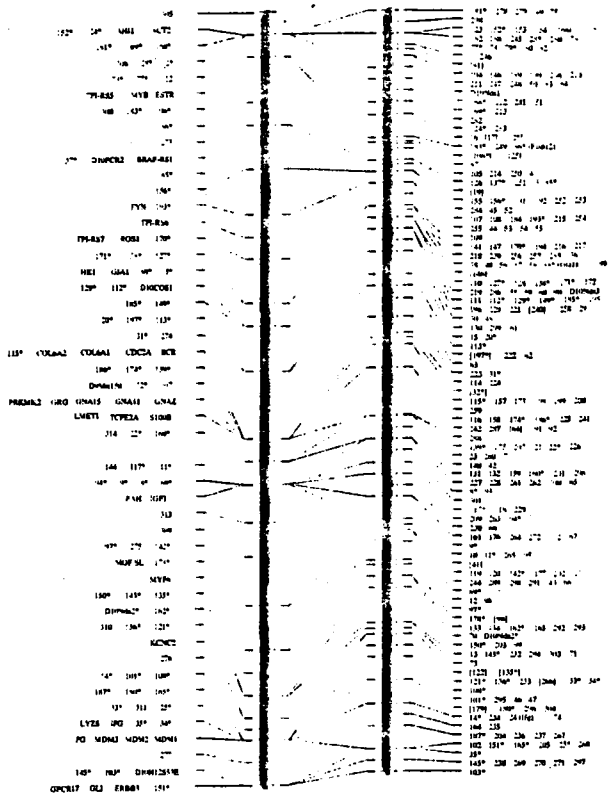
Chromosome 4



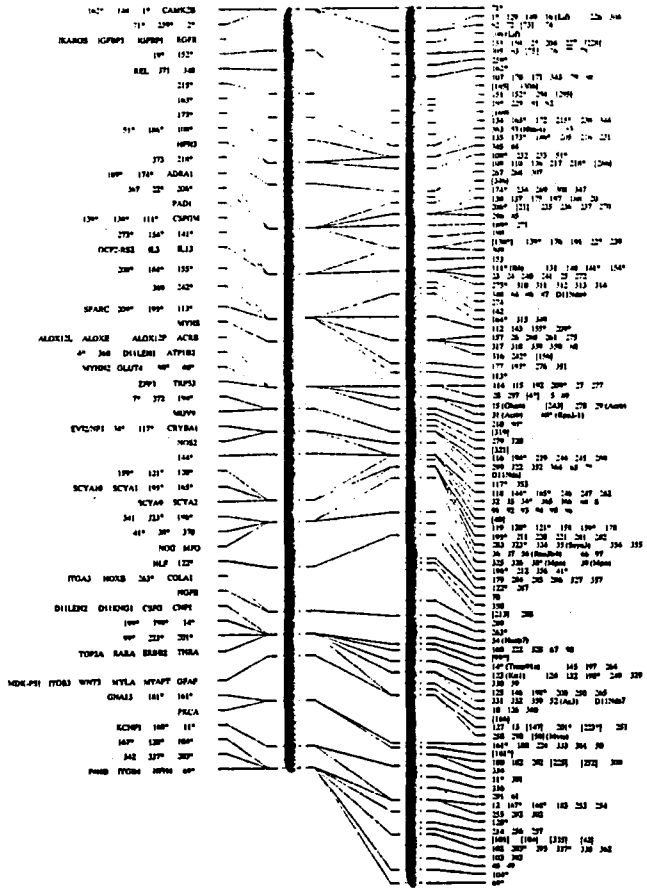
Chromosome 9



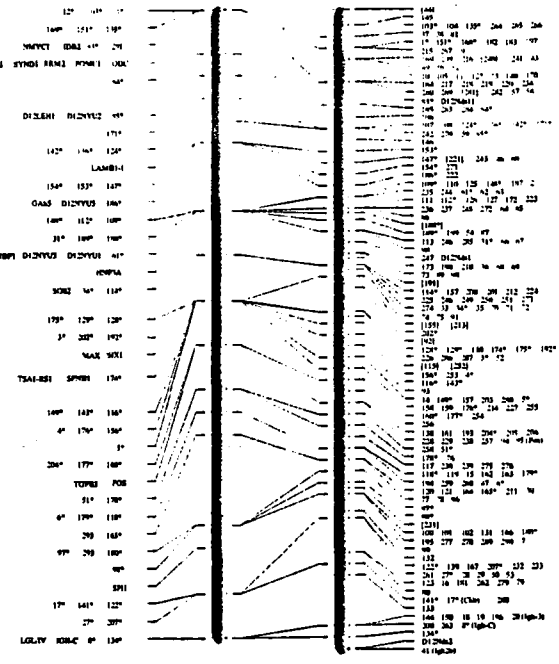
Chromosome 10



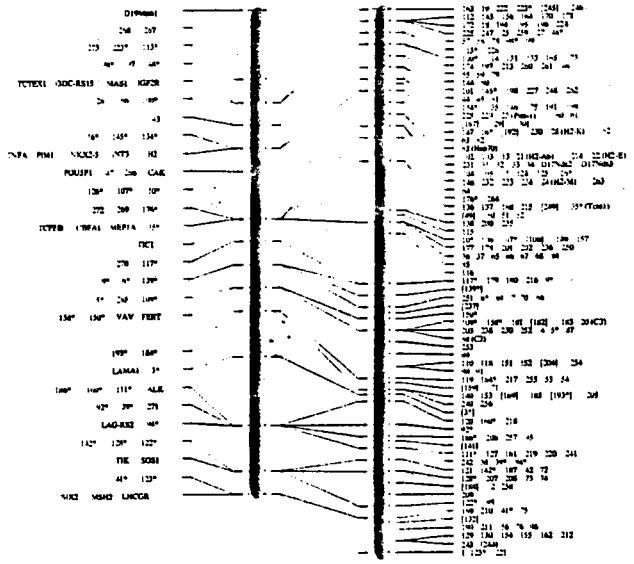
Chromosome 11



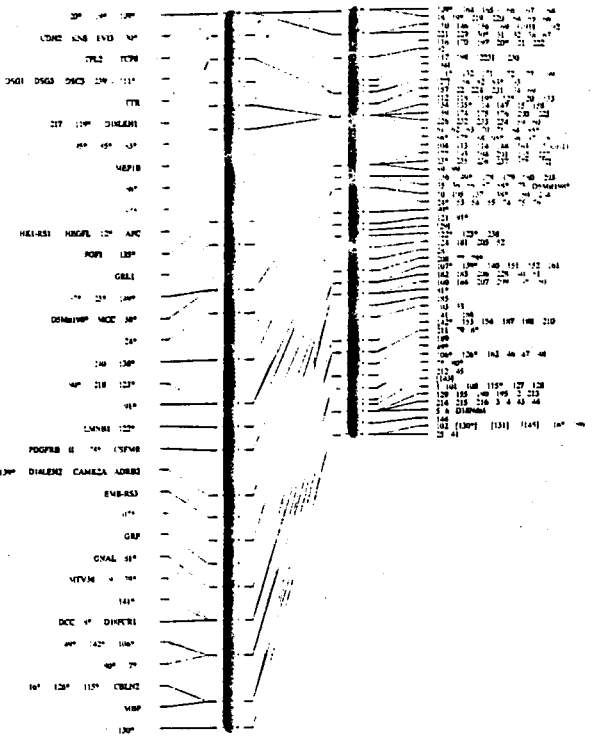
Chromosome 12



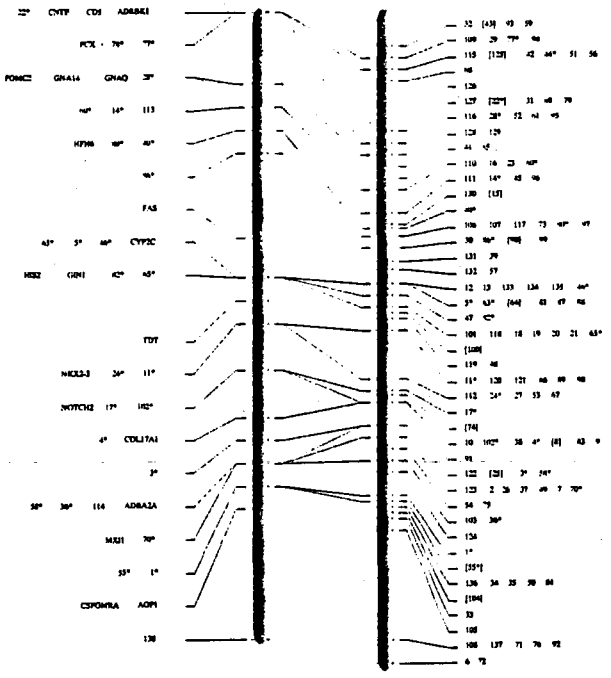
Chromosome 17



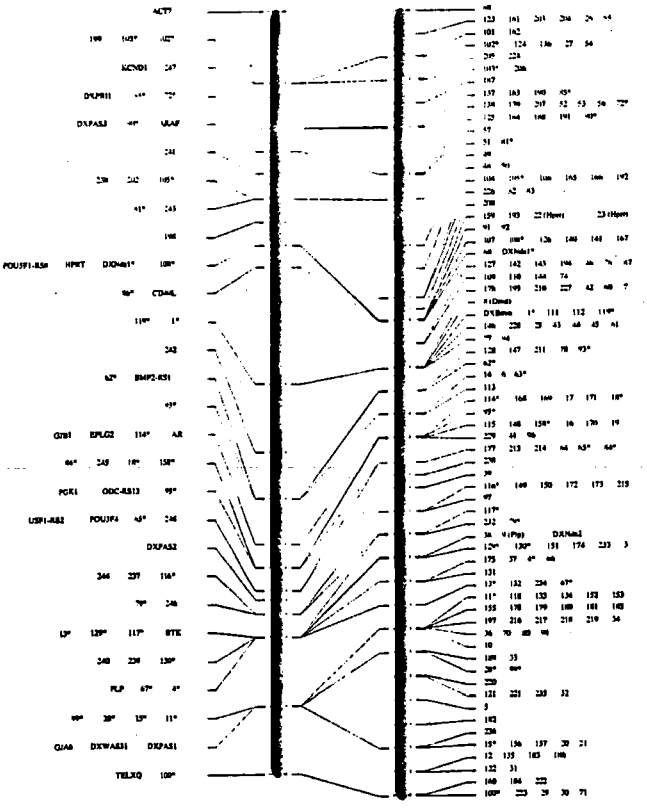
Chromosome 18



Chromosome 19



Chromosome X



A comprehensive genetic map of the mouse genome

William F. Dietrich*, Joyce Miller*, Robert Steen*, Mark A. Merchant*, Deborah Damron-Boles*, Zeeshan Husain*, Robert Dredge*, Mark J. Daly*, Kimberly A. Ingalls*, Tara J. O'Connor*, Cheryl A. Evans*, Margaret M. DeAngelis*, David M. Levinson*, Leonid Kruglyak*, Nathan Goodman*, Neal G. Copeland†, Nancy A. Jenkins†, Trevor L. Hawkins*, Lincoln Stein*, David C. Page*‡§ & Eric S. Lander*‡||

* Whitehead/MIT Center for Genome Research, Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA
† Mammalian Genetics Laboratory, ABL-Basic Research Program, NCI-Frederick Cancer Research and Development Center, Frederick, Maryland 21702, USA
‡ Department of Biology, and § Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA
|| To whom correspondence should be addressed.

The availability of dense genetic linkage maps of mammalian genomes makes feasible a wide range of studies, including positional cloning of monogenic traits, genetic dissection of polygenic traits, construction of genome-wide physical maps, rapid marker-assisted construction of congenic strains, and evolutionary comparisons[1,2]. We have been engaged for the past five years in a concerted effort to produce a dense genetic map of the laboratory mouse[3–6]. Here we present the final report of this project. The map contains 7,377 genetic markers, consisting of 6,580 highly informative simple sequence length polymorphisms integrated with 797 restriction fragment length polymorphisms in mouse genes. The average spacing between markers is about 0.2 centimorgans or 400 kilobases.

To construct a simple sequence length polymorphism (SSLP) map, we identified more than 9,000 sequences from random genomic clones and public databases containing simple sequence repeats (mostly, $(CA)_n$-repeats), designed polymerase chain reaction (PCR) primers flanking the repeat, and tested each for polymorphism by measuring the allele sizes in 12 inbred mouse strains. Of the successful PCR assays, we genotyped the 90% of loci that revealed different alleles between the OB and CAST strains in an (OB × CAST) F_2 intercross with 46 progeny. These data were assembled into a map by performing genetic linkage analysis with the MAPMAKER computer package[7,8].

A total of 6,336 SSLP loci were scored in the F_2 intercross, with 6,111 derived from anonymous sequence and 225 from known genes (Table 1). Of these, 5,905 were scored as codominant markers and 431 as dominant markers (because the pattern of one allele obscured the other). The map provides dense coverage of all 20 mouse chromosomes, with a total genetic length of 1,361 centimorgans (cM). Because the cross involves 92 meioses, the mean spacing between crossovers is 1.1 cM and thus loci can be mapped to 'bins' of this average size. The map has 1,001 occupied bins (Table 3(*a*)), with an average of 6.3 markers per bin and an average spacing of 1.36 cM between consecutive bins.

We next sought to integrate the map of largely anonymous SSLPs with the locations of known genes, because this information can suggest candidates for the genes underlying mouse mutations. We analysed a (B6 × SPRET) backcross that has been extensively used for restriction fragment length polymorphism (RFLP) mapping[9–11]. The backcross has been genotyped for 797 RFLPs. To integrate the maps, we genotyped 1,245 SSLPs from our map in 46 progeny from the SPRET backcross, providing a common reference point approximately every 1.1 cM. We also genotyped 244 additional SSLPs that were not polymorphic—and thus could not be mapped—in the (OB × CAST) intercross, but were polymorphic in the (B6 × SPRET) backcross. The SPRET cross was thus scored for a total of 1,543 SSLPs and 797 RFLPs.

The final map with 7,377 loci is shown in Fig. 1, with the SSLP map on the right and the integration with the RFLP map on the left. A full description of the markers—including primer sequences, locus sequence, genotypes in each cross, and allele

TABLE 1 Genetic markers, genetic length and polymorphism* by chromosome

Chromosome	No. of markers	No. of random markers	No. from GENBANK	'Consensus' genetic length†	Observed genetic length‡	Polymorphism among lab strains (%)§‖	Lab strains versus SPR or CAST (%)‖
1	511	494	17	98	109.9	57	92
2	507	491	16	107	95.7	49	94
3	343	332	11	100	67.5	51	95
4	350	342	8	81	74.2	51	93
5	402	391	11	93	82.9	48	95
6	368	349	19	74	59.1	46	94
7	357	341	16	89	59.8	48	94
8	350	345	5	81	72.0	44	94
9	336	318	18	70	62.9	52	95
10	293	286	7	78	73.0	35	96
11	350	326	24	78	82.0	53	94
12	278	268	10	68	61.5	50	94
13	303	296	7	72	60.2	48	95
14	259	246	13	53	65.6	49	94
15	264	257	7	62	62.2	51	94
16	215	214	1	59	51.0	43	94
17	255	239	16	53	51.0	56	93
18	231	226	5	57	39.7	53	95
19	134	131	3	42	57.2	52	93
X	230	219	11	88	73.5	33	95
Total	6,336	6,111	225	1,503	1,360.9¶	48	94

* Polymorphism survey was based on visual comparisons of fragments across large acrylamide gels and was thus subject to mobility differences among lanes. To assess the accuracy of data in our database, 3,000 individual pairwise comparisons were repeated. Some 6% of reported polymorphic pairs turn out to be monomorphic upon careful comparison, while 4% of reported monomorphic pairs turn out to be polymorphic. The data are thus accurate enough to allow selection of markers for crosses, but geneticists wishing to know every polymorphic marker in a narrow region (for fine-structure genetic mapping and positional cloning, for example) are advised to recheck each locus.
† Based on 'consensus' genetic map in Encyclopedia of the Mouse Genome, http://www.informatics.jax.org.encyclo.html (1993).
‡ Distance between most proximal and most distal markers in the map reported here.
§ Pairwise comparisons of OB, B6, DBA, A, C3H, BALB, AKR, NON, NOD and LP.
‖ Standard error of the mean for each chromosome depends on number of markers studied, but is < 1% in all cases.
¶ Distance is shorter than in previously published versions of this map (ref. 6) because final error checking reduced the number of apparent crossovers.

TABLE 2 Distribution of random markers based on cytogenetic length of chromosomes

Chromosome	No. of random markers†	Based on cytogenetic length* Percentage of total length	Expected number of markers‡	Z-score§
Autosomes only				
1	494	7.68 ± .15	452.7 ± 22.4	1.84
2	491	7.42 ± .15	437.0 ± 21.9	2.47
3	332	6.39 ± .13	376.7 ± 20.2	−2.20
4	342	6.29 ± .13	360.4 ± 20.0	−1.41
5	391	6.06 ± .12	356.2 ± 19.7	1.73
6	349	5.90 ± .12	347.7 ± 19.4	0.07
7	341	5.54 ± .11	326.4 ± 18.7	0.79
8	345	5.30 ± .11	312.5 ± 18.3	1.78
9	318	5.11 ± .10	301.2 ± 17.9	0.94
10	286	5.06 ± .10	298.1 ± 17.8	−0.67
11	326	5.04 ± .10	296.8 ± 17.8	1.65
12	268	5.21 ± .10	306.9 ± 18.1	−2.14
13	293	4.67 ± .09	275.4 ± 17.1	1.21
14	246	4.76 ± .10	280.5 ± 17.3	−1.99
15	257	4.32 ± .09	254.7 ± 16.4	0.15
16	214	4.07 ± .08	239.6 ± 15.9	−1.60
17	239	4.12 ± .08	242.7 ± 16.0	−0.22
18	226	4.14 ± .08	244.0 ± 16.0	−1.11
19	131	2.91 ± .06	171.7 ± 13.4	−3.04
Total	5,892	100.0	5,892.0	
Autosomes versus X chromosome				
Autosomes	5,892	93.76 ± .12	5,729.7 ± 20.4	7.96
X	219	6.24 ± .12	381.3 ± 20.4	−7.96
Total	6,111	100.0	6,111.0	

* Cytogenetic length taken from previous measurements[19]. Standard error of the mean was calculated directly from the raw data on chromosome measurements, generously provided by E. Evans.

† Only random markers are considered to avoid biases in chromosomal distribution of known genes.

‡ Mean ± standard deviation. Standard deviation in number of markers expected combines both standard error in the measurement of chromosome length and sampling error given to the total number of loci examined. Uncertainty in the precise length of chromosomes was not included in previous analyses[6], owing to its small magnitude, but it becomes relevant as the number of loci increases and sampling error correspondingly decreases. For comparison of autosomes to X chromosome, the expectation reflects the fact that 5% of the random markers were derived from male DNA (thus underrepresenting the X chromosome by a factor of two) and 95% from female DNA.

§ Z-score = (observed − expected)/standard deviation. For the autosomes, all of the Z-scores are significant at the $P = 0.05$ level after Bonferroni correction for multiple testing. For the comparison of autosomes to X chromosome, the Z-score is significant at $P < 10^{-14}$.

sizes in the characterized strains—would require over 500 pages of this journal. The complete information is available electronically on the WorldWide Web (see Fig. 1 legend).

The maps constructed in the CAST intercross and SPRET backcross maps have similar lengths (1,361 and 1,385 cM respectively), despite the fact that the intercross reflects sex-averaged recombination rates and the backcross reflects female recombination rates (because heterozygous mothers were used). Because there is typically about 80% more recombination in females than males, the SPRET backcross map might be expected to be about 40% longer. That it is not probably reflects recombinational suppression owing to structural heterogeneity (inasmuch as the laboratory mouse is evolutionarily twice as distant from SPRET as from CAST).

The SSLP map constructed in the cross was subjected to rigorous quality control and quality assessment[3,8]. All obligate double crossovers were identified and rechecked. The final data set contained no obligate double crossovers involving markers separated by less than 21 cM, indicating strong crossover interference in the mouse. (In the absence of interference, about 100 such events would be expected.) We also filled in any missing genotypes that could alter the position of a locus (by virtue of being adjacent to the site of a crossover). Despite our best efforts, some errors surely remain: in particular, an incorrect genotype adjacent to the site of a crossover would not necessarily produce a double crossover, and could shift a locus by 1.1 cM. Each chromosome is thus likely to contain a handful of loci that are slightly misplaced. The SSLPs used for integration with the SPRET backcross provided a different assessment of accuracy. We checked whether these 1,245 loci mapped to the same location in both crosses. There were ten apparent discrepancies. In five cases (*D5Mit198*, *D7Mit173*, *D9Mit132*, *D9Mit150* and *D19Mit61*), the loci were found to reproducibly amplify polymorphic fragments at different chromosomal locations in the two crosses. This probably occurs because strain variation creates an alternative target for amplification, although the possibility that CAST and SPRET differ by small insertional translocations cannot be excluded. In remaining five cases, the results from the CAST cross were found not to be reproducible. These probably arose from laboratory errors that unfortunately cannot be identified in retrospect. These five loci were removed from the map. Based on the frequency (5 of 1,245), we would expect that 20 further erroneous loci remain, which corresponds to about one per chromosome.

We used several criteria to analyse the genomic distribution of loci. The spacing between SSLPs agrees reasonably well with expectation under a random distribution, although some deviation from randomness can be detected. The relative positions of markers and crossovers can be inferred completely in an experimental cross, and the entire data set can be reduced to a string of the form 'mmcccmmmcccmcmcm...', with each m and c denoting the occurrence of a marker or a crossover, respectively. The hypothesis that markers are randomly distributed with respect to crossovers can be tested by comparing the observed clustering of consecutive markers and crossovers to that expected for tossing a biased coin with the probability of a marker being $p_m = M/(M + C)$, where M is the number of markers and C the number of crossovers[6]. There is some statistically significant evidence of clustering by this test (Table 3). The map contains an interval with eight consecutive crossovers (on chromosome 19) and a block of 54 recombinationally inseparable markers (on chromosome 2); the probability of such clusters of crossovers and markers occurring at random somewhere in the map is 0.5% and 3.4%, respectively. More generally, the frequency of both large and small clusters slightly exceeds expectation. Nonetheless, the distribution is not far from random expectation, at least at the level of resolution provided by the meioses studied here.

The chromosomal distribution of SSLPs among the autosomes agrees well with expectation under the assumption that loci are uniformly distributed with respect to cytogenetic length.

(Chromosome 19 shows a slight deficit, which is not statistically significant after correction for multiple testing; it may reflect the unusually large proportion of heterochromatin on this chromosome.) In contrast, chromosome X shows a clear deficit, with only about 57% as many as expected (Table 2). This phenomenon appears to be general in mammalian genomes, as we have also found a similar deficit in an SSLP map of the rat[12] (62% of expectation), and Weissenbach and colleagues report a slightly less pronounced deficit in the human genome[13] (75% of expectation). In principle, the deficit of SSLPs on chromosome X could occur if $(CA)_n$-repeats were either less frequent on chromosome X, or were equally frequent but less polymorphic. The latter hypothesis would predict that the deficit of polymorphic loci on chromosome X would be offset by a great excess of non-polymorphic repeats. Of the SSLPs monomorphic between OB and CAST, 37% would have to lie to chromosome X to explain the observed data. We determined the chromosomal location of > 100 monomorphic loci (by genetic mapping for those that were polymorphic between B6 and SPRET and by somatic cell hybrid mapping for those that were not), but we found no significant excess on chromosome X. Accordingly, the deficit appears to be primarily due to an actual shortage of $(CA)_n$-repeats on chromosome X.

The SSLPs show a polymorphism rate of about 50% among inbred laboratory strains surveyed and about 95% between laboratory strains and CAST or SPR (Table 1). The pairwise polymorphism rates among the 12 strains surveyed have not changed significantly from our previous report[6] and are not presented here. Interestingly, the distribution of polymorphism across the genome is not uniform[11]. The average polymorphism rate among the *Mus musculus* strains surveyed was just under 50%, but two chromosomes showed substantially lower polymorphism rates: chromosome X at 33%, and chromosome 10 at 35% (Table 1). Decreased polymorphism could reflect recent selection for specific ancestral chromosomes. For the X chromosome, it could also reflect a different mutation rate (inasmuch as each chromosome X resides in males only two-thirds as often each autosome, and most mutations are thought to occur in male germline) or different population genetic forces (with hemizygosity affecting selection and effective population size).

Our mouse genetic-mapping project is now at its conclusion. Although more SSLPs remain to be found (newly

TABLE 3 Clusters of consecutive crossovers and markers

(a) Number of crossovers between consecutive random markers*

No. of crossovers per interval	Observed No.	Observed (percentage)	Expected† No.	Expected† (percentage)	P(longest run ≥ n) (%)†
0	5,095	(83.85)	5,035.5 ± 29.6	(82.59)	
1	784	(12.90)	876.7 ± 27.4	(14.38)	100.0
2	151	(2.49)	152.6 ± 12.2	(2.50)	100.0
3	27	(0.44)	26.6 ± 5.1	(0.44)	100.0
4	14	(0.23)	4.6 ± 2.2	(0.08)	99.6
5	4	(0.07)	0.8 ± 0.9	(0.01)	62.2
6	0	(0.00)	0.1 ± 0.4	(< 0.01)	15.6
7	0	(0.00)	0.0 ± 0.2	(< 0.01)	2.9
8	1	(0.02)	0.0 ± 0.1	(< 0.01)	0.5
Total	6,076				

(b) Random markers occurring between consecutive crossovers‡

No. of markers per block	Observed No.	Observed (percentage)	Expected§ Number	Expected§ (percentage)	P(longest run ≥ n) (%)§
0	288	(22.3)	227.9 ± 13.7	(17.4)	100.0
1	208	(16.1)	188.2 ± 12.7	(14.4)	100.0
2	126	(9.8)	155.5 ± 11.7	(11.9)	100.0
3	111	(8.6)	128.4 ± 10.8	(9.8)	100.0
4	84	(6.5)	106.0 ± 9.9	(8.1)	100.0
5	73	(5.7)	87.6 ± 9.0	(6.7)	100.0
6	62	(4.8)	72.3 ± 8.3	(5.5)	100.0
7	51	(4.0)	59.7 ± 7.6	(4.6)	100.0
8	36	(2.8)	49.3 ± 6.9	(3.8)	100.0
9	38	(2.9)	40.7 ± 6.3	(3.1)	100.0
10	32	(2.5)	33.7 ± 5.7	(2.6)	100.0
11	37	(2.9)	27.8 ± 5.2	(2.1)	100.0
12	19	(1.5)	23.0 ± 4.7	(1.8)	100.0
13	28	(2.2)	19.0 ± 4.3	(1.4)	100.0
14	18	(1.4)	15.7 ± 3.9	(1.2)	100.0
15	7	(0.5)	12.9 ± 3.6	(1.0)	100.0
16	12	(0.9)	10.7 ± 3.3	(0.8)	100.0
17	5	(0.4)	8.8 ± 3.0	(0.7)	100.0
18	5	(0.4)	7.3 ± 2.7	(0.6)	100.0
19	6	(0.5)	6.0 ± 2.4	(0.5)	100.0
20	10	(0.8)	5.0 ± 2.2	(0.4)	100.0
21	3	(0.2)	4.1 ± 2.0	(0.3)	100.0
22	5	(0.4)	3.4 ± 1.8	(0.3)	100.0
23	7	(0.5)	2.8 ± 1.7	(0.2)	100.0
24	4	(0.3)	2.3 ± 1.5	(0.2)	100.0
25	0	(0.0)	1.9 ± 1.4	(0.1)	100.0
26	5	(0.4)	1.6 ± 1.3	(0.1)	99.9
27	1	(0.1)	1.3 ± 1.1	(0.1)	99.8
28	1	(0.1)	1.1 ± 1.0	(0.1)	99.3
29	0	(0.0)	0.9 ± 0.9	(0.1)	98.4
30	1	(0.1)	0.7 ± 0.9	(0.1)	96.7
31	1	(0.1)	0.6 ± 0.8	(< 0.1)	94.0
32	0	(0.0)	0.5 ± 0.7	(< 0.1)	90.3
33	0	(0.0)	0.4 ± 0.6	(< 0.1)	85.4
34	1	(0.1)	0.3 ± 0.6	(< 0.1)	79.6
35	1	(0.1)	0.3 ± 0.5	(< 0.1)	73.1
38	1	(0.1)	0.2 ± 0.4	(< 0.1)	52.2
40	1	(0.1)	0.1 ± 0.3	(< 0.1)	39.6
54	1	(0.1)	< 0.1 ± 0.1	(< 0.1)	3.4
Total	1,289				

*The intervals with ≥ 1 crossover represent the 981 gaps between consecutive bins of recombinationally inseparable markers. Only random markers are considered to avoid biases in distribution of known genes.

†The probability of the longest run is calculated in ref. 6. Briefly, if a coin with heads probability P is tossed n times, the length R_n of the longest head run has expected value $\mu = \log_{1/p}[(n-1)(1-p)+1]$ and the distribution of R_n is given approximately by $\text{Prob}(R_n - \mu > t) = 1 - \exp(-p^t)$. In this case, $p = 0.17$.

‡The blocks with ≥ 1 marker represent the 1,001 bins of recombinationally separable markers. Only random markers are considered to avoid biases in distribution of known genes.

§The probability of the longest head run is calculated with $p = 0.83$.

isolated repeats show < 10% overlap with our current set), we have reached the point of diminishing returns. The map covers the entire mouse genome, with the markers being sufficiently abundant, polymorphic and stable to allow the mapping of monogenic or polygenic traits in virtually any mouse cross of interest[5,8]. Moreover, the markers are sufficiently dense to facilitate positional cloning of most mouse mutations. With > 90% of the mouse genome being within 750 kb of a marker, and current mouse yeast artificial chromosome (YAC) libraries[14,15] having a mean insert size > 750 kb, the map affords ready access to the vast majority of the genome with little need for chromosomal walking, and provides a preliminary scaffold for constructing a genome-wide physical map[16].

The map also provides a common framework for the mapping of mutations and cloned genes. In addition to our integration with the Frederick cross, the SSLP map is being used as a framework for other mapping crosses, including public resources at the Jackson Laboratory[17] and the European Collaborative Interspecific Backcross (EUCIB)[18]. The EUCIB project (http://www.hgmp.mrc.ac.uk/MBx/MBxHomepage.html) is rescoring our SSLP markers in a cross with 1,000 meioses, which should yield finer resolution of order and correct remaining errors.

Together with the final report on the human genetic map[13], this paper marks the close of the first phase of the Human Genome Project: the construction of dense genetic maps of mouse and man. □

Received 23 October 1995; accepted 19 February 1996.

1. Copeland, N. G. *et al. Science* **262,** 57–66 (1993).
2. Copeland, N.G. *et al. Science* **262,** 67–82 (1993).
3. Dietrich, W. F. *et al. Genetics* **131,** 423–447 (1992).
4. Dietrich, W. F. *et al.* in *Genetic Maps 1992* (ed. O'Brien, S.) 4.110–4.142 (Cold Spring Harbor Laboratory Press, NY, 1992).
5. Miller, J. C. *et al.* in *Genetic Variants and Strains of the Laboratory Mouse* 3rd edn. (eds Lyons, M. F. & Searle, A.) (Oxford Univ. Press, New York, 1994).
6. Dietrich, W. F. *et al. Nature Genet.* **7,** 220–245 (1994).
7. Lander, E. S. *et al. Genomics* **1,** 174–181 (1987).
8. Lincoln, S. E. & Lander, E. S. *Genomics* **14,** 604–610 (1992).
9. Copeland, N. G. & Jenkins, N. A. *Trends Genet.* **7,** 113 (1991).
10. Ceci, J. D. *et al. Genomics* **5,** 699–709 (1989).
11. Buchberg, A. M. *et al. Genetics* **122,** 153–161 (1989).
12. Jacob, H. J. *et al. Nature Genet.* **9,** 63–69 (1995).
13. Dib, C. *et al. Nature* **380,** 152–154 (1996).
14. Larin, Z., Monaco, A. P. & Lehrach, H. *Proc. natn. Acad. Sci. U.S.A.* **88,** 4123 (1991).
15. Kusumi, K. *et al. Mamm. Genome* **4,** 391–392 (1993).
16. Hudson, T. *et al. Science* **270,** 1945–1955 (1995).
17. Rowe, L. B. *et al. Mamm. Genome* **5,** 253–274 (1994).
18. The European Backcross Collaborative Group. *Hum. molec. Genet.* **3,** 621–627 (1994).
19. Evans, E. in *Genetic Variants and Strains of the Laboratory Mouse* 3rd edn (eds Lyons, M. F. & Searle, A.) (Oxford Univ. Press, New York, 1994).

ACKNOWLEDGEMENTS. We thank L. Wangchuk, D. Tsering, G. Farino and K. Norbu for technical assistance; D. Gilbert and L. Maitais for help in ascertaining official nomenclature for gene loci; and Research Genetics Inc. for making SSLP primers available to the community. This work was supported in part by a grant from the National Center for Human Genome Research (to E.S.L.). L.K. was supported by a Special Emphasis Research Career Award from the National Center for Human Genome Research.

A comprehensive genetic map of the human genome based on 5,264 microsatellites

Colette Dib*, Sabine Fauré*, Cécile Fizames*, Delphine Samson*, Nathalie Drouot*, Alain Vignal*, Philippe Millasseau*, Sophie Marc*, Jamilé Hazan*, Eric Seboun*, Mark Lathrop†, Gabor Gyapay*, Jean Morissette*‡ & Jean Weissenbach*§

* Généthon and CNRS URA 1922, 1 rue de l'Internationale, 91000 Evry, France
† INSERM U358, Hôpital Saint-Louis, Paris, France
‡ Centre de Recherche du Centre Hospitalier de l'Université Laval, Québec G1V 4G2, Canada
§ To whom correspondence should be addressed

The great increase in successful linkage studies in a number of higher eukaryotes during recent years has essentially resulted from major improvements in reference genetic linkage maps[1–6], which at present consist of short tandem repeat polymorphisms of simple sequences or microsatellites[7,8]. We report here the last version of the Généthon human linkage map[6]. This map consists of 5,264 short tandem (AC/TG)$_n$ repeat polymorphisms with a mean heterozygosity of 70%. The map spans a sex-averaged genetic distance of 3,699 cM and comprises 2,335 positions, of which 2,032 could be ordered with an odds ratio of at least 1,000 : 1 against alternative orders. The average interval size is 1.6 cM; 59% of the map is covered by intervals of 2 cM at most and 1% remains in intervals above 10 cM.

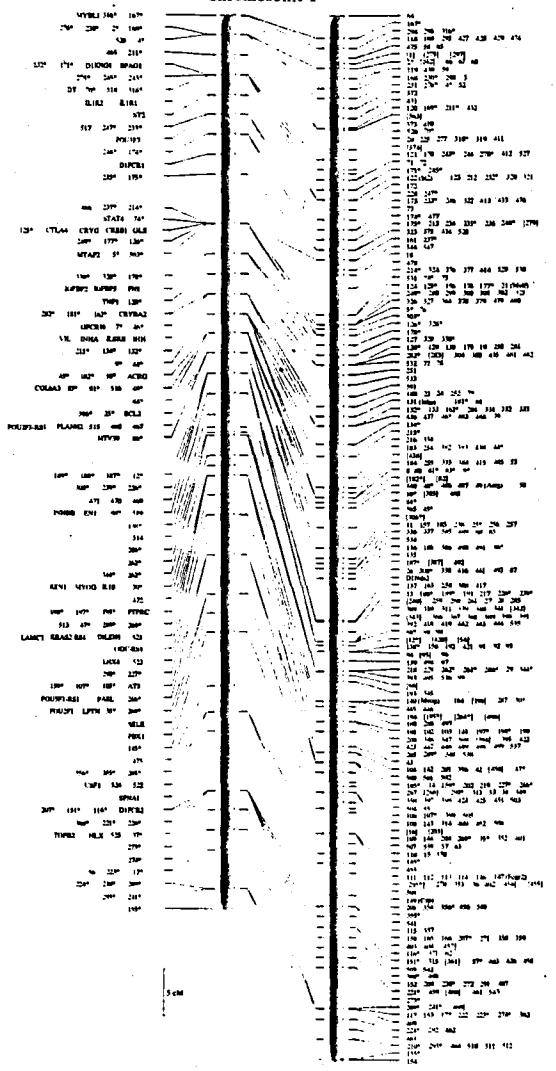
Microsatellite markers were obtained as described previously[5,6]. A heterozygosity above 0.5 was observed for 93% of the markers and above 0.7 for 58%. These values remain very close to those of our previous version[6]. Average heterozygosity per chromosome varied from 0.65 (chromosome X) to 0.73 (chromosome 19), with a mean value of 0.70 for the entire collection of markers (Table 1). Database sequence comparisons and searches detected matches of AFM (Association Française contre les Myopathies) markers with 19 genes and 74 anonymous markers.

Genotyping of the microsatellite markers was performed as described previously on the same eight CEPH (Centre d'Etudes du Polymorphisme Humaine) families (20 for the X chromosome), which comprised a total of 134 individuals and 186 meioses[5,6] (304 individuals and 291 meioses for the X chromosome). Genotypes were submitted to the same error-checking procedures as reported earlier[6]. These procedures consisted of (1) a reinvestigation of families with abnormally elevated recombination frequencies between pairs of markers, and (2) correction or elimination of all double recombinant genotypes of markers placed in short linkage intervals. Such apparent double recombinations probably result from mutation events that converted an allele of one individual into the other allele. A more detailed analysis of double-recombination events and mutations in microsatellites is in preparation.

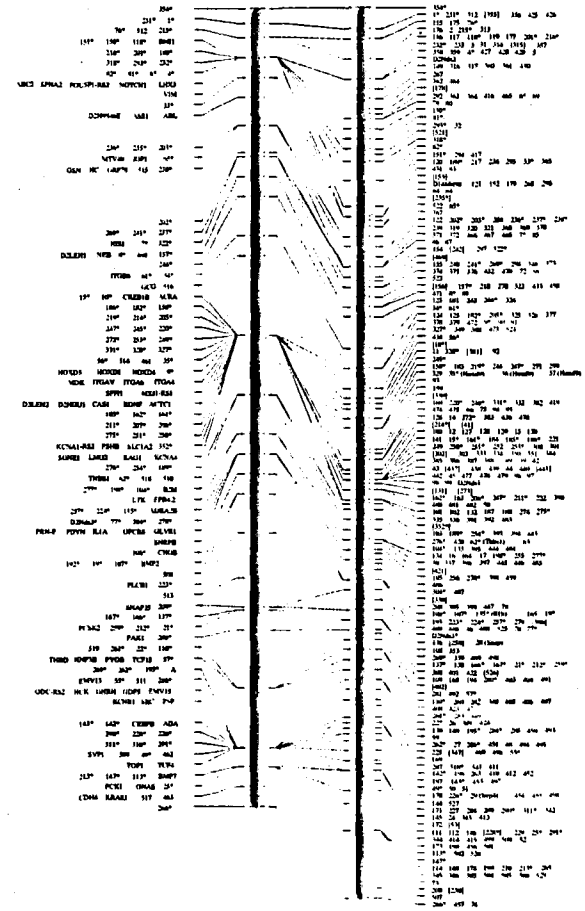
Map construction was done in a stepwise manner with multiple controls at each step. The total length of this map as evaluated from the CILINK algorithm[9] is 3,699 cM (Table 1). This is almost identical in length to our previous version, despite the addition of new terminal markers that extend the 93/94 chromosome maps by 145 cM (4%). The absence of increase in length probably results from a very thorough error-checking process and from elimination of apparent double-recombinant genotypes. The 5,264 markers are distributed in 2,335 positions (Fig. 1), 2,032 of which are ordered with odds ratios against alternative orders of at least 1,000 : 1. The mean interval size is 1.6 cM. The fraction of the map in intervals above 10 cM represents only 1 per cent of the total linkage distance and consists of 3 intervals spanning 11 cM. Fifty-nine per cent of the map is covered by intervals of 2 cM at most, and 92 per cent by intervals of 5 cM at most. Markers from the CEPH and CHLC databases have been integrated into this map as shown in Fig. 2, which presents the map of chromosome 22 as an example. Detailed information, including integrated maps of all chromosomes, a list of markers, their primer sequences, heterozygosity, number and size-range of alleles observed in the 8 (or 20) genotyped CEPH families, sex-specific distances, and mutations, will be presented in an extended reprint available on request and on an electronic server (http://www.genethon.fr).

The total sex-specific lengths of autosomes estimated by CILINK[9] show only slight variations when compared to the lengths of the previous map[6]. The length excess observed for the female map is comparable to other published maps. This excess

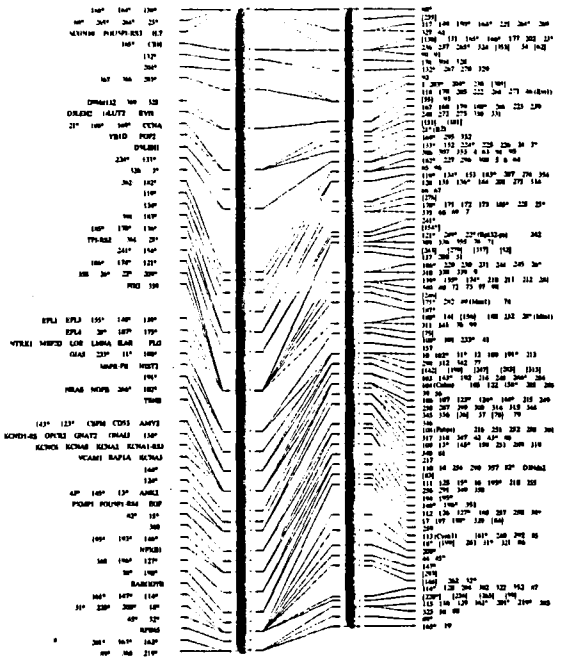
Chromosome 1



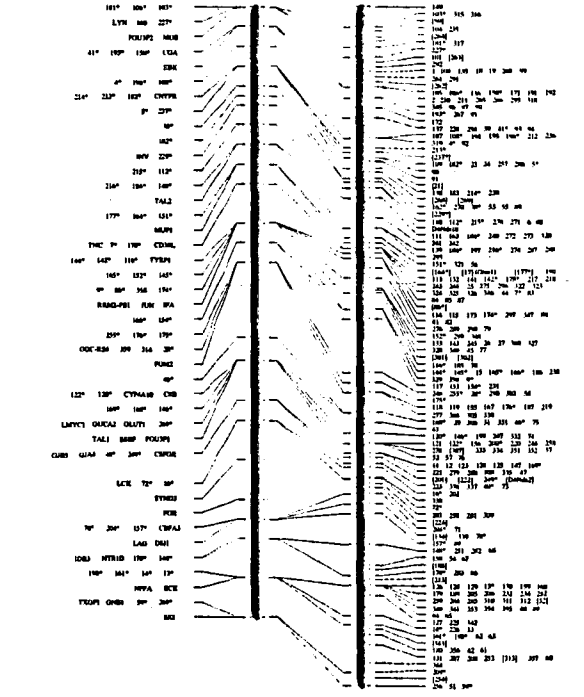
Chromosome 2



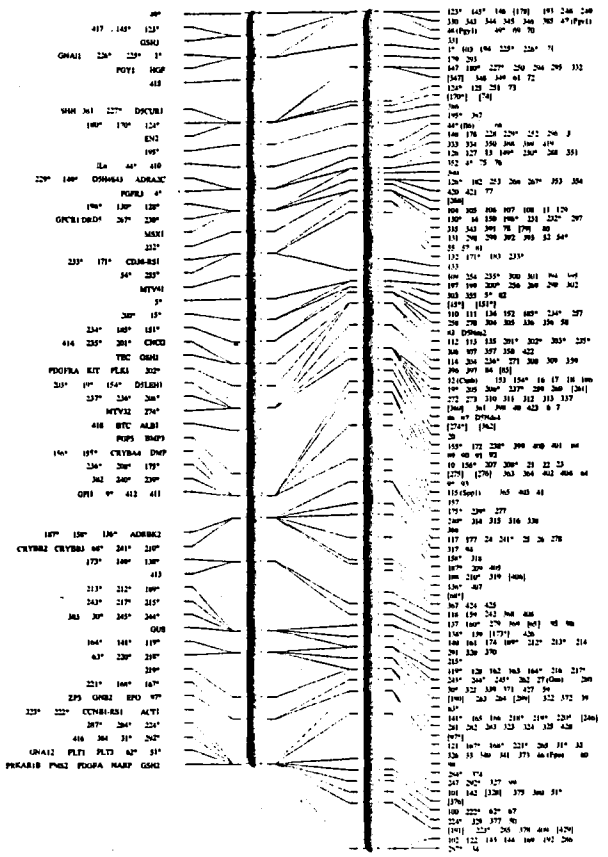
Chromosome 3



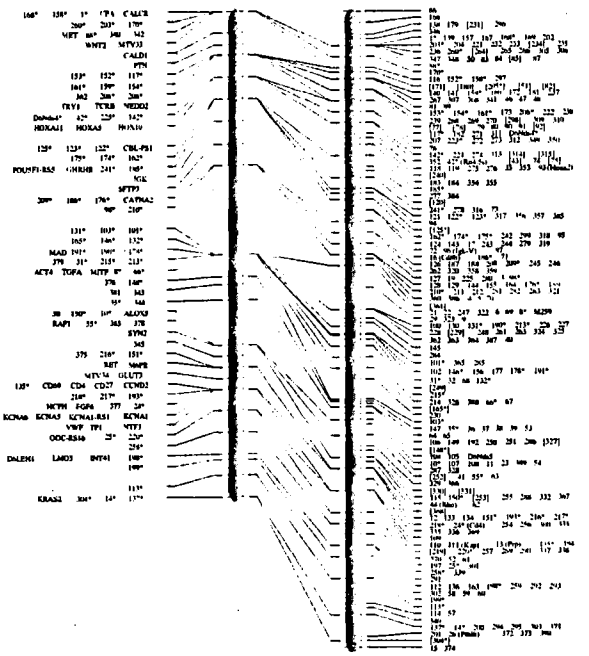
Chromosome 4



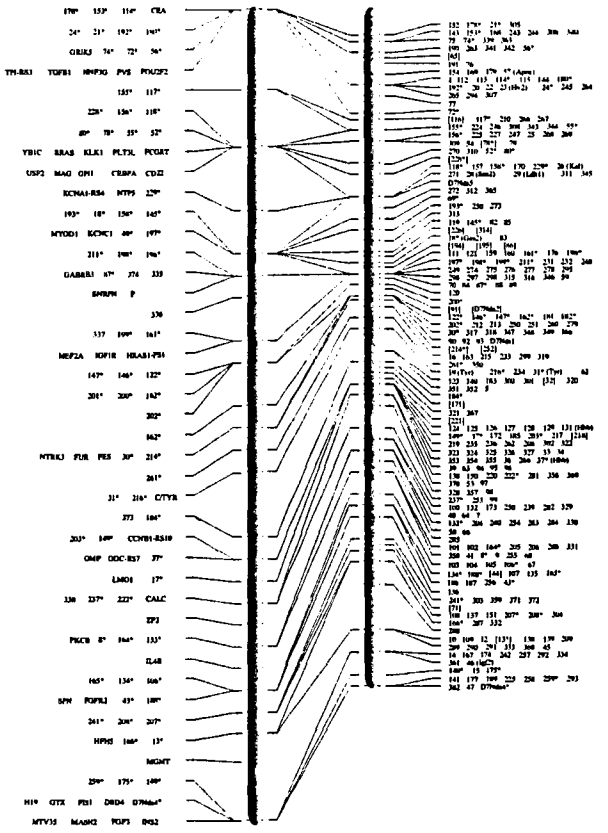
Chromosome 5



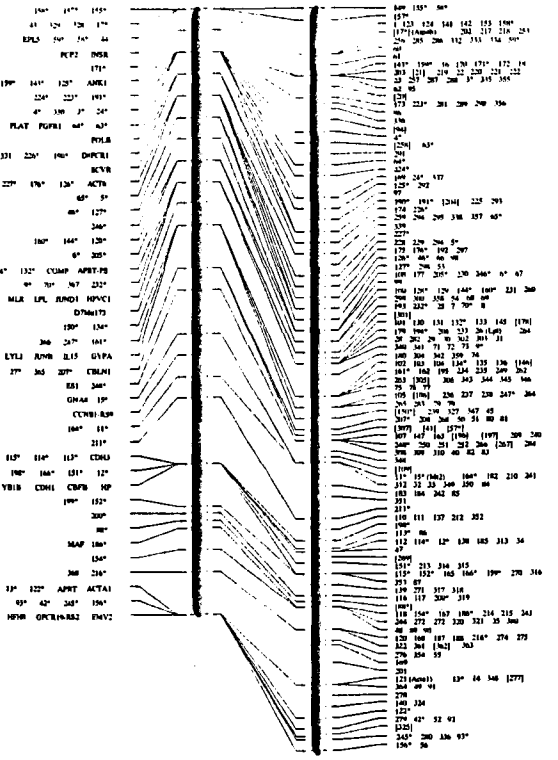
Chromosome 6



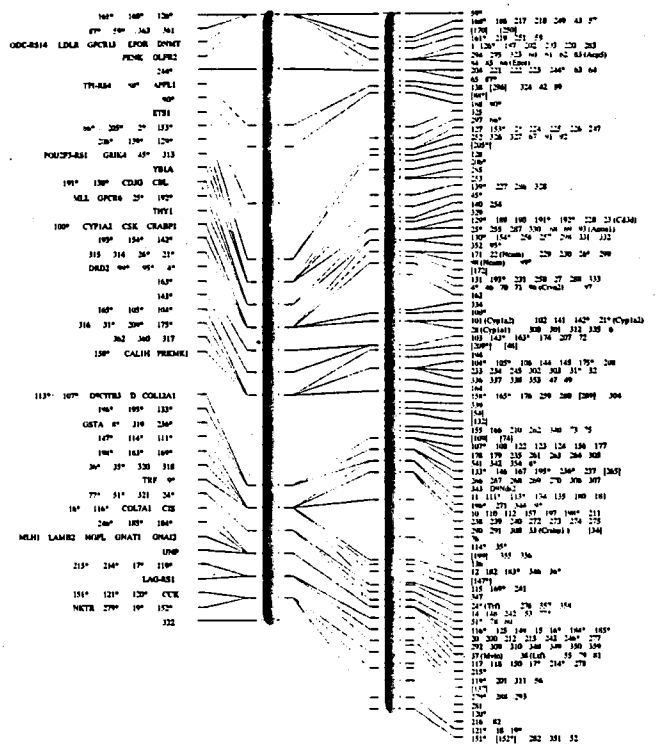
Chromosome 7



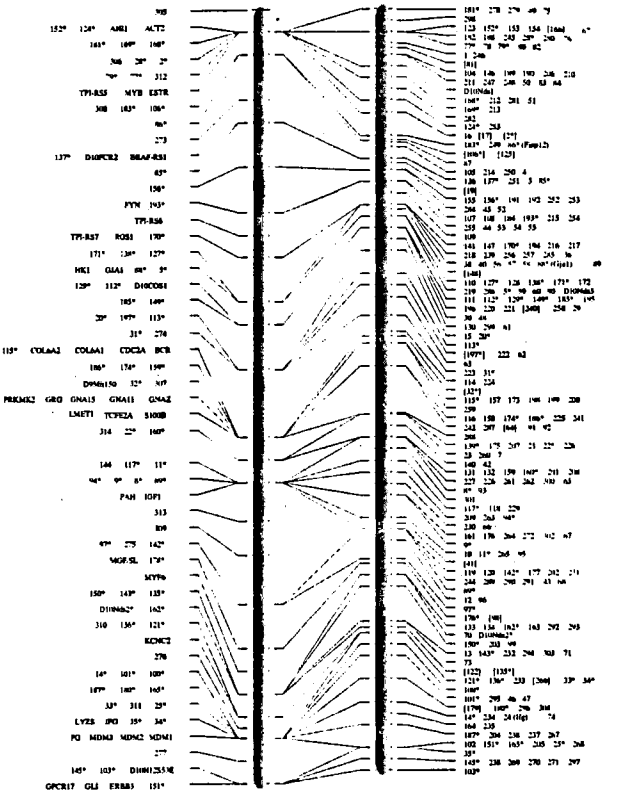
Chromosome 8



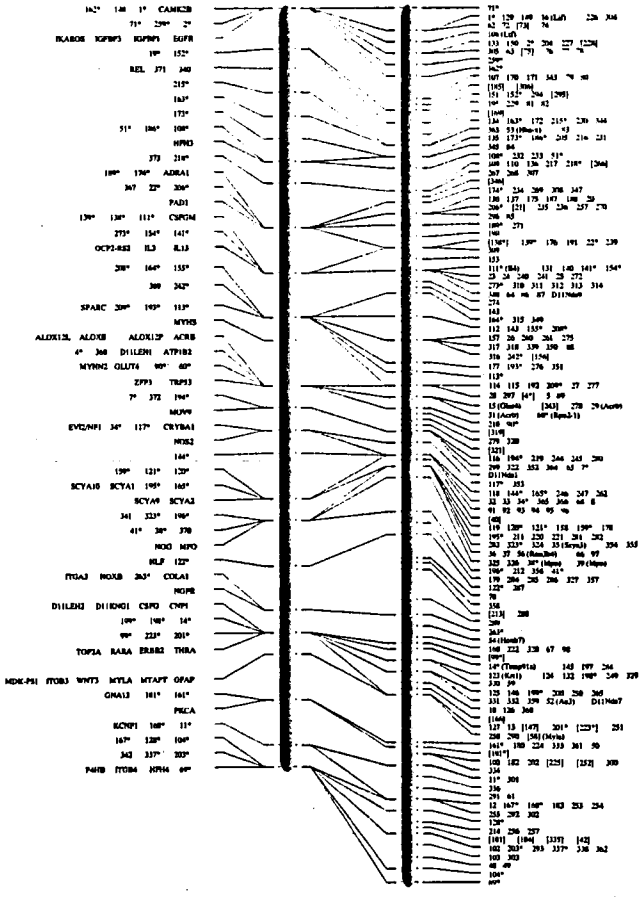
Chromosome 9



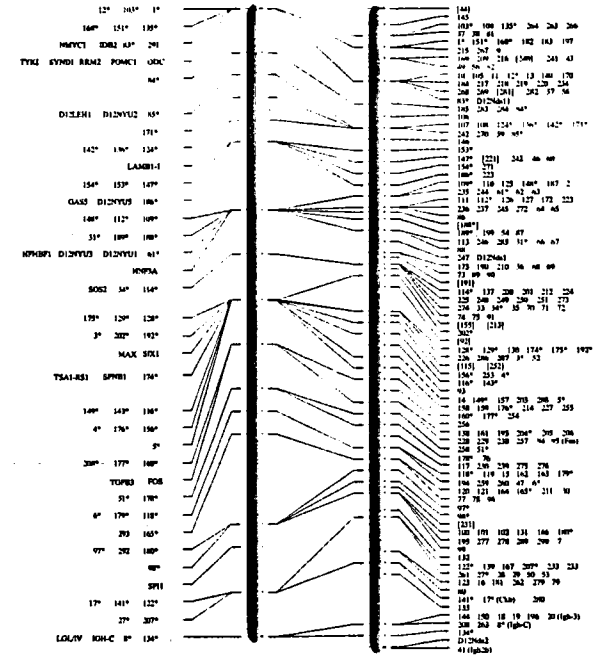
Chromosome 10



Chromosome 11



Chromosome 12



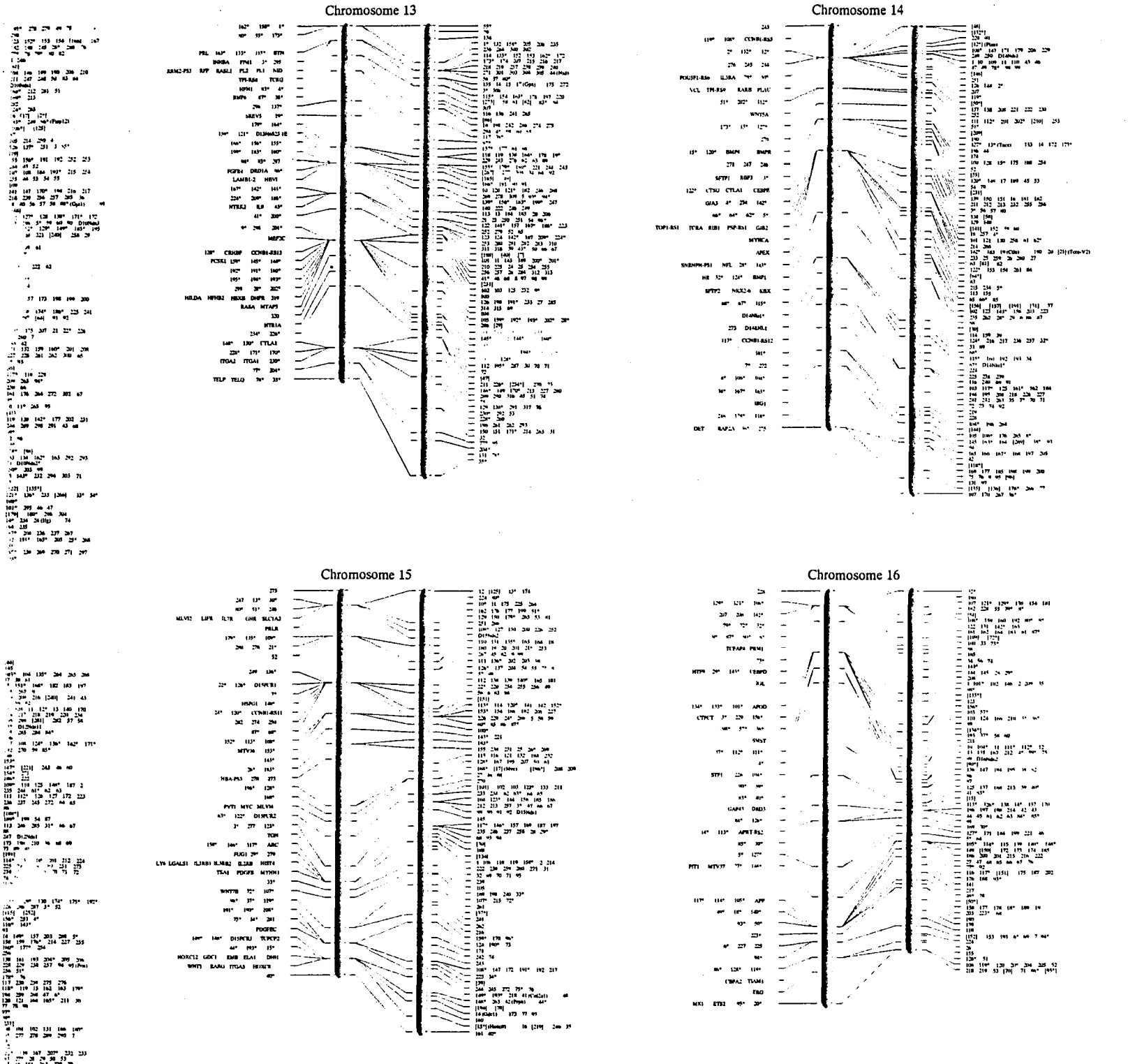
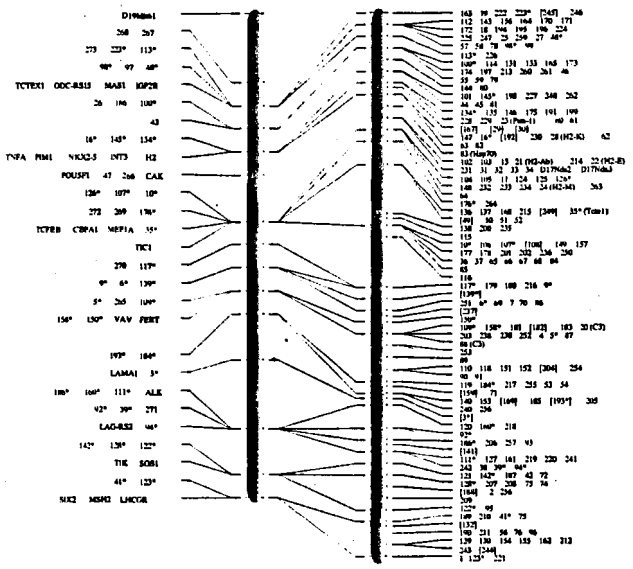
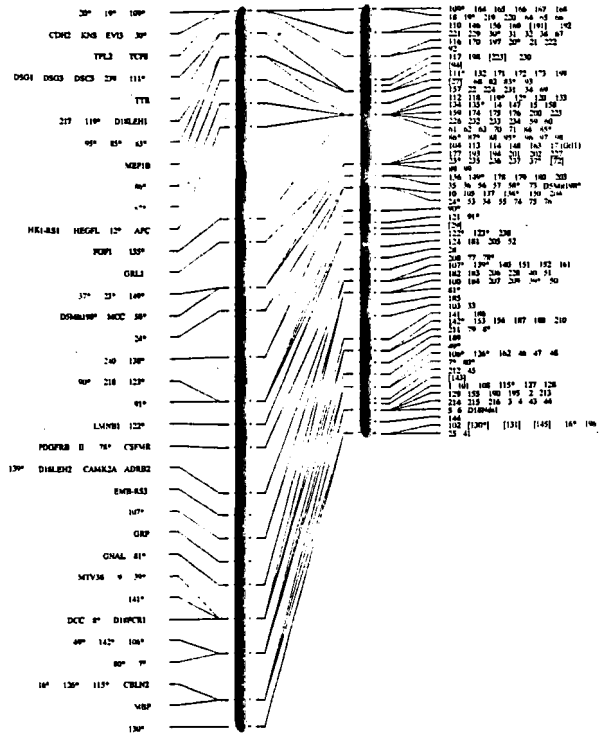


FIG. 1 Genetic map of the mouse. The map on the right was constructed in the Whitehead Institute/MIT CAST intercross; that on the left was constructed in the SPR backcross. For SSLPs, formal locus names have been abbreviated; for example, the locus *D7Mit3* is simply denoted by 3 on chromosome 7. For the few SSLP loci developed elsewhere, the laboratory designation is retained (for example, *D4Nds1* is denoted Nds1). For SSLPs developed from genes for which a gene symbol has been assigned by the Mouse Nomenclature Committee, the gene symbol is given in parentheses. For RFLPs in genes, the gene symbol is given. Linkage groups are represented by lines, with the centromere at the top. Loci genotyped in both crosses are followed by an asterisk; a line connects the respective positions in the two crosses. Loci genotyped as dominant systems are indicated by brackets. The map position of such loci is not certain because the meioses are not fully informative. The range of possible positions for the locus can be found by examining our electronic database, which contains the underlying genotypes for all loci. Loci that did not recombine in the meioses studied are listed together in a block. Distances were calculated by using Kosambi's map function[6]. Map scale is shown to the side of the map of chromosome 1. One anomaly should be noted: in the (B6 × SPRET) backcross, an SSLP (*D9Mit33*) derived from the sequence of the *Crabp1* gene maps to a different region of chromosome 9 than does an RFLP detected by a probe for this gene. This does not appear to be due to genotype errors, but may arise because the two assays detect different loci for technical reasons. The underlying data are all available electronically from our World-Wide Web site (http://www-genome.wi.mit.edu), an electronic mail server (send a message with the single word 'help' to genome_database@genome.wi.mit.edu), and anonymous ftp (at ftp-genome.wi.mit.edu). The analysis tools are also available electronically.

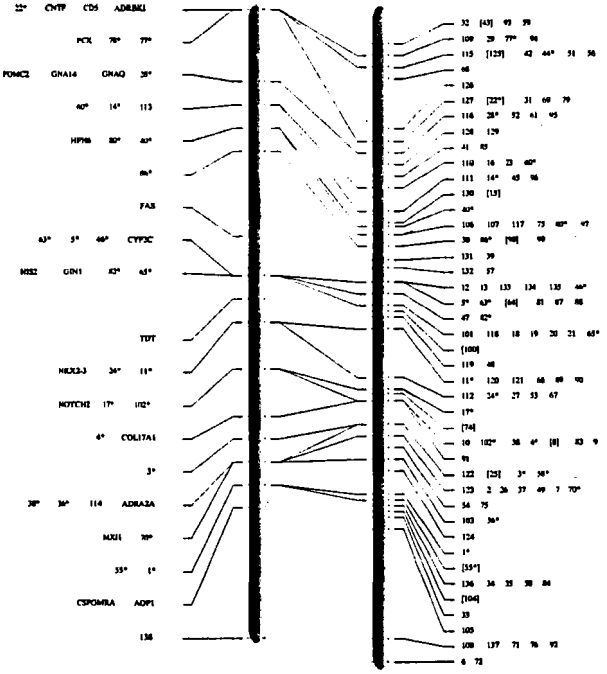
Chromosome 17



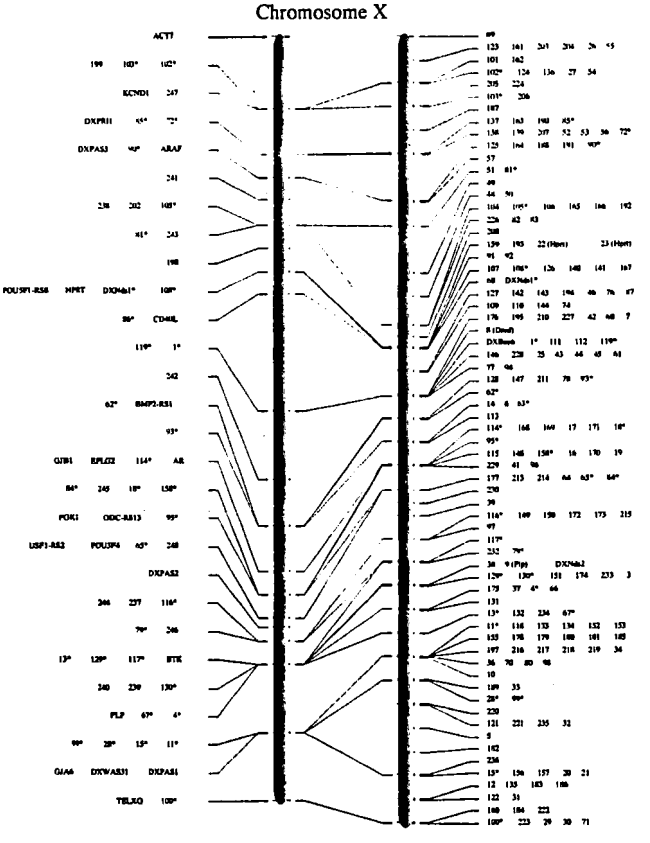
Chromosome 18



Chromosome 19



Chromosome X



4742–4743 Nucleic Acids Research, 1995, Vol. 23, No. 22

Solid-phase reversible immobilization for the isolation of PCR products

Margaret M. DeAngelis, David G. Wang and Trevor L. Hawkins*

Whitehead Institute/MIT, Center for Genome Research, One Kendall Square, Cambridge, MA 02139, USA

Received March 15, 1995; Revised and Accepted October 12, 1995

Large numbers of templates for DNA sequencing can be produced via PCR directly from plaques, colonies or genomic DNA. Sequencing directly from PCR products has many advantages over subcloning; the ability to PCR directly from plaques or colonies removes the need for template preparation and is highly amenable to automation. The main problem with this approach is the subsequent purification of the amplified products prior to DNA sequencing, especially since the sequence quality is proportional to the purity of the template. This is especially important when sequencing PCR products to identify sequence polymorphisms.

The advantages of using magnetic particles in molecular and diagnostic biology have been described previously (1–4). The use of solid phase techniques has significantly increased over the past few years as more biochemical methods have become adapted for use with magnetic particles.

We introduce here a general method for producing quality DNA sequencing template from PCR products. This procedure is rapid and inexpensive ($0.15 per prep.). The method termed SPRI (solid-phase reversible immobilization) avoids organic extraction, filtration and centrifugation steps (5). The SPRI method employs a carboxyl coated magnetic particle manufactured by PerSeptive Diagnostics, Cambridge, MA. (cat no #8-4125). We discovered that these particles could reversibly bind DNA in the presence of polyethylene glycol (PEG) and salt.

This solid phase has no streptavidin, making the use of biotinylated primers or probes attached to the particles unnecessary. When using biotinylated primers one must exercise caution since excess primer will compete for streptavidin particle binding (6). This in turn may also contribute to lower yield and quality of the template.

Here we describe a general PCR isolation procedure which is amenable to automation, rapid and yields double-stranded PCR product suitable for DNA sequencing. The method is as follows.

PCR primers. Forward primers are tailed with –21M13 sequences. TGTAAAACGACGGCCAGT (18 nt).

PCR reagents.

1. 10× PCR buffer [100 mM Tris–HCl (pH 9.3); 500 mM KCl; 15 mM $MgCl_2$; 0.01% gelatin].
2. 10 mM dNTPs.
3. 10 µM forward and reverse primers.
4. 20 ng/µl genomic DNA.

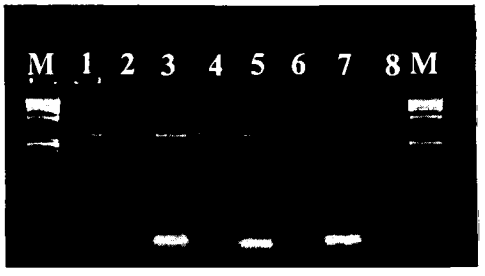


Figure 1. This gel shows an example of PCR products before and after purification using SPRI. Lanes M are 200 ng φX174 *Hae*III digest, lanes 1 and 2 are one-tenth of a PCR product before and after SPRI purification, lanes 3 and 4 are one tenth of a PCR product spiked with 100 nmol of excess primer (36 nt) prior to purification, lane 3 is before and lane 4 is after SPRI. Lanes 5 and 6 are identical to lanes 3 and 4 using a different PCR product. Lanes 7 and 8 are 100 nmol excess primer (36 nt) before and after purification using SPRI.

Standard PCR reaction (50 µl).

1. 6.5 µl PCR MIX [10× PCR buffer, 5 µl; 10 mM dNTPs, 0.5 µl; Taq, 1 U; dH_2O to add up to 6.5 µl].
2. 41 µl primer dilution [10 µM F&R primers, 0.5 µl; dH_2O, 40.5 µl].
3. 2.5 µl genomic DNA (50 ng).

PCR conditions (35 cycles). 96°C, 5 min; 96°C, 30 s; 57°C or 55°C, 2 min; 72°C, 2 min; 72°C, 5 min; 4°C.

Solid-phase reversible immobilization for the purification of PCR products (96-well format).

1. Wash 10 mg/ml carboxyl coated magnetic particles three times with WASH BUFFER [0.5 M EDTA (pH 8.0)].
2. For each PCR reaction (50 µl), add 10 µl of washed particles and 50 µl of HYB BUFFER (2.5 M NaCl/20% PEG 8000). Mix well and incubate at room temperature for 10 min.
3. Place the microtitre plate on a magnet for 2 min and wash the particles twice with 150 µl of 70% EtOH.
4. Air dry for 2 min, and resuspend the particles in 20 µl of ELUTION BUFFER [10 mM Tris-acetate (pH 7.8)] and incubate at room temperature for 5 min.
5. Magnetically separate the particles and remove the supernatant for testing and sequencing.

* To whom correspondence should be addressed

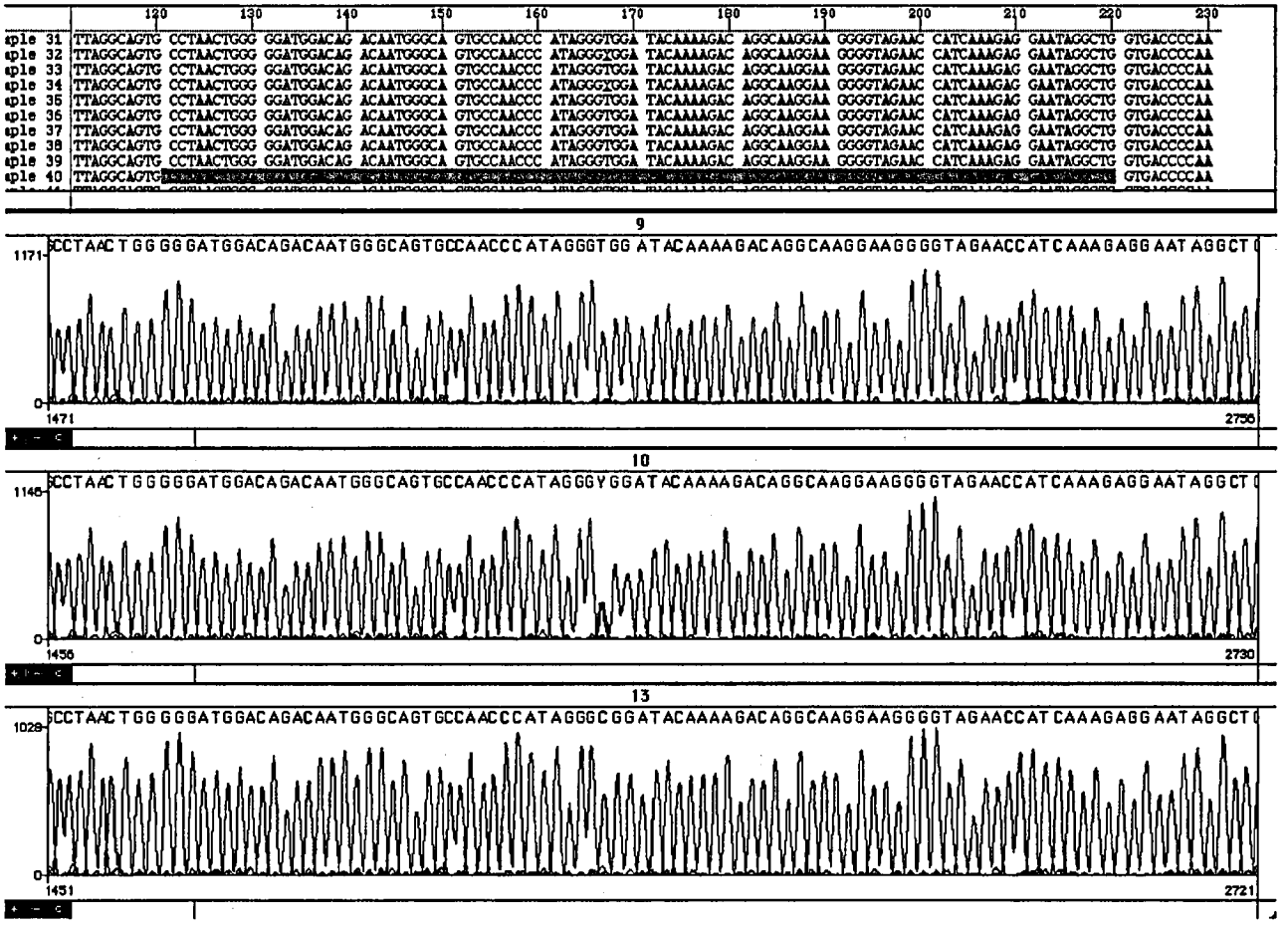


Figure 2. Sequencing traces derived from three related individuals are aligned to show the reliability of calling heterozygous bases. The cycle sequencing reactions of purified double-stranded SPRI purified PCR products were performed using AmpliTaq FS DNA polymerase (Applied Biosystems Division of Perkin Elmer, CA) using dye-labeled –21M13 primers. The reactions were then run on an ABI 373A following the manufacturers protocols.

The SPRI PCR method binds DNA based upon size as shown in Figure 1. This figure shows that for a 2 kb PCR product, the final yield is 80–90% whereas the yield from a PCR primer <50 nt in length is almost undetectable. We have shown previously that the lower limit at which yields in excess of 80% are achieved is 200 bp, the maximum limit is in excess of 200 kb (BAC DNA isolation).

Overall this solid-phase procedure is fast, simple and highly automatable. Over the past year, this method has been used to isolate >5000 PCR products for DNA sequencing, the majority of which have been purified on our robotic systems. As shown in Figure 2, the sequence data is of the highest quality, allowing the identification of single base pair polymorphisms.

ACKNOWLEDGEMENTS

We would like to thank Dr Chris Burns for supplying the magnetic particles. In addition we would like to thank the members of the DNA Sequencing group for help and support during this project. This work was supported by NIH and DOE funding.

REFERENCES

1 Uhlen,M. (1989) *Nature* **340**, 733–744.
2 Hultman,T., Stahl,S., Hornes,E. and Uhlen,M. (1989) *Nucleic Acids Res.* **17**, 4937–4946.
3 Hawkins,T.L (1992) *J. DNA Seq. Mapping* **3**, 65–69.
4 Hawkins, T.L. (1994) In Craig Venter J. (ed.) *Automated DNA sequencing and analysis.*
5 Hawkins,T.L., O'Connor-Morin,T., Roy,A. and Santillan,C. (1994) *Nucleic Acids Res.* **22**, 4543–4544.
6 Sibson, D.R. (1994) In Craig Venter J. (ed.) *Automated DNA sequencing and analysis.*

LRA, Vol. 7, pp. 117–122

Thermal Cycle DNA Sequence Setup Using a Modified Lab Workstation

T. L. Hawkins,† *S. R. Banerjee, C. Brodowski, F. Days, C. A. Evans, D. Levinson, and K. Ingalls*

Novel biochemical approaches and the modification of a commercially available robotic device has led to the development of a small flexible system that can perform the setup of thermal cycle DNA sequencing reactions in a high-throughput manner. The system is highly flexible without the need for large or expensive automation. Our results from using this small footprint robotic system open up the possibility of using this system for other molecular biology tasks.

INTRODUCTION

Widespread laboratory automation is set to be the next evolutionary step for molecular biology. Basic robotic workstations are becoming available although, at first inspection, their day one usefulness is rather limited (Kristensen et al. [1], Watson et al. [2], and Smith et al. [3]). Molecular biology projects such as the Human Genome initiative and associated model organism studies have led to the need for useful and effective automation. For our research, we utilize three basic molecular biology techniques: clone picking, DNA purification, and DNA sequencing. When looking at ways to automate these tasks, we failed to find an all-purpose commercially available system. Rather than designing our own system, which could be redundant and out of date by completion, we decided to purchase the most flexible *XYZ* system that would allow us to develop our applications. We started with a Tecan RSP 5032, which had a working area of 434 × 300 mm with two robotic arms. One arm had a single fixed tip, while the second had four fixed tips. All tips were capable of capacitance sensing.

Our aim was to provide our lab with a small, highly flexible, robotic system that could initially perform the task of DNA sequence setup to feed six ABI 373 DNA sequencers per day. The device needed to be usable by personnel without knowledge of robotics, programming, or biological processes being carried out. To achieve this goal, we made a number of significant modifications to the Tecan system. To aid the user, we wrote our own menu-driven soft-

*T. L. Hawkins, S. R. Banerjee, C. Brodowski, C. A. Evans, D. Levinson, and K. Ingalls are with Whitehead Institute/MIT Center for Genome Research, One Kendall Square, Bldg. 300, Cambridge, MA 02139. F. Days is with Tecan US, Research Triangle Park, NC 27709.

†Author to whom correspondence should be addressed.

CCC 0895-7533/95/030117-06

ware that allowed procedures to be called via a point and click environment. We also developed a 12-channel pipette tip to replace the Tecan's more standard four fixed tips, which dramatically decreased the time spent in liquid handling. To allow flexibility, we designed various workstation layouts for different procedures, the designs of which are all available free from the authors (e-mail: tlh@genome.wi.mit.edu).

This article describes the modification of the Tecan RSP 5032 robot to facilitate high-throughput DNA sequencing setup. Here we discuss the approaches used, the time taken, and the results from using the device.

MATERIALS AND METHODS

Tecan RSP 5032 Dimensions and Specifications

The Tecan RSP 5032 is a small footprint robotic workstation (Figure 1). The system has two arms: one with a single fixed tip and the other with 12 fixed tips (tip spacing 9 mm, own design). All tips are capable of liquid detection using capacitance sensing.

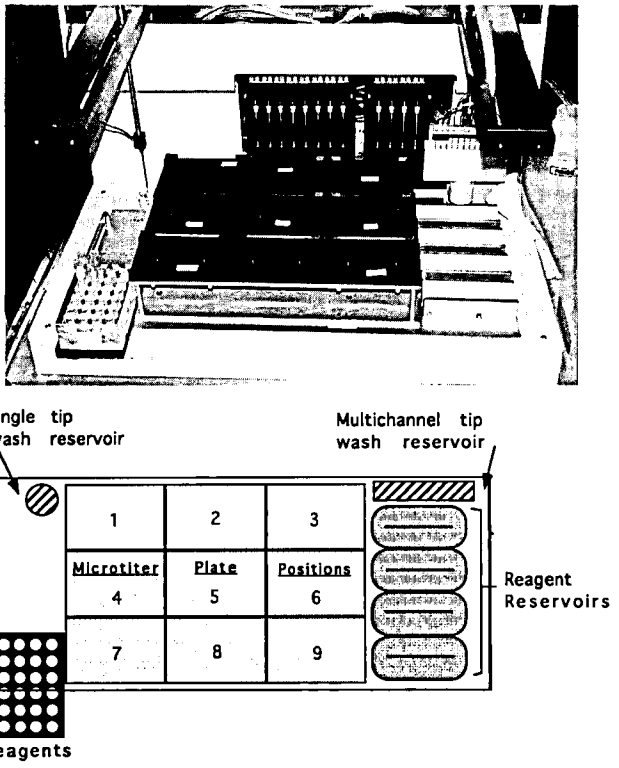


Figure 1. The Tecan microtiter plate layout, with the Cavro syringes shown at the back of the instrument in two banks of 8 and 4 syringes. Both arms can access the common 9 microtiter plate positions. The reagent tubes can only be accessed by arm 1 while the reagent reservoirs can only be accessed by arm 2.

12-Channel Modifications

The RSP 5032 normally utilized a single tip on one arm and a four tip on the other. In order to expand to 12 channels, we removed the internal Tecan diluters and replaced these with an eight-channel diluter Cavro XL 3000-8 with RS-232 communication (Cavro #724522) and a four-channel diluter Cavro XL 3000-4 (Cavro #724510). The two diluters were connected in parallel so that from the software point of view they were one unit. A Newark 24 V power supply was used to power the units. The Cavro units were different from the Tecan diluters in that one stepper motor drives all syringes in each unit. Each syringe input and output was controlled using different switch valves. This allowed no individual volume control but facilitated on/off switching of all channels. For the 12-channel device, this approach was suitable since our applications always used all 12 channels with identical volumes.

Time Parameters of the Modified 12-Tip Tecan

General

Purging time for syringes and tips: 20 seconds

Preparing duplicates of 96 samples into a microtiter plate aspirating 10 μL of sample: 2 minutes
- Transfer: 35 seconds
- Tip washing: 1 minute, 25 seconds

Dispensing 100 μL of reagent into microtiter plate: 20 seconds

Sequencing Setup

Transferring 50 μL of sample by aliquoting 10 and 15 μL volumes into cycle plates: 4 minutes, 30 seconds
- Transfer: 2 minutes
- Tip washing: 2 minutes, 30 seconds

Adding 10 μL dye-primer mix to a microtiter plate performed by the single tip: 10 minutes
- Mix addition: 9 minutes, 20 seconds
- Tip washing: 40 seconds

Software

The Tecan robot was available with INTEGRATOR, a PASCAL-like language. The compiler ran on a standard 486PC, though a 286 would have been suf-

TABLE 1. Tecan RSP 5032 Specifications

	Width (mm)	Depth (mm)	Height (mm)
Overall dimensions	770	647	500
Common workspace (two arms)	434	300	160
	X (mm)	*Y* (mm)	*Z* (mm)
Smallest addressable move	0.229	0.127	0.100
Maximum velocity	0.92 m/s	0.64 m/s	0.6 m/s

ficient. The INTEGRATOR package included low level procedures such as "movetopos (rack, position)" to which parameters could be passed in the normal manner. Racks such as microtiter arrays were defined using the INTEGRATOR teach mode. The teach mode allowed users to define positions such as maximum *z* displacement, dispensing height, and travel height. A series of racks (e.g., rectangular, circular, discrete) together comprised a layout.

We used the layout system of the INTEGRATOR to define specific areas of the common workspace for wash sites, positions of microtiter plates, DNA plates, pipette tip holders, reagents, etc. Specific racks were defined by a starting point (position 1), a final point (position *n*), and the number of columns and rows of points (rectangular rack) between the start and end points. The INTEGRATOR then calculated the separation distance between the wells and assigned each well a specific location in the rack. The rack system alleviated the problem of having to define well positions individually. In the case of 96-well microtiter plates, we defined the first well and the last well, as well as the number of rows (8) and columns (12).

The INTEGRATOR also offered the possibility of external device control through a standard serial port. Upon designating the auxiliary devices in the INTEGRATOR, we were able to directly send commands to any external devices via the programming environment. Because the software allowed us to manipulate any three external devices simultaneously, our system rapidly became amenable to expansion.

DNA Sequencing Chemistry Modifications

Taq Dye-Primer-Cycled Sequencing Reactions for the ABI 373A DNA Sequencer: Buffers and Solutions

Cycle buffer: 400 mM Tris-HCl pH 8.9, 100 mM ammonium sulfate, and 25 mM magnesium chloride.

Mix		Mix	
A:	62.5 μM dATP	C:	250 μM dATP
	250 μM dCTP		62.5 μM dCTP
	250 μM dGTP		250 μM dGTP
	250 μM dTTP		250 μM dTTP
	1.5 mM ddATP		0.75 mM ddCTP
Mix		Mix	
G:	250 μM dATP	T:	250 μM dATP
	250 μM dCTP		250 μM dCTP
	62.5 μM dGTP		250 μM dGTP
	250 μM dTTP		62.5 μM dTTP
	0.125 mM ddGTP		1.25 mM ddTTP

Primers: HPLC purified and made up to 1 pmol/μL.
DNA: single or double stranded at 15 ng/μL.
Taq polymerase: ca. 5 U/μL

Thermal Cycling

Reaction mix for 96 reactions. This required the following:

A Stock	C Stock	G Stock	T Stock
400 μL buffer	400 μL buffer	400 μL buffer	400 μL buffer
104 μL A mix	104 μL C mix	200 μL G mix	200 μL T mix
104 μL Taq mix	104 μL Taq mix	200 μL Taq mix	200 μL Taq mix
32 μL primer	32 μL primer	64 μL primer	64 μL primer
340 μL water	340 μL water	100 μL water	100 μL water

Taq mix: 120 μL Taq, 96 μL buffer, 480 μL water.

These mixes were placed on the robot which added the following to each well:

Reagent	A	C	G	T
Stock solution	10 μL	10 μL	10 μL	10 μL

Then the following was added using the 12-channel manifold:

DNA template	10 μL	10 μL	15 μL	15 μL

After completion, the cycle plates were removed from the Tecan, capped, and then placed on Techne GeneE thermal cyclers, which had heated lids to prevent evaporation. The cycle times for the M13 clones were as follows:

96°C, 30 seconds; 51°C, 1 second; and 72°C, 1 minute for 1 cycle followed by:

96°C, 1 second; 51°C, 1 second; and 72°C, 1 minute for 15 cycles followed by:

96°C, 1 second; and 72°C, 1 minute for 15 cycles.

RESULTS

Software Developments

The Cavro diluter units required independent control from the RSP workstation since the functions within INTEGRATOR were specific to the Tecan diluters. In our system, we retained the Tecan dual diluter for arm one and only used the Cavro systems for arm two. This required the writing of new functions within INTEGRATOR to control the Cavro diluters and to move the arm to correct positions when removing or adding liquids. These functions were added to the standard INTEGRATOR architecture.

Hardware Developments

To utilize the 12-channel system, we designed a suitable head and mounting assembly, as shown in Figure 2. The tips were stainless steel with siliconized internal and external surfaces and had an internal diameter of 1 mm. The main problems in assembly of this new part involved the elimination of torque on the head. This was achieved by adding a second *z* rod and fixing assembly. With these modifications, we were able to achieve reproducible positioning of all 12 tips.

The Cavro diluter units, fitted with 1 mL syringes, gave a working accuracy of +/−5% on volumes of 10 μL well within our required specifications. We were able to control volumes, ramp speeds, cutoff speeds, and individual switching to enable liquids of various viscosities to be dispensed. In all experiments, we found no edge effects or variations from tip to tip. Washing the tips involved flushing of the tip to a waste position followed by immersion of the tips into water and further dispensing of water through the tips. This had the effect of cleaning both the internal and external surfaces of the tips.

DNA Sequence Setup

The 12-channel system has been in full operation in this lab since September 1994 with over 25,000 DNA sequence reactions performed to date. The setup for the DNA sequencing process (Figure 3) required four cycling plates, the reaction mixes for A, C, G, and T, and the purified DNA presented in a 96-well plate. For our system, the DNA were M13 clones containing either human or mouse DNA that had been isolated using a magnetic particle purification process (Hawkins [4, 5]). The DNA sequencing chemistry employed was taq dye-primer sequencing that required 600–800 ng DNA and could be set up and then thermal cycled to provide linear amplification. This system could, however, be used for other types

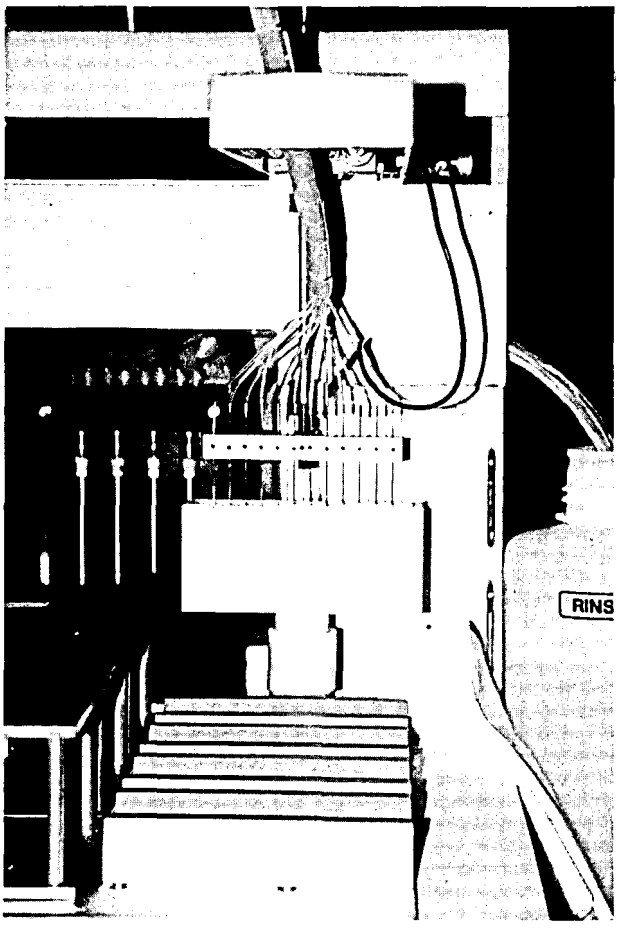


Figure 2. The 12-channel pipette head. The head was constructed from aluminum with stainless steel pins used as tips. To prevent torque, we fitted an additional *z* rod, which was secured to the *z*-actuator mounting. The 12-channel wash reservoir is also shown.

of DNA sequencing chemistry, such as primer and terminator approaches (Hawkins et al. [6]) and those employed by the Pharmacia ALF or the LiCor systems.

The robot took 14 minutes to set up 96 dye-primer sequencing reactions using four wells per reaction. This is limited by the addition of the reagent mixes by the single tip, which takes 10 minutes. We rejected the use of the 12 tip for mix addition since this would require time to aliquot mixes from tubes to plates to allow the 12 tip access. The setup of ABI style dye terminators would be much faster because only one well would be used per clone. In using the robotic system, we found that the overall quality of reactions increased compared to the manual system, especially because the robot required no further intervention until the cycle plates were removed and placed on thermal cyclers. We decided not to integrate thermal cyclers into the system since this would "lock up" the robot for many hours while the re-

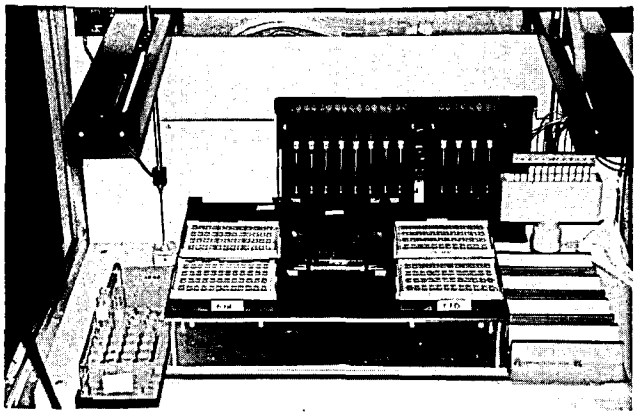


Figure 3. The DNA sequencing setup. Five of the available nine microtiter plate positions are occupied. Centrally, one can see the DNA plate presented as a flexible Falcon 9311 plate. The remaining four plates are composed of 0.2 mL thin-walled tubes. In the reagent position, the four sequencing mixes, A, C, G, and T are placed.

actions were being cycled. An example of the data quality from this approach is shown in Figure 4.

DISCUSSION

Using the modified RSP system, we can set up the 432 sequencing reactions, necessary to fill our 6 ABI sequencers (36 lanes × 2 runs/day), in less than 2 hours. The automation of the DNA sequencing process highlighted two key points in process automation. First, designing Peltier-driven heat blocks into the workspace would have allowed us to set up and cycle our sequencing reactions, however, this would have incapacitated the robot for several hours during thermal cycling. Second, pipetting small volumes is difficult to reproduce; therefore, it was easier to redesign the biochemistry than reengineer the robot. As with all process automation, there are limitations from both the biochemical and the engineering facets. In many ways, our tasks have been to expedite the biochemical techniques so that they are compatible with specifications of the robot, while not compromising the yield or the quality of the DNA products.

The features of the INTEGRATOR software such as the rack and layout system, the database, and external device control provided us a basic system from which to expand. The Tecan *XYZ* system furthermore furnished us with a flexible system that allowed us to customize our applications through its simple liquid handling capabilities and through the addition of the Cavro diluters.

The design of a small highly flexible system with both single- and 12-tip pipetting heads has ob-

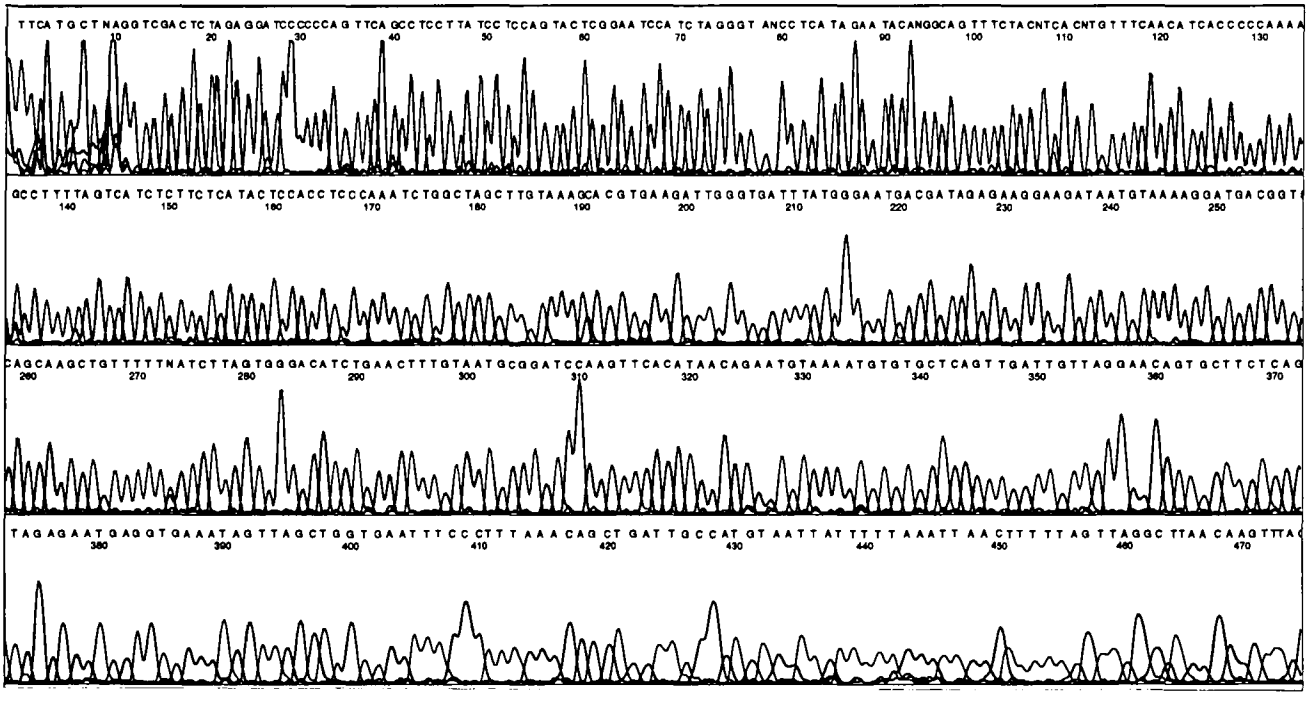


Figure 4. Shown here is an example of the data quality produced by samples set up on this device. The sequence chromatograph is from a taq dye-primer reaction, which was electrophoresed on an ABI 373A following the manufacturers protocols.

viously provided us with a unit to automate many other biochemical tasks. Currently under development are the automation of clone picking, PCR set up, and DNA purification using magnetic particles as a solid phase.

ACKNOWLEDGMENTS

The authors would like to thank Mr. Kais Arfaoui from Tecan US for his help in INTEGRATOR use and Cavro control. This work was supported by NIH Grant HG#000948.

REFERENCES

[1] T. Kristensen, H. Voss, and W. Ansorge, "A simple and rapid preparation of M13 sequencing templates for manual and automated dideoxy sequencing," Nucl. Acids Res., *15*, 5507–5516 (1987).

[2] A. Watson, N. Smaldon, R. Lucke, and T. Hawkins, "The Caenorhabditis elegans genome project: first steps in automation." Nature, *362*, 569–560 (1993).

[3] V. Smith, C. Brown, A. Bankier, and B. Barrell, "Semiautomated preparation of DNA templates for large-scale sequencing projects." J. DNA Seq. Mapping, *1*, 73–78 (1990).

[4] T. Hawkins, "M13 single-strand purification using a biotinylated probe and streptavidin coated magnetic beads." J. DNA Seq. Mapping, *3*, 65–69 (1992).

[5] T. L. Hawkins, "The use of custom magnetic particles in molecular biology," in J. Craig Venter (ed), Automated DNA Sequencing and Analysis (1986).

[6] T. Hawkins, Z. Du, N. Halloran, and R. Wilson, "Fluorescence chemistries for automated primer-directed DNA sequencing." Electrophoresis, *13*, 552–559 (1992).

DNA purification and isolation using a solid-phase

Trevor L.Hawkins, Tara O'Connor-Morin, Aparna Roy[1] and Cynthia Santillan
Whitehead Institute/MIT, Center for Genome Research, One Kendall Square, Cambridge, MA 02139 and [1]PerSeptive Diagnostics, 735 Concord Avenue, Cambridge, MA 02138, USA

Reprinted from **Nucleic Acids Research,**
Volume 22 No. 21 (1994) pp 4543–4544

Original article published by Oxford University Press,
Walton Street, Oxford OX2 6DP, UK.
Tel: +44 (0)1865 56767, Fax: +44 (0)1865 267773

DNA purification and isolation using a solid-phase

Trevor L.Hawkins*, Tara O'Connor-Morin, Aparna Roy[1] and Cynthia Santillan
Whitehead Institute/MIT, Center for Genome Research, One Kendall Square, Cambridge, MA 02139
and [1]PerSeptive Diagnostics, 735 Concord Avenue, Cambridge, MA 02138, USA

Received August 4, 1994; Revised and Accepted September 17, 1994

Preparation and manipulation of high quality DNA is a vital step in molecular biology. Although there are many methods reported for single and double stranded DNA isolations (1, 2, 3, 4, 5) there are few procedures that are rapid, low cost and procedurally identical for all DNA types, from plasmids to single copy BAC clones.

We have noted that under conditions of high polyethylene glycol (PEG) and salt concentration (10% PEG 8000 and 1.25 M NaCl final concentrations) (6), DNA would bind to the surface of carboxyl coated magnetic particles. Once bound, the DNA-bead complex could be extensively washed and finally eluted in water to yield purified DNA.

We have used carboxyl coated magnetic particles (Cat No. #8-4125B) available from PerSeptive Diagnostics (Cambridge, MA) for all our applications. These particles are 1 μM in diameter, are superparamagnetic and display iron as well as the carboxyl groups on the surface. These surface features are important since we have obtained reduced yields when either the iron or the negative charge are removed from the solid-phase surface.

We have adapted the procedure, called SPRI (solid-phase reversible immobilization) for use in all scales of template preparations and manipulations. As well as the standard mini, midi and maxi prep scales, we have also demonstrated the use of the SPRI technology in the extraction of DNA from agarose.

The method for the mini prep protocol is as follows:

1. Take 1 ml of overnight culture containing the plasmid clone in an Eppendorf™ tube.
2. Create a cleared lysate.
   Centrifuge for 2 minutes to pellet the cells.
   Pour off the supernatant and resuspend the pellet in 30μl Solution 1 (50mM Glucose, 25mM Tris.Cl pH 8, 10mM EDTA pH 8, 100μg/ml RNase).
   Add 60μl Solution 2 (0.2N NaOH, 1% SDS) and mix by shaking. Leave at room temperature for 5 minutes.
   Add 45μl Solution 3 (3M KOAc), mix by shaking and leave on ice for 10 minutes.
   Centrifuge for 10 minutes and remove 100μl of the supernatant to a new Eppendorf tube.
3. Take 10μl (@20mg/ml) carboxyl coated magnetic particles, wash three times in 0.5M EDTA pH 7.2 and resuspend in 10μl 0.5 EDTA. Add to the cleared lysate.
4. Add 100μl of the binding buffer (20% PEG 8000, 2.5M NaCl) and mix.
5. Allow to incubate at room temperature for 5 minutes.
6. Wash the magnetic particles twice with 5M NaCl and once with wash buffer (25mM Tris.Acetate pH 7.8, 100mM KOAc, 10mM Mg_2OAc, 1mM DTT). There is no need to resuspend the particles during each wash.
7. Resuspend the particles in 50μl water and incubate at room temperature for 1 minute.
8. Magnetically separate the particles and remove the DNA to a new tube.

From our results, shown in Figure 1, we estimate the initial binding efficiency to be approximately 100% under conditions in which the particles were in excess. Washing the solid phase then reduces this yield to approximately 80%, dependent on the types of wash solutions used. The solid-phase will bind to all types of DNA, which facilitates the same procedure to be applied to all DNA vectors and fragments. We have demonstrated the application of SPRI for the isolation of PCE products, M13 single stranded DNA, plasmids, cosmids and BACs with inserts up to 240Kb. The procedures for all these DNA isolations are identical.

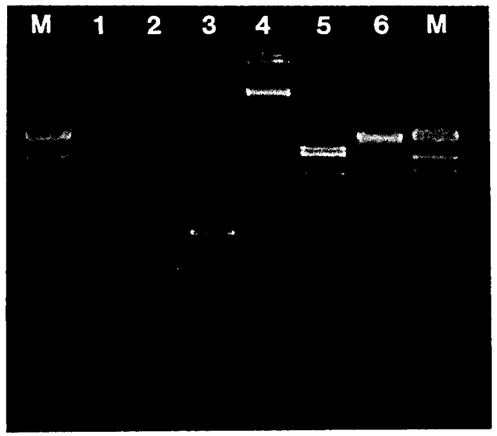


Figure 1. This gel shows the various types of DNA that can be isolated using the SPRI technique. M is a 200ng Lambda HindIII marker. Lane 1 shows the pUC18 supernatant after DNA binding to the solid-phase,demonstrating 100% binding. Lane 2 shows uncut pUC18, lane 3 SmaIA cut pUC18, lane 4 uncut pWE15 cosmid with 30kb insert, lane 5 cosmid cut with BamHI, lane 6 BAC clone with 240 Kb insert cut with NotI to excise the 7.2 Kb vector.

* To whom correspondence should be addressed

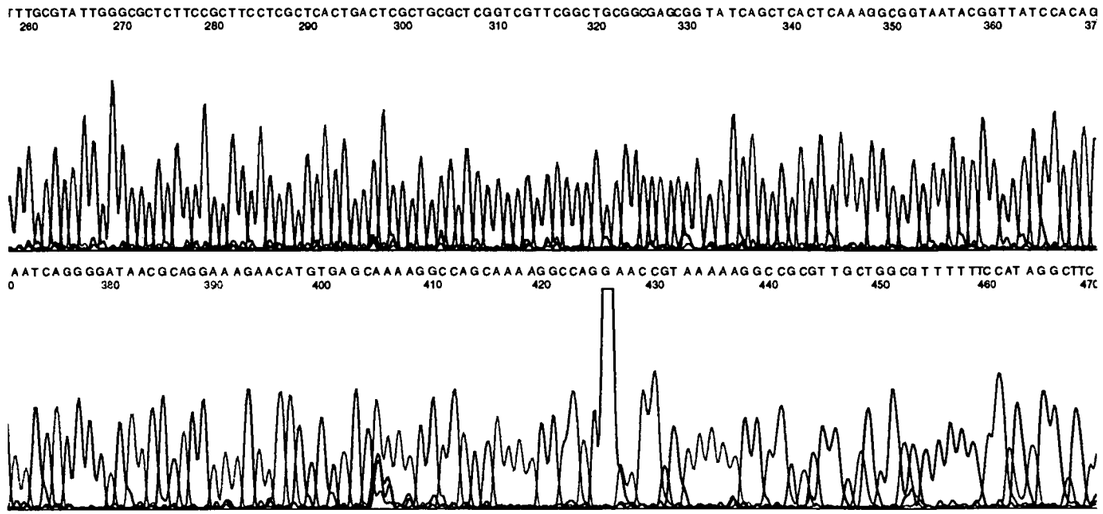


Figure 2. This figure shows an ABI DNA sequencing trace from a SPRI isolated pUC18 clone. Bases 260–470 are shown.

However, for M13 and PCR the procedure is further simplified since a cleared lysate is not necessary. Examples of various DNA isolations are also shown in Figure 1.

Our main interest in this application is the use of SPRI for high-throughput pUC isolations suitable for DNA sequencing. We have been using this approach to isolate plates of 96 clones in microtitre plate format in 1 hour which are then sequenced using the M13-21 universal dye-primer and electrophoresed on the ABI 373. The high quality of the sequence information, as shown in Figure 2, is typical of the data produced using this method.

This procedure allows the manipulation and isolation of all DNA types using a single protocol. The SPRI method is fast, low cost (approximately $0.15 per sample in solid phase costs) and amenable to automation. The key to the SPRI technology is its potential to standardize all DNA manipulations and isolations and to enable us to perform solid-phase biochemistry to encompass many steps in molecular biology.

ACKNOWLEDGEMENTS

We would like to thank Dr Chirs Burns for her help with magnetic particle characterization and Dr Bruce Birren for supplying BA-C clones. This work was supported by NIH grant HG #00098.

REFERENCES

1. Bankier, A., Weston,D. and Barrell,B. (1979). In Wu,R. (ed.) *Methods Enzymol.* **155**, 52–93.
2. Birnboim,H. and Doly,J. (1979) *Nucleic Acids Res.* **9**, 1513.
3. Ish-Horowicz,D. and Burke,J.F. (1981) *Nucleic Acids Res.* **9**, 2989.
4. Kristensen,T., Voss,H. and Ansorge,W. (1987) *Nucleic Acids Res.* **15**, 5507–5516.
5. Smith,V., Brown,C., Bankier,A. and Barrell,B. (1990) *J. DNA Seq. and Mapping* **1**, 73–78.
6. Lis,J. (1980) *Methods Enzymol.* **65**, 347–353.

11. Herrup, K. & Mullen, R. J. *Brain Res.* **227,** 475–485 (1981).
12. Herrup, K. *Brain Res.* **313,** 267–274 (1983).
13. Zanjani, H. S., Mariani, J., Delhaye-Bouchaud, N. & Herrup, K. *Brain Res. Dev.* **67,** 153–160 (1992).
14. Dietrich, W. F. *et al. Nature Genet.* **7,** 220–245 (1994).
15. Kusumi, K., Smith, J. S., Segre, J. A., Koos, D. S. & Lander, E. S. *Mamm. Genome* **4,** 391–392 (1993).
16. Becker-André, M., André, E. & DeLamarter, J. F. *Biochem. biophys. Res. Commun.* **194,** 1371–1379 (1993).
17. Giguere, V. *et al. Genes Dev.* **8,** 538–553 (1994).
18. Giguere, V., McBroom, L. D. & Flock, G. *Molec. cell. Biol.* **15,** 2517–2526 (1995).
19. McBroom, L. D., Flock, G. & Giguere, V. *Molec. cell. Biol.* **15,** 796–808 (1995).
20. Herrup, K. & Mullen, R. J. *Brain Res.* **172,** 1–12 (1979).
21. Hatten, M. E. & Heintz, N. *A. Rev. Neurosci.* **18,** 385–408 (1995).
22. Trenkner, E. & Hoffmann, M. K. *J. Neurosci.* **6,** 1733–1737 (1986).
23. Legrand, J. *Trends Neurosci.* **2,** 234–236 (1979).
24. Messer, A. *J. Neurochem.* **51,** 888–891 (1988).
25. Zou, L., Hagen, S. G., Strait, K. A. & Oppenheimer, J. H. *J. biol. Chem.* **269,** 13346–13352 (1994).
26. Tini, M., Tsui, L. C. & Giguere, V. *Molec. Endocr.* **8,** 1494–1506 (1994).
27. Forman, B. M. *et al. Molec. Endocr.* **8,** 1253–1261 (1994).
28. Tini, M., Fraser, R. A. & Giguere, V. *J. biol. Chem.* **270,** 20156–20161 (1995).
29. Forman, B. M., Umesono, K., Chen, J. & Evans, R. M. *Cell* **81,** 541–550 (1995).
30. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989).

ACKNOWLEDGEMENTS. We thank H. Sweet for sharing mice from the Jackson Laboratory Mouse Mutant Resource; E. Rinchik, M. Davisson and P. D'Eustachio for discussion and for sharing DNAs and data; K. Herrup for discussion and for examination of histological sections; J. Barker laboratory for a cDNA library, P. Lane, K.-F. Lee and A. Messer for discussion and advice; J. Lee, K. Devon and E. Sun for technical assistance; and C. Dunbar for animal care. B.A.H. was supported by a fellowship from the Helen Hay Whitney Foundation. This work was supported in part by grants from the National Center for Human Genome Research to E.S.L., L.K. and B.W.B., and by a grant from the National Institute of Neurological Disorders and Stroke and Klingenstein fellowship to W.N.F.

Disruption of the nuclear hormone receptor RORα in *staggerer* mice

Bruce A. Hamilton*, Wayne N. Frankel†, Anne W. Kerrebrock*, Trevor L. Hawkins*, William FitzHugh*, Kenro Kusumi*‡, Liane B. Russell§, Ken L. Mueller*, Victor van Berkel*, Bruce W. Birren*, Leonid Kruglyak* & Eric S. Lander*‡

* Whitehead Institute for Biomedical Research, Nine Cambridge Centre, Cambridge, Massachusetts, 02142, USA
† The Jackson Laboratory, Bar Harbor, Maine 04609, USA
‡ Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA
§ Biology Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831-8077, USA

Homozygous *staggerer* (*sg*) mice show a characteristic severe cerebellar ataxia due to a cell-autonomous defect in the development of Purkinje cells[1,2]. These cells show immature morphology, synaptic arrangement, biochemical properties and gene expression, and are reduced in numbers[3–12]. In addition, *sg* heterozygotes show accelerated dendritic atrophy and cell loss[13], suggesting that *sg* has a role in mature Purkinje cells. Effects of this mutation on cerebellar development have been studied for 25 years, but its molecular basis has remained unknown. We have genetically mapped *staggerer* to an interval of 160 kilobases on mouse chromosome 9 which was found to contain the gene encoding RORα, a member of the nuclear hormone-receptor superfamily. *Staggerer* mice were found to carry a deletion within the RORα gene that prevents translation of the ligand-binding homology domain. We propose a model based on these results, in which RORα interacts with the thyroid hormone signalling pathway to induce Purkinje-cell maturation.

We set out to identify the defective gene in *staggerer* by positional cloning, using genetic and physical mapping (Fig. 1). Apart from some large chromosomal deletions (such as Df 10FDFoD; Fig. 1), there is only one allele of *sg*. We mapped this allele relative to simple sequence length polymorphisms (SSLPs) from the Whitehead/MIT map[14] in 2,497 informative meioses. Marker loci that remained tightly linked to *sg* after 1,000 meioses were used to isolate clones from yeast artificial chromosomes (YACs)[15]. Genetic mapping localized *sg* to a 610-kilobase (kb) YAC and subsequently to a 162-kb bacterial artificial chromosome (BAC), 28A1.

In the course of sequencing clones from this region, we discovered that the RORα gene was at least partially contained in BAC 28A1 and a 135-kb BAC, 287E5. RORα is an orphan nuclear hormone receptor of the class that can bind DNA and activate transcription as a monomer in the apparent absence of ligand[16–19]. Although RORα has been extensively studied, its ligand, if any, remains unknown.

To study the genomic region in more detail, we used shotgun sequencing (with ninefold redundancy) to determine the nucleotide sequence of BAC 287E5. In addition, we isolated and sequenced a complementary DNA clone that contains the entire coding region for the mouse RORα1 isoform. A full description of these sequences will be presented elsewhere. Comparison of the cDNA and genomic sequences revealed that 1,370 base pairs (bp) of the 1,566-bp coding region is spread over ~35 kb of BAC 287E5 and provided the complete exon–intron structure of this portion of the gene. The 5′ end extends some 20 kb further, beyond BAC 28A1.

Because chimaera experiments have shown that *sg* affects Purkinje cells in a cell-autonomous fashion[12,20], we expected that the *sg* gene should be expressed in cerebellar Purkinje cells in normal mice and should be altered in either its expression or coding sequence in the mutant. To evaluate RORα by these criteria (Fig. 2), we first hybridized northern blots to a probe from the hinge and ligand-binding domains of RORα. Of three transcripts seen in wild-type cerebellum (~10.5, 6.5 and 2.4 kb), the larger two were of approximately normal length but greatly reduced abundance in *sg/sg* cerebella. Analysis using reverse

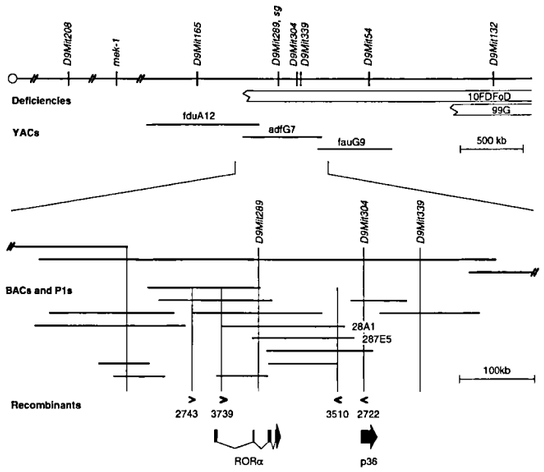


FIG. 1 Genetic and physical map of the *staggerer* locus. Top line, genetic map based on 2,497 informative meioses typed for SSLP markers and an SSCP in the *mek-1* gene. A chromosomal deficiency that removes sg (10FDFoD) and one that does not (99G) are shown as open bars. A minimum tiling path of YACs used as a scaffold for physical mapping and the full complement of BAC and P1 clones used for high-resolution mapping and sequencing are shown as horizontal lines. The positions of DNA polymorphisms used as genetic markers are indicated by vertical lines; meiotic recombinants are indicated by an arrowhead and the corresponding animal ID number at the bottom of the figure. Arrows indicate the genes for RORα and calpactin p36 (p36), which was also found to lie nearby but outside the sg interval.
METHODS. Genetic mapping was performed in two crosses, B6CWD-sg/+ × MOLF/Ei and C57BL/6J-sg/+ × CAST/Ei. In the first cross, a single sg heterozygous male from an earlier mapping cross (C57BL/6J-sg/+ × CWD/Le) was mated with several MOLF/Ei females. Progeny were genotyped and sg heterozygotes were mated with each other as an intercross or with B6CWD-sg/+ partners as a backcross. In the second cross, C57BL/6J-sg/+ animals were mated with CAST/Ei and the resulting sg heterozygotes were intercrossed. Marker order was inferred by minimizing double recombinants and verified (or refined, for markers between which no recombinations were observed) by physical map STS content. STSs from the ends of YACs were isolated by inverse PCR. Additional STSs were obtained from YACs by sequencing of random *Sau*3AI and *Eco*RI subclones made from CHEF gel-isolated YAC DNA. P1 clones were obtained from Genome Systems (St Louis) using PCR assays. BAC clones (B.W.B. unpublished results) were identified by a combination of PCR assays and oligonucleotide hybridization. Ends of physical mapping clones were isolated by inverse PCR and sequenced to provide additional STS coverage.

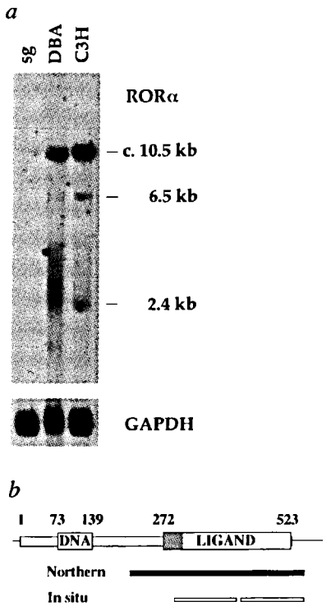


FIG. 2 Evaluation of RORα as a candidate gene for *staggerer*. *a*, Northern blot hybridization shows that RORα transcripts are expressed in adult *staggerer* cerebellum at low levels relative to wild-type controls. The blot was stripped and rehybridized with GAPDH as a control for the amount of RNA loaded in each lane. *b*, RORα gene product, showing the positions of the DNA-binding domain and ligand-binding domain homology. The line indicates sequences used to probe the northern blot shown in *a*; shaded bar indicates sequence missing from sg transcripts. Numbered amino-acid positions refer to the mouse RORα1 isoform. *c*, Key sequence features of the region deleted in sg. Exons are shown in capital letters, introns in lower case. Omitted sequences are represented by ellipses. Deletion breakpoints are marked with arrows; only one copy of the underlined sequence GCTA is retained in *sg*. Wild-type coding sequence translation is shown above the nucleotide sequence of the exons, and the shifted frame translation of the predicted *staggerer* protein is given below the nucleotide sequence. Conceptual translation of the *sg* RORα coding sequence from both genomic and RT-PCR sequencing predicts a frameshift at amino-acid 273 of the α1 isoform, leading to a truncation 27 residues later. Sequence of PCR fragments from sg, C3H/HeJ, BALB/cJ and subclones from a 129/Sv-derived BAC demonstrates that the sequence deleted from the RNA corresponds to a single 122-bp exon removed by a 6.5-kb genomic deletion.
METHODS. RNA was isolated from frozen tissue with the Trizol reagent and polyadenylated RNAs selected by oligo(dT)–cellulose chromatography. For northern blots, 3 μg poly(A)$^+$ RNA were treated with glyoxal, electrophoresed through 1% agarose gels and transferred to nylon membranes as described[30]. For RT-PCR, 1 μg poly(A)$^+$ RNA was reverse-transcribed and 1:100 of the resulting cDNA was used for each amplification. For sequencing, PCR products were reamplified using chimaeric primers carrying M13 sequencing primer sites and gel-isolated before cycle sequencing with fluorescently labelled primers.

c

```
A  D  S  A  V  S  S  F  Y  L  D  I  Q  P  S  P  D  Q  S  G
GCCGACTCAGCCGTCAGCAGCTTCTACCTGGACATCCAGCCCTCCCCAGACCAGTCGGGA

L  D  I  N  G  I  K  P  E  P  I  C  D  Y  T  P  A  S  G  F
TTGGACATCAATGGGATCAAACCCGAACCCATATGTGACTACACACCAGCATCTGGCTTC

F  P  Y  C  S  F  T  N  G  E  T  S  P  T  V  S  M  A  E  L
TTCCCCTACTGTTCCTTCACCAACGGAGAGACTTCCCCAACCGTGTCCATGGCAGAACTA

E
Ggtaaggcagtatgcagtgctcctcctgg.............gccttaccccagcacatg
A
gtattttg|gctaacagtgcacaaatccatgactagggct.............tgatct

                    H  L  A  Q  N  I  S  K  S  H  L  E  T  C
cctgttttctccccagAACACCTTGCCCAGAACATATCCAAATCCCACCTGGAAACCTGC

Q  Y  L  R  E  E  L  Q  Q  I  T  W  Q  T  F  L  Q  E  E  I
CAGTACTTGCGGGAAGAGCTCCAGCAGATAACGTGGCAGACCTTCCTGCAGGAGGAGATT

E  N  Y  Q  N  K
GAAAACTACCAGAACAAGgtagagttggaggacttttctcttccca.............

ctccattccatcaccttcaaactgtaa|gctaggaatctcagtaggcagaaaggatacgg

aagcaggcttttataacctaagtaacagagactatttttttttaatcctataatctttttt

taacccatttccaaagaaagctttttacaaactaggaagccagcccgaggctggctgcac

                                                Q  R  E  V  M
cagggttcccagctagtcttatgctaatgtgccttcctctttcacagCAGAGAGAGGTGA
                                                   E  R  G  D

  W  Q  L  C  A  I  K  I  T  E  A  I  Q  Y  V  V  E  F  A  K
TGTGGCAGCTGTGTGCCATCAAGATTACAGAAGCTATCCAGTATGTGGTGGAGTTTGCCA
  V  A  A  V  C  H  Q  D  Y  R  S  Y  P  V  C  G  G  V  C  Q

  R  I  D  G  F  M  E  L  C  Q  N  D  Q  I  V  L  L  K  A
AACGCATTGATGGATTTATGGAGCTGTGTCAAAATGATCAAATTGTGCTTCTAAAAGCAG
  T  H  *
```

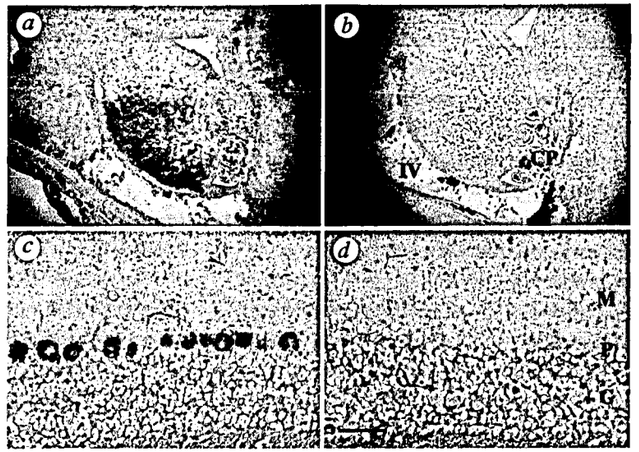



FIG. 3 *In situ* hybridization of RORα to sagittal sections of E15 mouse embryos and coronal sections through adult mouse hindbrain. The positions of the probe sequences within the RORα gene are shown in Fig. 2*b*. *a*, c, RORα antisense probe; *b*, *d*, sense probe. *a*, *b*, Hybridization signals detected in E15 cerebellum. The fourth ventricle (IV) and choroid plexus (CP), which lie below and posterior to the cerebellum, are marked for orientation. Anterior is to the left. Bar, 200 μm. c, *d*, Hybridization to the Purkinje cell layer detected in coronal sections through adult cerebellum. Letters denote the molecular (M), Purkinje cell (P) and granule cell (G) layers. Bar, 50 μm.
METHODS. Two cDNA fragments ~ 500 bp were used for each sense and antisense probe. cDNA corresponding to the RORα constant region were amplified by RT-PCR (primers 6680U.2: AGTTGGTCGGATGTCCAAG and rora.2: GGCTGCAGAAATGCCTGG), reamplified with internal chimaeric primers bearing priming sites for dye-primer sequencing, gel-purified and sequenced. Sequenced fragments were reamplified with chimaeric primers, either with the antisense strand carrying a T7 RNA polymerase promoter (T7rora.8: TGTAATACGACTCACTATAGGGCGAG-TCAAAGGCACGGCAC or T7rora.6: TGTAATACGACTCACTAT-AGGGCGATCTAGAAGTGCTCGG- GCG), or with the sense strand carrying a T3 RNA polymerase promoter (T3rora.3: AATTAACCCTCACTAAA-GGGCGACTCAGCCGTCAGCAG or T3rora.7: AATTAACCCTCACTAAAGGGA-CTTTGACGGGAAGTATGCG), and gel-isolated. RNA probes incorporating 11-digoxygenin–UTP were synthesized with the appropriate polymerase. Hybrid-ready tissue sections (from Novagen) were processed according to the protocol of A. Nieto (Mill Hill), and visualized with an alkaline-phosphatase-conjugated anti-digoxygenin antibody using 5-bromo-4-chloro-3-indolyl phosphate and nitroblue tetrazolium substrates.

transcription and the polymerase chain reaction (RT-PCR) revealed a 122-bp deletion, which removes the start of the ligand-binding homology domain and shifts the reading frame, causing a stop codon after 27 amino acids. Comparison with the genomic sequence showed that the deletion corresponds to a single exon.

Southern blot and long-range PCR analysis revealed a 6.5-kb genomic deletion relative to several wild-type chromosomes. The precise deletion breakpoints were localized by sequencing genomic PCR products (Fig. 2). The deletion is not present on C3H/HeJ chromosomes, on which *sg* apparently arose. (Specifically, *sg* arose in F_2 progeny of a non-inbred V-*ob*/*ob* male and a C3H/He × BALB/cHm female (P. W. Lane, personal communication). Of existing strains related to these progenitors, the *sg* chromosome most closely resembles C3H/HeJ, having identical allele sizes at 16 SSLPs within 1 cM of *sg*.)

The deletion would be predicted to create a null or severe hypomorphic allele, because truncation of the ligand-binding homology domain of RORα has been shown to preserve DNA binding but greatly to reduce or abolish transactivation activity of the unliganded receptor *in vitro*[19]. The predicted loss-of-function is consistent with our observation that *sg* behaves genetically as a null allele, with the behavioural and histological phenotype in *sg*/Df 10FDFoD mice being no more severe than in C57BL/6J-*sg*/*sg*.

We examined the pattern of RORα RNA expression in mouse embryos and adult brain to determine whether it is consistent with the cell-autonomous Purkinje cell defects seen in young homozygotes and the accelerated atrophy seen in heterozygotes (Fig. 3). Purkinje neurons arise from a proliferative zone above the fourth ventricle beginning on embryonic day 13 (E13) in normal mice and migrate along radial glia from E14 to E17; they form a laminar shell in the presumptive cerebellar cortex and position themselves in a monolayer during postnatal development[21]. *In situ* hybridization to sagittally sectioned embryos reveals prominent expression in large cells of the cerebellar anlage by E14. The position and size of the RORα-positive cells in embryonic cerebellum is consistent with Purkinje cell precursors. In addition, hybridization to coronal sections through adult hindbrain reveals expression only in the Purkinje cell layer. Prominent hybridization is also detected in embryonic midbrain, with weaker signals in thymus, whisker follicles, eye, lung, kidney tubules, gut. The expression in thymus is particularly interesting as *sg*/*sg* mice have been reported to have delayed thymic development and a defect in terminating T-cell responses[22].

The data provide compelling evidence that the deletion in RORα is the cause of the *sg* defect: the deletion lies in the small (162-kb) region containing *sg*, is not found on the presumed ancestral chromosome (or in 20 other strains examined) and eliminates a domain required for transcriptional activity *in vitro*, and the gene is expressed in Purkinje cells at the expected time in development. Ultimate proof will require transgenic rescue.

Identification of the *sg* gene as a nuclear hormone receptor is particularly informative given the key role that thyroid hormone (TH) plays in cerebellar development. Hypothyroidism causes reduced dendritic arborization of Purkinje cells and decreased granule-cell proliferation similar to that seen in *sg* mutants, whereas TH replacement alleviates these defects in a dose-dependent fashion[23]. Interestingly, the *sg* mutation blocks Purkinje cell response to TH[24]. One apparently direct target of TH action is the Purkinje cell protein-2 (*pcp-2*) gene, which responds to TH levels *in vivo* and which is activated by ligand-bound TH receptor-β (TRβ) in transfection assays[25]. Accordingly, we used RT-PCR to test whether *sg* affects expression of either TRβ or *pcp-2* RNA. We found that *pcp-2* expression is undetectable in *sg* mutants, despite significant levels of TRβ expression (not shown). Taken together, this strongly suggests that RORα interacts with a thyroid-hormone signalling pathway in the maturation of cerebellar Purkinje cells. Such a model would be consistent with recent examples of crosstalk among nuclear hormone receptor signalling pathways[26–29]. Further elucidation of the RORα signalling pathway should provide new insights into these regulatory interactions and the morphological, synaptogenic and trophic events that are defective in *staggerer*. □

Received 6 December 1995; accepted 11 January 1996.

1. Sidman, R. L., Lane, P. W. & Dickie, M. M. *Science* **137,** 610–612 (1962).
2. Herrup, K. & Mullen, R. J. *Brain Res.* **178,** 443–457 (1979).
3. Sotelo, C. & Changeux, J.-P. *Brain Res.* **67,** 519–526 (1974).
4. Sotelo, C. *Adv. Neurol.* **12,** 335–351 (1975).
5. Crepel, F., Delhaye-Bouchaud, N., Guastavino, J. M. & Sampaio, I. *Nature* **283,** 483–484 (1980).
6. Landis, D. M. & Sidman, R. L. *J. comp. Neurol.* **179,** 831–863 (1978).
7. Mariani, J. & Changeux, J.-P. *J. Neurobiol.* **11,** 41–50 (1980).
8. Hatten, M. E. & Messer, A. *Nature* **276,** 504–506 (1978).
9. Trenkner, E. *Nature* **277,** 566–567 (1979).
10. Edelman, G. M. & Chuong, C. M. *Proc. natn. Acad. Sci. U.S.A.* **79,** 7036–7040 (1982).

GENOMICS 27, 306–311 (1995)

41 Kilobases of Analyzed Sequence from the Pseudoautosomal and Sex-Determining Regions of the Short Arm of the Human Y Chromosome

L. SIMON WHITFIELD,*,[1] TREVOR L. HAWKINS,[2] PETER N. GOODFELLOW,* AND JOHN SULSTON†

*Department of Genetics, University of Cambridge, Downing Street, Cambridge CB2 3EH, United Kingdom; and †Sanger Centre, Hinxton Hall, Hinxton, Cambridgeshire, CB10 1RQ, United Kingdom

Received December 9, 1994; accepted March 17, 1995

Determination of 41.2 kb of Y chromosome genomic sequence has been made from a cosmid that spans the Yp pseudoautosomal boundary and includes 18.5 kb of sequence from the patient-defined sex-determining region of the Y chromosome. An AceDB database of the sequence and the analysis data have been produced as a resource for studies of the evolution and population genetics of the Y chromosome. Comparison of the 18.5 kb from the sex determining region to the sex determining region of mouse does not locate any areas of similarity outside *SRY/Sry*. Indeed, no coding regions other than those previously reported can be detected anywhere in the 41 kb. The Y-specific and pseudoautosomal portions of this sequence have different repeat sequence and GC contents: this may have relevance both to the events defining the pseudoautosomal boundary and to the course of sequence evolution in the absence of recombination.

INTRODUCTION

Different portions of the human Y chromosome have different modes of inheritance. The regions at the telomeres of the long and short arms are pseudoautosomal in character; the telomeric 2.6 Mb of Yp and 0.4 Mb of Yq are identical to the corresponding regions of Xp and Xq, respectively, and these regions recombine during male meiosis (Simmler *et al.*, 1985; Goodfellow *et al.*, 1987; Freije *et al.*, 1992). The remainder of the Y chromosome normally does not exchange with the X chromosome, and this portion remains male-specific. The lack of recombination and unique mode of inheritance of these sequences are associated with their startling lack of polymorphism within *Homo sapiens*. Conventional restriction fragment length polymorphisms (RFLPs) are found on the Y chromosome at a many-fold lower frequency than they are found on autosomes (Jakubiczka *et al.*, 1989; Malaspina *et al.*, 1990; Spurdle and Jenkins, 1992) although Y-linked hypervariable loci do exist (Jobling, 1994). There may be, therefore, a marked difference in the levels of polymorphism between adjacent regions on either side of the pseudoautosomal boundary. Studies of Y-chromosome evolution and population genetics are hampered not only by the lack of Y-specific sequence polymorphism but also by the lack of Y-chromosome sequence, a raw material for all manner of such studies.

Aberrant X–Y recombination events can transfer Y-specific material onto the X chromosome. A zygote containing the product of such an event can give rise to a sex-reversed "XX male" if the transferred Y-chromosome material includes the Y-borne component of the male sex-determining pathway. The study of such XX males enabled this sex-determining locus to be mapped to the 35 kb of Y chromosome immediately proximal to the Yp pseudoautosomal boundary. The region was cloned in a cosmid and phage "walk" beginning at the pseudoautosomal gene *MIC2* (Ellis *et al.*, 1989). Cosmid cAMF3.1, isolated during this walk, lies across the pseudoautosomal boundary and contains 18.5 kb of the patient-defined sex-determining region. The DNA used in the cosmid library was that of an Italian Caucasian. The single-exon gene *SRY* was located in a Y-specific subclone of this cosmid. *SRY* lies approximately 5 kb from the pseudoautosomal boundary with the 3′ end closer to the boundary (Sinclair *et al.*, 1990). Although a growing body of evidence suggests that *SRY* is necessary and sufficient for the sex-determining function of the Y chromosome (Goodfellow and Lovell-Badge, 1993; Hawkins, 1994), it remains formally possible that the 35 kb of Y chromosome proximal to the pseudoautosomal boundary contain other loci involved in sex-determination.

The mouse sex-determining region has also been defined. Chromosomally female mice transgenic for a 14-kb fragment of the Y chromosome, which contains mouse *Sry*, are phenotypically male (Koopman *et al.*,

[1] To whom correspondence should be addressed. Telephone: (44) 1223 333986. Fax: (44) 1223 333992.
[2] Current address: Whitehead Institute/MIT, Center for Genome Research, Cambridge, MA 02139.

0888-7543/95 $6.00

1991). *Sry* does not map close to the pseudoautosomal boundary of the mouse Y chromosome.

The Y-chromosome location of cAMF3.1 is shown in Fig. 1. cAMF3.1 contains, in addition to *SRY*, exons 2 and 3 of *XG*, which encodes the X-linked blood group antigen Xg_a (Mann *et al.*, 1962; Ellis *et al.*, 1994a,b). *XG* is truncated on the Y chromosome by the pseudoautosomal boundary (Weller *et al.*, 1994).

In this paper we report the sequencing and analysis of cosmid cAMF3.1. The analysis has been performed with two aims; first, to discover whether this portion of the sex-determining region contains genes other than *SRY*, and second, to identify organizational, structural, or functional features that might either illuminate the evolution of the pseudoautosomal boundary region or provide comparative markers for mapping these evolutionary events in primates.

MATERIALS AND METHODS

Shotgun sequencing of cAMF3.1. Details of the shotgun sequencing strategy can be found in Wilson *et al.* (1994) and references therein. In outline the procedure was as follows. cAMF3.1 was shotgun cloned into m13mp18. M13 minipreps were prepared for approximately 800 clones as per Hawkins (1992). These were sequenced by linear amplification dideoxy termination cycle sequencing using dye-labeled M13 forward primers. Electrophoresis was performed on Applied Biosystems 373A sequencers. Sequences of M13 clones were assembled and edited into the sequence of cAMF3.1 using the STADEN package (Dear and Staden, 1991; Gleeson and Staden, 1991). Security against misplacement of *Alu*-containing sequences was achieved by identifying *Alu* repeat regions prior to assembly so that assembly based solely on *Alu–Alu* sequence matches could be carefully examined. Initially the sequences from 800 random clones (of average read length 271 bases) resolved into 20–25 "contigs."

The contigs were joined and the sequence was completed on both DNA strands using a combination of strategies: (i) primer walking in individual M13 clones, sequencing with dye-labeled dideoxy analogues (Applied Biosystems) rather than dye-labeled primers; (ii) sequencing the opposite end of the insert of appropriate M13 clones via a PCR amplification product of the single-stranded clone; (iii) resequencing and electrophoresing samples on a stretch-liner ABI 373 sequencer to generate longer reads. Additional measures required to sequence a minisatellite region in cAMF3.1 are described in the next section.

Cloning of PCR products and generation of DNase digestion subclones. The sequence of the minisatellite region could not be obtained from the 800 random M13 clones. A PCR product containing this region was therefore cloned, and nested subclones were generated by DNase digestion (Jones and Sulston, 1995). The PCR product was treated with mung bean nuclease to generate blunt-ended fragments that were ligated into the plasmid pSC. The recombinant plasmid was linearized with a single cut with I-Sce meganuclease (Boehringer Mannheim), and the end of the insert was made susceptible to Exonuclease III digestion by the ligation of oligonucleotides. Brief Exonuclease III digestion was used to shorten the insert before recircularization and retransformation of the plasmid. The sequence of the portion of insert newly adjacent to the M13 primer site was then obtained. *Taq* polymerase cycle sequencing of this region proved difficult with any template.

Computer-aided sequence analysis. BLASTN (B = 1,000,000) was used to compare nucleotide sequences (Altschul *et al.*, 1990). BLASTX (B = 1,000,000, S = 50, M = BLOSUM62-12) was used to produce hypothetical six-frame translations of nucleotide sequences and compare these to protein sequences (Altschul *et al.*, 1990; Gish and States, 1993). The sequence of cAMF3.1 was compared to three sequence databases to identify coding regions and other features of interest: EMBL release 38, SWIR version 5 (SWIR 5 is a Sanger Centre nonredundant compilation of SWISSPROT 28, PIR 39, and WORMPEP 5), and dbEST version 2.1. Prior to the execution of database comparisons, human repeat sequence family members were identified (using the database assembled by J. Jurka, available from the PYTHIA server (Milosavljevic and Jurka, 1993a)) and replaced with the character "N". A hidden Markov model (Krogh *et al.*, 1994) was used to identify *Alu* elements (Micklem and Eddy, unpublished. Software available by anonymous ftp from /pub/sre at cele.mrc-lmb.cam.ac.uk.), while BLASTN was used to identify LINE and MER sequences. The program ALUS from the PYTHIA package (Milosavljevic and Jurka, 1993a, 3b) was used to assign *Alu* elements to subfamilies. All maximal segment pairs produced by BLAST were screened with MSPcrunch to identify significant matches as previously defined (Sonnhammer and Durbin, 1994). Tandem repeats were detected using the programs QUICKTANDEM and TANDEM (R. Durbin, unpublished).

To detect regions with the characteristics of coding sequence, cAMF3.1 (without repeat elements masked) was examined using GRAIL II, an artificial intelligence-based technique trained with non-Y chromosome sequences (Xu *et al.*, 1994a,b). All GRAIL II-predicted exons were entered into AceDB rather than just its "final predictions." The program CPG (Micklem and Durbin, unpublished) was used to identify "CpG islands," regions in which the CG dinucleotide is as uncommon as it normally is in the genome of mammals and that are generally located within or near genes (Bird, 1986).

The sequence and the outputs of all analysis programs were entered into AceDB (Durbin and Thierry-Mieg, unpublished) for cross-comparison, storage, and presentation.

RESULTS AND DISCUSSION

Sequencing of cAMF3.1

From cAMF3.1, 41,155 bases of Y chromosome genomic sequence were obtained with 7.95-fold coverage per consensus base. The sequence of both DNA strands has been completely determined, except for the central 796 bp of a minisatellite repeat for which only one strand has been sequenced.

An estimate of the sequencing error frequency can be made because 5684 bp of sequence from within the region covered by cAMF3.1 have been previously reported (EMBL Accession Nos. HSSRYZ and HSPABY01). Comparison of these two sequences to cAMF3.1 identified an incorrectly called base in the sequence of cAMF3.1. One basepair error in 5684 bp suggests approximately 99.98% sequencing accuracy. This is in line with estimates of error frequency published by large-scale sequencing projects (~99.97% for *Saccharomyces cerevisiae* chromosome XI (Dujon *et al.*, 1994)).

Sequence Analysis: The Search for Novel Coding Regions

The first aim of the sequence analysis was to discover whether the sequence of cAMF3.1—in particular the portion from the sex-determining region—contains any coding regions not previously reported.

The 18.5 kb of sequence from the sex-determining region were compared to the 14 kb of the mouse sex-determining region in both untranslated and translated forms to identify sequences common to both re-

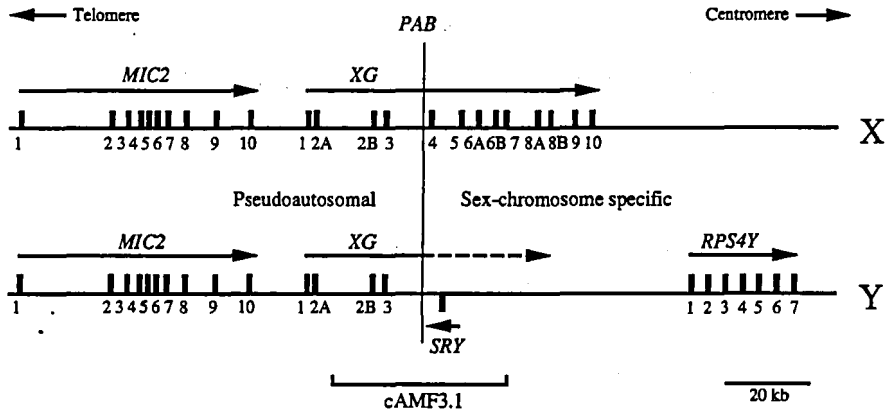


FIG. 1. The region of the Y chromosome covered by cAMF3.1 is depicted with respect to the pseudoautosomal boundary, PAB, the sex-determining gene, *SRY*, the gene for the Xga blood group antigen, *XG*, and *MIC2*. Exons are not to scale.

gions. First, the nucleotide sequences were compared using BLASTN; second, cAMF3.1 was translated in all six reading frames before BLASTX was used to similarly translate the mouse sequence and compare the two. The only matches between the human and the mouse sequences were either in the high mobility group (HMG) domains of *SRY* and *Sry* or, as expected, between low-complexity repetitive sequences. Matches were seen at the nucleotide level and at the amino acid level.

Coding regions were also sought by searching protein and nucleotide sequence databases, by using GRAIL II, and by identifying CpG islands. The assessment of candidate coding regions was made by simultaneous cross-comparison of these data in AceDB as shown in Fig. 2. Figure 2 is derived from the AceDB display.

The three pieces of coding sequence and one pseudogene known to exist within cAMF3.1 (see Figs. 1 and 2) acted as internal controls for gene-finding. The single 612-base exon of *SRY* lies 5 kb from the pseudoautosomal boundary on the bottom DNA strand as depicted in Figs. 1 and 2. Exons 2 (41 bp) and 3 (23 bp) of *XG*, the gene encoding the Xg_a blood group antigen (Ellis *et al.*, 1994a,b), lie in the pseudoautosomal portion of cAMF3.1 on the top strand. The pseudogene reported by Behlke *et al.* (1993) (hereafter referred to as T6) lies 3 kb from *SRY* toward the pseudoautosomal boundary but on the opposite DNA strand (see Fig. 2).

Y-specific sequence. *SRY* is detected by GRAIL II although the predicted gene is shorter at the 3′ end than the actual coding sequence. A CpG island and two high-scoring GRAIL II predicted exons fall within the region of the pseudogene T6. It is notable that an RT-PCR product has been found in which a region encompassing the GRAIL II hits and the CpG island (shown in Fig. 2) is spliced onto the pseudoautosomal exons of the X chromosome gene *XG* (Weller *et al.*, 1994). This transcript is not thought to be functional. A weaker GRAIL II-predicted exon is found in the same position as the strong hits on T6 but on the opposite DNA strand. There are no other predicted exons in the Y-specific portion of cAMF3.1.

BLASTX detects one region of Y-specific sequence other than *SRY* that, when translated, matches any SWIR entry. The score is marginally above the MSPcrunch cutoff score and the sequence contains stop codons, so no further characterization has been attempted. BLASTN detected many similarities between cAMF3.1 (involving either Y-specific or pseudoautosomal sequence) and DNA sequence database entries, but these matches were generally to noncoding portions of entries, many of which, if not already known to be repetitive areas, seemed to be of low sequence complexity. As expected, therefore, the vast majority of the data accrued from nucleotide database searches was not useful for gene searching. The BLASTN search does indicate, however, that two Y-specific sequence tagged sites, EMBL entries HSPH09R and HSPH09F, come from the region covered by cAMF3.1. They are close to one another, 7.5 kb proximal to *SRY* (see Fig. 2).

Pseudoautosomal sequence. GRAIL II successfully detects only exon 2 of *PBDX/XG;* the 23-bp exon 3 is missed. Low-confidence prediction of coding sequence is made for five other short pseudoautosomal regions. The only prediction deemed "good" by GRAIL II is made for the bottom DNA strand, although a prediction is made for the same place on the other strand; a top-strand exon could be part of *XG*. The other predicted exons are low scoring and lie in areas for which there is no other evidence for coding sequence. They have not been further analyzed. The higher number of predicted exons in the pseudoautosomal sequence may reflect its higher GC content.

The only region of the pseudoautosomal portion of cAMF3.1 for which BLASTX detects a similarity to a protein database entries seems to be repetitive, giving a proline-rich sequence when translated. For this reason this sequence also has not been analyzed further.

In summary, it seems that there is no strong evidence for any coding sequence in cAMF3.1 other than that previously reported, although this conclusion can and should be reviewed as gene-finding techniques improve and sequence databases expand. cAMF3.1 is cur-

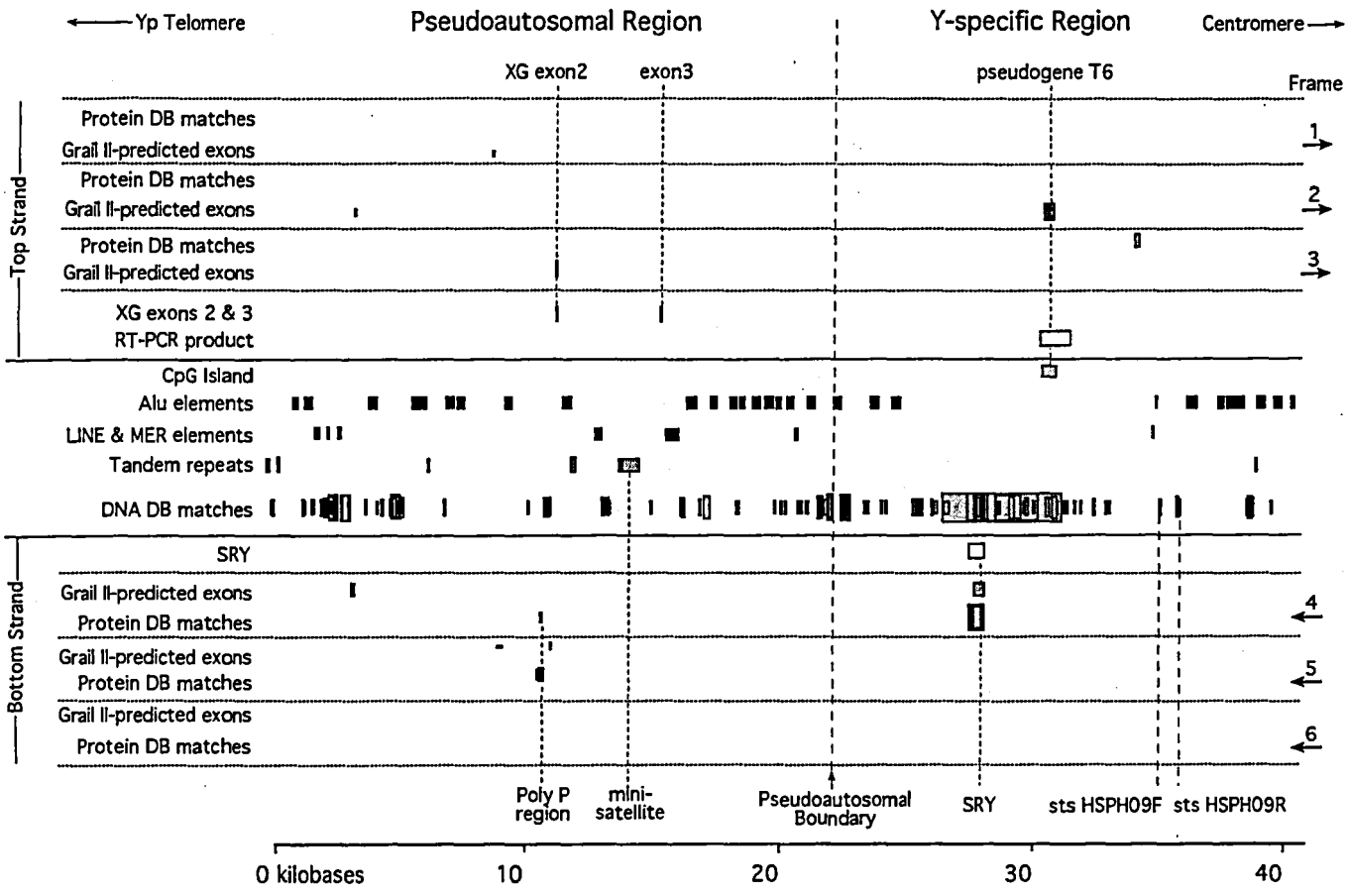


FIG. 2. The principal sequence features of cAMF3.1. The figure is derived from the AceDB display. cAMF3.1 is in the same orientation as in Fig. 1. (Middle) In this panel boxes show the location of CpG islands, *Alu*, LINE, and MER elements, tandem repeats, and "DNA DB matches." DNA DB matches are regions where the sequence of cAMF3.1 matches one or more entries in the EMBL database. The height of the box denotes the percentage similarity between the sequence and the database entry sequence. When one region of sequence matches multiple database entries, the overlaid boxes cause dark shading. (Top and bottom) Above the central panel, boxes mark the locations of GRAIL II-predicted exons and "protein DB matches" for the three reading frames of the top strand of cAMF3.1. The three reading frames of the bottom strand of cAMF3.1 are shown below. Protein DB matches are regions where the sequence of cAMF3.1 translated in that frame matches an entry in the SWIR5 database (see Materials and Methods). The height of the boxes denoting GRAIL II-predicted exons denotes the confidence of the prediction. The height of the boxes marking protein DB matches also denotes the percentage similarity of the sequence to the database sequence. In addition, *XG* exons and the Y-specific region found spliced to these exons in a RT-PCR product (see text) are marked on the top strand. *SRY* is marked on the bottom strand.

rently known to contain, therefore, less than 2% coding sequence.

Sequence Analysis: Repetitive Sequence Elements

The second aim of the sequence analysis was to identify organizational, structural, and functional features useful in the investigation of the evolution of the sex-determining and pseudoautosomal boundary regions. Overall base composition of large regions can reflect aspects of origin and function, whereas length variation in tandem repeats and insertions of transposable elements can provide useful makers for inter- or intra-species studies. For example, the *Alu* present at the pseudoautosomal boundary is found in great apes but is absent in Old World monkeys (Ellis *et al.*, 1990).

Transposable elements. There are several *Alu* subfamilies of different "ages." Alu-J and Alu-S are the major subfamilies, Alu-J being the older. The Alu-S subfamily is divided into subgroups -Sb, -Sc, -Sp, -Sq, and -Sx. Alu-Sb, the youngest subset of Alu-S, contains two even younger groups Sb1 (also known as the HS subfamily) and Sb2 (Jurka and Milosavljevic, 1991; Jurka, 1993). cAMF3.1 contains 30 whole or truncated copies of the *Alu* element; 5 are Alu-J, 1 is Alu-Sp, 12 are Alu-Sx, 4 are Alu-Sc, and 8 are Alu-Sb. In addition there are two regions of cAMF3.1 that appear to have LINE (long interspersed nuclear element) sequence similarity and six areas that match known medium reiteration frequency (MER) repeats. The locations of all transposable elements are shown in Fig. 2.

It is clear that *Alu* (and other) elements are nonrandomly distributed in the human genome. Large sequencing projects have reported a range of *Alu* frequencies, from 0.1 elements/kb for the *KAL-X* region of Xp22.3 (Legouis *et al.*, 1991) to 1.4 elements/kb in the *ERCC1* region of 19q13.3 (Martin-Gallardo *et al.*, 1992). They appear to be concentrated in the G/C-rich,

"reverse" or "Giemsa light" bands of chromosomes (Korenberg and Rykowski, 1988) and may comprise up to 30% of these regions (Bernardi, 1989). The Y-specific portion of cAMF3.1 contains 10 whole or truncated *Alus* in 18.5 kb (i.e., 0.54 *Alus*/kb. This figure does not include the *Alu* that was inserted at the pseudoautosomal boundary (Ellis *et al.*, 1990). and is 39.1% G/C. The pseudoautosomal portion has higher *Alu* and G/C content; it has 19 whole or truncated *Alus* in 22.7 kb (0.84 elements/kb) and is.47.5% G/C. The contrast lends support to the notion that the pseudoautosomal boundary was generated by a chromosomal rearrangement. There is as yet insufficient evidence to say whether the pseudoautosomal boundary also marks the junction between the light and the dark bands observed at the tip of Yp; *Alu* density and G/C content can vary greatly within a chromosome band; the β-globin cluster on 11p15.5 (EMBL entry HSHBB), for example, has a low G/C and *Alu* content (39.5% G/C and 0.1 *Alus*/kb) but is located in a Giemsa light band.

Tandem repeats. Although microsatellites and minisatellites have been sought on the Y chromosome as potentially highly polymorphic markers for Y-specific population studies, those detected seem to be almost exclusively pseudoautosomal. The pseudoautosomal portion of cAMF3.1 is indeed much richer in tandem repeats than the Y-specific region. It contains five regions of tandem repeats totaling 1.2 kb. The largest is a minisatellite sequence located 8 kb from the boundary. It is composed of 45 copies of a 16-bp element, with sequence similarity between elements averaging 81%, and it does appear to be polymorphic (data not shown). A region of related sequence occurs 4 kb further toward the telomere. It is composed of 3 copies of a 16-base element with 88% sequence conservation. Two other tandem arrays close to the telomeric end of cAMF3.1 are composed of elements that have compositions similar to each other. It is not known whether any of these sequences are polymorphic. The Y-specific portion of cAMF3.1 contains a single, short microsatellite repeat sequence, 17 kb from the pseudoautosomal boundary. Unfortunately it did not show any repeat number variation in a sample of 42 males, which included representatives from a variety of ethnic groups (data not shown).

Total repetitive sequence content of cAMF3.1. cAMF3.1 contains two known repeat sequences in addition to those discussed above: first, the sequence of the human sex chromosome repeat pDP316 (Fisher *et al.*, 1990) (EMBL Accession No. HSSEXRPA) matches cAMF3.1, 5 kb from the telomeric end of the clone, with only a single difference, suggesting that this is the genomic origin of pDP316; second, 248 bp of pseudoautosomal sequence located approximately 5 kb from the boundary are 85% similar to part of the Yp pseudoautosomal telomere clone HSTARS7A (Brown *et al.*, 1990). When this sequence was used to probe Southern blots of digested genomic DNA, multiple copies were detected (data not shown).

In total, 12.7 kb or 30.9% of cAMF3.1 is repetitive sequence. *Alu*, LINE, and MER elements account for 10.4 kb of sequence. The pseudoautosomal portion is particularly rich in repetitive sequence: 42.5% of this area but only 16.8% of the Y-specific region is composed of repetitive DNA. It is likely that the totals will increase as the more rare repetitive elements are characterized. Many of the matches already noted between cAMF3.1 and noncoding portions of nucleotide database entries may be caused by this type of sequence. The sequence at the pseudoautosomal boundary (not including the *Alu* element) makes one such match.

CONCLUSION

The sequence of the cosmid cAMF3.1 has been determined with high accuracy. The cosmid straddles the Yp pseudoautosomal boundary, and the sequence is approximately half pseudoautosomal and half Y specific. The Y-specific sequence is from the genetically defined sex-determining region of the Y chromosome but does not appear to contain any coding sequences other than *SRY*. This supports the conclusion that *SRY* alone is necessary and sufficient for the sex-determining function of the Y chromosome. Moreover, there is no evidence that the genomic structure of *SRY* is other than that which has been previously described (Behlke *et al.*, 1993).

The Y-specific and pseudoautosomal portions of cAMF3.1 seem to differ dramatically in both GC content and repeat content; the pseudoautosomal sequence has a much higher GC content and is much richer in repetitive sequences than is the Y-specific region. This may indicate that a chromosomal rearrangement generated the pseudoautosomal boundary at this location and may further reflect differences in the processes of sequence evolution in nonrecombining regions. It is hoped that this project has generated materials with which this and other questions might be addressed.

ACKNOWLEDGMENTS

We thank G. Micklem for patient assistance with computerized sequence analysis and for critical reading of the manuscript, and N. Smaldon and the staff of the Sanger Centre for assistance with sequencing. L.S.W. is supported by an Imperial Cancer Research Fund Bursary; P.N.G. is supported by the Wellcome Trust. This work was supported by the Wellcome Trust (Grant 036522/Z/92 to J.S.).

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215:** 403–410.

Behlke, M. A., Bogan, J. S., Beer-Romero, P., and Page, D. C. (1993). Evidence that the SRY protein is encoded by a single exon on the human Y chromosome. *Genomics* **17:** 736–739.

Bernardi, G. (1989). The isochore organisation of the human genome. *Annu. Rev. Genet.* **23:** 637–671.

Bird, A. (1986). CpG-rich islands and the function of DNA methylation. *Nature* **321:** 209–213.

Brown, W. R. A., MacKinnon, P. J., Vilasanté, A., Spurr, N., Buckle, V. J., and Dobson, M. J. (1990). Structure and polymorphism of human telomere-associated DNA. *Cell* **63:** 119–132.

Dear, S., and Staden, R. (1991). A sequence assembly and editing program for efficient management of large projects. *Nucleic Acids Res.* **19:** 3907–3911.

Dujon, B., Alexandraki, D., Andre, B., Ansorge, W., Baladron, V., Ballesta, J. P. G., Banrevi, A., Bolle, P. A., Bolotinfukuhara, M., and Bossier, P. (1994). Complete DNA-sequence of yeast chromosome-XI. *Nature* **369:** 371–378.

Ellis, N. A., Goodfellow, P. J., Pym, B., Smith, M., Palmer, M., Frischauf, A. M., and Goodfellow, P. N. (1989). The pseudoautosomal boundary in man is defined by an Alu repeat sequence inserted on the Y chromosome. *Nature* **337:** 81–84.

Ellis, N., Yen, P., Neiswanger, K., Shapiro, L. J., and Goodfellow, P. N. (1990). Evolution of the pseudoautosomal boundary in Old World monkeys and Great Apes. *Cell* **63:** 977–986.

Ellis, N. A., Tippett, P., Petty, A., Reid, M., German, J., Weller, P. A., Goodfellow, P. N., Thomas, S., and Banting, G. (1994a). *PBDX* is the *XG* blood group gene. *Nature Genet.* **8:** 285–290.

Ellis, N. A., Ye, T. Z., Patton, S., German, J., Goodfellow, P. N., and Weller, P. (1994b). Cloning of *PBDX*, and *MIC2*-related gene that spans the pseudoautosomal boundary on chromosome Xp. *Nature Genet.* **6:** 394–400.

Fisher, E. M. C., Alitalo, T., Luoh, S. W., La Chapelle, A., and Page, D. C. (1990). Human sex-chromosome-specific repeats within a region of pseudoautosomal/Yq homology. *Genomics* **7:** 625–628.

Freije, D., Helms, C., Watson, M. S., and Doniskeller, H. (1992). Identification of a second pseudoautosomal region near the Xq and Yq telomeres. *Science* **258:** 1784–1787.

Gish, W., and States, D. J. (1993). Identification of protein coding regions by database similarity search. *Nature Genet.* **3:** 266–272.

Gleeson, T. J., and Staden, R. (1991). An X-windows and UNIX implementation of our sequence analysis package. *Comput. Appl. Biosci.* **7:** 398.

Goodfellow, P. J., Pritchard, C., Tippett, P., and Goodfellow, P. N. (1987). Recombination between the X and Y chromosomes: Implications for the relationship between MIC2, XG and YG. *Ann. Hum. Genet.* **51:** 161–167.

Goodfellow, P. N., and Lovell-Badge, R. (1993). *SRY* and sex determination in mammals. *Annu. Rev. Genet.* **27:** 71–92.

Hawkins, J. R. (1994). Sex determination. *Hum. Mol. Genet.* **3:** 1463–1467.

Hawkins, T. L. (1992). M13 single-stranded purification using a biotinylated probe and streptavidin coated magnetic beads. *J. DNA Seq. Map.* **3:** 65–69.

Jakubiczka, S., Arnemann, J., Cooke, H. J., Krawczak, M., and Schmidtke, J. (1989). A search for restriction fragment length polymorphism on the human Y chromosome. *Hum. Genet.* **84:** 86–88.

Jobling, M. A. (1994). A survey of long-range DNA polymorphisms on the human Y-chromosome. *Hum. Mol. Genet.* **3:** 107–114.

Jones, M. C., and Sulston, J. (1995). A new vector and strategy for unidirectional sequential subcloning of DNA. Submitted for publication.

Jurka, J. (1993). A new subfamily of recently retroposed human alu repeats. *Nucleic Acids Res.* **21:** 2252.

Jurka, J., and Milosavljevic, A. (1991). Reconstruction and analysis of human Alu genes. *J. Mol. Evol.* **32:** 105–121.

Koopman, P., Gubbay, J., Vivan, N., Goodfellow, P., and Lovell-Badge, R. (1991). Male development of chromosomally female mice transgenic for *Sry*. *Nature* **351:** 117–121.

Korenberg, J. R., and Rykowski, M. C. (1988). Human genome organization: Alu, Lines, and the molecular structure of metaphase chromosome bands. *Cell* **53:** 391–400.

Krogh, A., Brown, M., Mian, I. S., Sjolander, K., and Haussler, D. (1994). Hidden Markov-models in computational biology—Applications to protein modelling. *J. Mol. Biol.* **235:** 1501–1531.

Legouis, R., Hardelin, J.-P., Levilliers, J., Claverie, J.-M., Compain, S., Wunderle, V., Millasseau, P., Le Paslier, D., Cohen, D., Caterina, D., Bougueleret, L., Delemarre-Van de Waal, H., Lutfalla, G., Weissenbach, J., and Petit, C. (1991). The candidate gene for the X-linked Kallmann syndrome encodes a protein related to adhesion molecules. *Cell* **67:** 423–435.

Malaspina, P., Persichetti, F., Novelletto, A., Iodice, C., Terrenato, L., Wolfe, J., Ferraro, M., and Prantera, G. (1990). The human Y chromosome shows a low level of DNA polymorphism. *Ann. Hum. Genet.* **54:** 297–305.

Mann, J. D., *et al.* (1962). A sex-linked blood group. *Lancet* **i:** 8–10.

Martin-Gallardo, A., McCombie, W. R., Gocayne, J. D., Fitzgerald, M. G., Wallace, S., Lee, B. M. B., Lamerdin, J., Trapp, S., Kelley, J. M., and Liu, L. I. (1992). Automated DNA sequencing and analysis of 106 kilobases from human chromosome 19q13.3. *Nature Genet.* **1:** 34–39.

Milosavljevic, A., and Jurka, J. (1993a). Discovering simple DNA-sequences by the algorithmic significance method. *Comput. Appl. Biosci.* **9:** 407–411.

Milosavljevic, A., and Jurka, J. (1993b). Discovery by minimal length encoding—A case-study in molecular evolution. *Mach. Learn.* **12:** 69–87.

Simmler, M.-C., Rouyer, F., Vergnaud, G., Nystrom-Lahti, M., Ngo, K. Y., de la Chapelle, A., and Weissenbach, J. (1985). Pseudoautosomal DNA sequences in the pairing region of the human sex chromosomes. *Nature* **317:** 692–697.

Sinclair, A. H., Berta, P., Palmer, M. S., Hawkins, J. R., Griffiths, B. L., Smith, M. J., Foster, J. W., Frischauf, A. M., Lovell-Badge, R., and Goodfellow, P. N. (1990). A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* **346:** 240–244.

Sonnhammer, E. L. L., and Durbin, R. (1994). A workbench for large-scale sequence homology analysis. *Comput. Appl. Biosci.* **10:** 301–307.

Spurdle, A. B., and Jenkins, T. (1992). The Y chromosome as a tool for studying human evolution. *Curr. Opin. Genet. Dev.* **2:** 487–491.

Weller, P. A., Goodfellow, P. N., German, J., and Ellis, N. A. (1994). The human Y chromosome homologue of *XG*:Transcription of a naturally truncated gene. *Hum. Mol. Genet.*, in press.

Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Burton, J., Connell, M., Bonfield, J., Copsey, T., and Cooper, J. (1994). 2.2 Mb of contiguous nucleotide-sequence from chromosome-III of *C. Elegans*. *Nature* **368:** 32–38.

Xu, Y., Einstein, J. R., Mural, R. J., Shah, M. B., and Uberbacher, E. C. (1994a). An improved system for exon recognition and gene modeling in human DNA sequences. *In* "Proceedings of the 2nd International Conference on Intelligent Systems for Molecular Biology," AAAI Press, Menlow Park, CA.

Xu, Y., Mural, R. J., Shah, M. B., and Uberbacher, E. C. (1994b). Recognizing exons in genomic sequence using GRAIL II. *in* "Genetic Engineering: Principles and Methods" (J. Setlow, Ed.), Plenum, New York.

GENOME MAPPING & SEQUENCING

May 8 - 12, 1996

ABSTRACT DEADLINE: February 21, 1996

Dear Colleague:

Cold Spring Harbor Laboratory is organizing the **9th Annual Genome Mapping & Sequencing Meeting,** which will begin at 7:30 pm on Wednesday, May 8th, 1996 and run through noon on Sunday, May 12th. We hope to bring together the major practitioners in the field of genome analysis.

The 1996 meeting will feature sessions on physical mapping, biological insights from multi-organismal genomics, human genetics and biology, informatics, mapping methods and technologies, gene discovery and transcript mapping. We will be highlighting large-scale DNA sequencing projects in a poster symposium session. In addition, last year's successful introduction of interactive demonstrations of genomic computational tools and databases will again form part of the program. The keynote address will be given by Dr. Shirley Tilghman before the banquet on Saturday evening.

We encourage you to submit your registration form and abstract materials by the deadline of **February 21, 1996.** In an effort to keep each session to a manageable size, we may need to limit the total number of abstracts presented at the meeting. As a result, we urge groups to avoid overlapping or redundant material and ask that each laboratory submit no more than 3 abstracts.

We anticipate that there may be more applicants than can be accommodated bythe facilities at Cold Spring Harbor. Should it be necessary to limit attendance, we will endeavor to ensure that the best possible science is presented and that every effort is made to provide for representation by all groups wishing to participate. Group leaders may be asked for a priority list to aid the organizers in making appropriate decisions.

We expect the availability of a small grant that can provide some support for attending the meeting. If your attendance depends upon the availability of such funding, please be sure to include that information when submitting your application. Cold Spring Harbor is planning to accept submission of abstracts as well as registration information over the World Wide Web. Complete details on how to register and submit abstracts online are available at http://www.cshl.org/meetings/gmas.html. For your convenience, hard copy abstract instructions and registration materials are enclosed.

We look forward to seeing you in May at Cold Spring Harbor.

Sincerely,

THE ORGANIZERS

David Bentley, Sanger Centre
Eric Green, National Institutes of Health
Phil Hieter, Johns Hopkins University

Cold Spring Harbor Laboratory, The Meetings & Courses Office, P.O. Box 100, Cold Spring Harbor NY 11724-2213
***Email: meetings@cshl.org** * Phone: (516) 367-8346 * Fax: (516) 367-8845 * W3 site: http://www.cshl.org/*

1996 Cold Spring Harbor Laboratory Meetings Application Form

Complete the following information and mail to The Meetings & Courses Office

First Name _____ Last Name _____ CSHL ID# (if known) _____

Affiliation _____

Department _____

Street/Building _____ City _____ State _____

Zip Code/Mail Code _____ Country _____ E-Mail Address _____

Phone Number _____ Fax Number _____

Please Indicate (for housing purposes)
☐ Female
☐ Male

Select One:
☐ Lab Head
☐ Professor
☐ M.D.
☐ Staff Scientist
☐ Post Doc
☐ Grad Student
☐ Other

Please Indicate where you heard about this CSHL meeting (choose one):
☐ Prior Attendee ☐ Word of Mouth ☐ Direct Mail ☐ Magazine Ad ☐ CSHL Poster ☐ Web Site ☐ Electronic Bulletin Board

MEETING TITLE _____ DATES _____

Abstract Information (Please **underline the presenter's name** in the listing of names & affiliations on your abstract. For further instructions, see enclosures.)

 # OF ABSTRACT(S) ENCLOSED THAT REGISTRANT WILL BE PRESENTING

☐ Zero ☐ One ☐ Two ☐ Three

Last name of Lab Head: _____

Accommodations: Most accommodations are double occupancy. Housing is for participants ONLY! Meetings participants will be housed at outside hotels if necessary.
Note: There is an additional charge of $90.00 per night for early arrivals and late departures (includes meals).

Housing Required: ☐ Yes ☐ No

COMPLETE INFORMATION BELOW ONLY IF APPLICABLE

Roommate Request (if desired) _____ Early Arrival Date _____ Late Departure Date _____

- **Special Diet & Health Needs**

☐ Kosher ☐ Vegetarian ☐ Other: _____

For further information contact:
Cold Spring Harbor Laboratory
The Meetings & Courses Office
1 Bungtown Rd. PO Box 100
Cold Spring Harbor, NY 11724-2213

phone: (516) 367-8346
fax: (516) 367-8845
e-mail: meetings@cshl.org
World Wide Web site
http://www.cshl.org/

PAYMENT MUST BE SUBMITTED WITH THIS FORM. Balance due upon arrival. Make checks payable to Cold Spring Harbor Laboratory. All payments must be in US Dollars drawn on a US Bank. Grad Students must provide photocopy of ID card with registration form to receive reduced rate.

- **Registration, Food & Housing cost:**

$200 housing deposit required with reg. form
(CIRCLE APPLICABLE RATE)

	Academic	*Grad Student*	*Corporate*
4 Day	$730	$595	$930
5 Day	$855	$695	$1075
7 Day	$1030	$855	$1280

- **Registration and Food Only cost:**

$100 deposit required with reg. form
(CIRCLE APPLICABLE RATE)

	Academic/Grad Student	*Corporate*
4 Day	$475	$605
5 Day	$555	$695
7 Day	$645	$810

Total Amount Due $ .....

Payment enclosed $ .....

Balance Due $ .....

Credit Card Information:
Amount Authorized $ .....

Account #: _____

_____ X _____
Exp. Date Cardholder Signature

☐ Discover ☐ Mastercard ☐ Visa

Cold Spring Harbor Laboratory

eg. Symposium (Poster)
Meeting Title

ABSTRACT OF PAPER TO BE PRESENTED

DO NOT FOLD

A. N. Other
Presenter's Last Name

PRINT TITLE HERE USING ALL CAPITAL LETTERS
Print Authors and Affiliations here. *(Underline presenting author's name)*

Begin text (Single spaced, using paragraph indent with no space between paragraphs.)

Do not draw box. For illustration purposes only

ABSTRACT INSTRUCTIONS:

- Use plain 8½" x 11" paper.
- Limit area of type to 5" width x 7½" height (12.7 cm x 19.05 cm).
- TITLE (all caps), author's name(s), and institution(s) should be flush left as close as possible to the top line.
- Indicate PRESENTER (person who will be attending and actually presenting material) on upper right hand corner of page and by underlining their full name in the list of authors and affiliations.
- Indicate MEETING and preference for *oral* or *poster* in upper left area
- Your abstract will be published exactly as typed.
- **Send ten copies of the abstract to the Meetings Office by the abstract deadline** in a flat mailing envelope.
- Abstracts must be accompanied by application form. Abstracts that do not follow the above format will be returned.

↑ DO NOT TYPE HERE ↓

↑ DO NOT TYPE HERE ↓

EXAMPLE

↑ 7 ½ inches (19.05 cm) ↓

← 5 inches (12.7 cm) →

← DO NOT TYPE HERE →

Mail*Link SMTP for Tina Whalen

To:
Genome Group

From:
Tom Hudson

Date:
Fri, Feb 16, 1996 10:54 AM

Subject:
CSHL Genome Mapping & Sequencing Meeting

RFC Header:
Received: by gatormail.wi.mit.edu with SMTP;16 Feb 1996 10:54:37 -0500
Received: by genome.wi.mit.edu (5.57/1.1.4/8Aug94)
id AA07927; Fri, 16 Feb 96 10:54:31 -0500
Message-Id: <9602161554.AA07927@genome.wi.mit.edu>
Date: Fri, 16 Feb 1996 10:54:50 -0500
To: Genome_Group@gatormail
From: thudson@genome.wi.mit.edu (Tom Hudson)
Subject: CSHL Genome Mapping & Sequencing Meeting

See below for those submitting abstracts.

Tom

>From: The Meetings & Courses Office <meetings@cshl.org>
>Subject: CSHL Genome Mapping & Sequencing Meeting
>Date: Fri, 16 Feb 1996 11:36:14 -0500
>Content-Type: text/plain; charset="us-ascii"
>Content-Transfer-Encoding: quoted-printable
>Apparently-To: <jmiller@genome.wi.mit.edu>
>Apparently-To: <lstein@genome.wi.mit.edu>
>Apparently-To: <rsteen@genome.wi.mit.edu>
>Apparently-To: <thudson@genome.wi.mit.edu>
>
>
>Our records indicate that you have attended or requested information for =
>the CSHL Genome Mapping & Sequencing Meeting. This year's meeting will =
>take place on May 8-12, 1996.=20
>=20
>The abstract deadline is February 21.
>=20
>For the first time, you can register and submit abstracts for the =
>meeting via the World Wide Web at =
>http://www.cshl.org/meetings/96genome.htm.
>=20
>Keep in mind that, like last year, some abstracts will be programmed for =
>projection-style computer demonstrations in Grace Auditorium. =20
>=20
>If you are NOT submitting your abstract electronically, please indicate =
>one of the following preferences at the lower right hand corner of your =
>abstract (note that these instructions were accidently left off of the =
>registration materials sent to you by mail):
>=20
> Oral or Poster Presentation
> Poster Presentation Only
> Computer Demonstration Only

> Oral or Poster Presentation or Computer Demonstration
>=20
>For additional information concerning this and other CSHL meetings, =
>contact:
>=20
> http://www.cshl.org/meetings
> e-mail: meetings@cshl.org
> fax: 516-367-8845
> phone: 516-367-8346
>
>

Jane

Privileged Communication DON'T CIRCULATE
Submitted to Hum. Molec. Genet.

Integration of physical, breakpoint and genetic maps of chromosome 22. Localization of 575 yeast artificial chromosomes with 235 mapped markers.

Pls. file REPRINT

Callum J. Bell[1]*, Marcia L. Budarf[1], Bart W. Nieuwenhuijsen[2], Barry L. Barnoski[1], Kenneth H. Buetow[4], Keely Campbell[1], Angela Colbert[3], Joelle Collins[1], Philippe R. Desjardins[2], Todd DeZwaan[2], Barbara Eckman[2], Simon Foote[4]**, Kyle Hart[2], Kevin Hiester[2], Marius J. Van Het Hoog[2], Elizabeth Hopper[1], Alan Kaufman[3], Heather E. McDermid[5], G. Christian Overton[2], Mary Pat Reeve[3], David B. Searls[2], Lincoln Stein[3], Edward Watson[1], Rachel Winston[2], Vinay H. Valmiki[2], Robert L. Nussbaum[2]***, Eric S. Lander[3], Kenneth H. Fischbeck[2], Beverly S. Emanuel[1] and Thomas J. Hudson[3]

1. Children's Hospital of Philadelphia, Division of Human Genetics and Molecular Biology, 34th Street & Civic Center Blvd., Philadelphia PA 19104

2. University of Pennsylvania School of Medicine, 415 Curie Blvd., Philadelphia, PA 19104-6146.

3. Center for Genome Research, Whitehead Institute for Biological Sciences/Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge MA 02142

4. Fox Chase Cancer Center, 7701 Burholme Ave, Philadelphia, PA 19111-2412.

5. Department of Biological Sciences, University of Alberta, Edmonton, Alberta T6G 2E9, Canada.

*To whom correspondence should be addressed phone:(215)-590-3856 Fax:(215)-590-3764

** Present address: Walter and Eliza Hall Institute of Medical Research, P.O. Royal Melbourne Hospital, Parkville 3050, Australia.

***Present address: Genetic Disease Research NCHGR/NIH, 9000 Rockville Pike, 49/4A72, Bethesda, MD 20892.

ABSTRACT

Detailed physical maps of the human genome are important resources for the identification and isolation of disease genes and for studying the structure and function of the genome. We used data from STS content mapping of YACs and natural and induced chromosomal breakpoints to anchor contigs of overlapping yeast artificial chromosome (YAC) clones spanning extensive regions of human chromosome 22. The STSs were assigned to specific bins on the chromosome using somatic hybrid mapping panels defining 6 and 25 intervals respectively. YAC libraries were screened by PCR amplification of hierarchical pools of yeast DNA with 235 STSs, and a total of 575 YAC clones were identified. These YACs were assembled into contigs based upon their shared STS content using a simulated annealing algorithm. Fifteen contigs, containing between 2 and 74 STSs were assembled, and ordered along the chromosome based upon the breakpoint, meiotic and PFG maps. Additional singleton YACs were assigned to unique chromosomal bins. These ordered YAC contigs will be useful for identifying disease genes and chromosomal breakpoints by positional cloning and will provide the foundation for higher resolution physical maps for large scale sequencing of the chromosome.

INTRODUCTION

Human chromosome 22 constitutes approximately 1.9% of the haploid autosomal genome (1). Clinical disorders associated with this chromosome include several acquired, tumor-related translocations such as the t(9;22) of chronic myelogenic leukemia and acute lymphocytic leukemia (3,4), the t(8;22) variant translocation of Burkitt's lymphoma (5) and the t(11;22) of Ewing sarcoma (6,7). Deletions of all or part of chromosome 22 are associated with meningiomas (8,9), acoustic neuromas (10,11), Neurofibromatosis type 2 (NF2) (12,13), and rhabdoid tumors (14,15). Chromosome 22 is also involved in the only recurrent constitutional chromosomal translocation in humans (16,17). In addition, a number of syndromes are caused by microdeletions or duplications of portions of 22q11, including DiGeorge syndrome (18-21), velo-cardio-facial syndrome (22), and cat-eye syndrome (23). Chromosome 22 has a high gene density and contains many duplicated sequences and gene families, which makes it an interesting model for mapping studies. The identification of new disease genes will be facilitated by the integration of detailed genetic and physical maps of this chromosome. Moreover, integrated maps can be used to make sequence-ready DNA templates, to facilitate the identification of novel structural elements and the study of chromosome structure.

We used STS-content mapping (24-26) to assemble contigs representing the majority of the chromosome. 313 STSs and 22 hybridization probes were developed from chromosome 22. Markers suspected of containing repeats, and others giving unsatisfactory results in control experiments were eliminated, and the remainder were localized by PCR or Southern hybridization to unique "bins", which are defined by hybrid mapping panels (27,47). These markers were used to identify YACs (28) in three libraries: the CEPH/Genethon YAC libraries (29,30), a chromosome 22-only hybrid cell line derived YAC library, and the Washington University YAC library (31). 213 STSs and 22 hybridization probes identified a total of 575 individual YACs which were then assembled into 15 contigs containing between 2 and 196 YACs.

RESULTS

Marker generation

The STSs and hybridization probes used in this study were derived from genes and other publicly available sequences, STRPs (simple tandem repeat polymorphisms) (32), expressed sequence tags (ESTs) (33,34), YAC vector-insert junction fragments (35), inter-Alu PCR fragments (36) and randomly sequenced plasmid clones (27,47). Table 1 lists the STSs and hybridization probes that were used. Also shown are the individual probes/STSs used at each locus, and the source laboratory of each. GGTX, GGTY and GGTZ refer to probes containing sequences homologous to γ-glutamyl transpeptidase 1 (GGT1) (60). These three GGT-like sequences have been shown to be physically linked to the BCR (break point cluster) -like sequences BCRL2 and BCRL4, and BCR itself, respectively, in 22q11 (unpublished observations). These BCR-like sequences can be distinguished from each other by HindIII polymorphisms (61), allowing assignment of the YACs detected by the GGT1 STS to be allocated to unique bins. Details of this study will be presented in a separate publication. Primer sequences for each STS may be found in the public FTP (file transfer protocol) sites of the Philadelphia (cbil.humgen.upenn.edu /pub/22/) and the Whitehead Institute / MIT (genome.wi.mit.edu /pub/human_STS_releases/) Genome Centers. Further details of the origin and primer sequences of the STSs are also presented in a separate publication (27).

YAC identification

Most YACs in the study were identified in the CEPH/Genethon libraries (original library (29) with an average insert size of 470 kb and megaYAC library with an average insert size of 0.9 megabases (30)) by PCR screening of yeast DNAs pooled in two or three dimensions. Additional YACs were isolated from the Washington University YAC library (31), and from a

chromosome 22 specific YAC library constructed with DNA from hybrid cell line GM10888 (chromosome 22 in a Chinese hamster background). The chromosome 22 specific YAC library contains approximately 300 YACs with an average insert size of 200 kb, equivalent to 1X coverage of the chromosome. YACs isolated from the Washington University library were kindly provided by Collaborators. In addition, limited use was made of a subset of YACs, kindly provided by Ilya Chumakov and Daniel Cohen, identified by hybridization of Alu-PCR products of a chromosome 22-only somatic cell hybrid to the CEPH megaYAC library. YACs from this subset, and from the chromosome 22-specific library were identified by colony hybridization.

Table 2 is a summary of the YAC/STS screening results, listed in order of cytogenetic "bin", and within each bin, ordered alphabetically by GDB locus name. The YACs that were found at each locus are listed. YAC addresses preceded by "A", "B", "C" or "D" are from the Washington University library. YAC addresses preceded by "l" are from the chromosome 22-specific (local) library constructed in the Philadelphia Genome Center. The remaining YACs are from the CEPH libraries. The majority of these results are YACs identified to single microtiter plate addresses, either from unequivocal PCR results in two or three dimensional screens, or from confirmatory PCR tests done on individual YACs. A YAC address consists of three dimensions: plate, row, and column. In initial screening of YAC pools, many of the addresses were incomplete (missing a dimension), or else had more than one possible value in a dimension, which occurs when there is more than one positive YAC per block of 8 microtiter plates (see materials and methods), or from false positive results. Such ambiguous addresses were resolved by several means including fingerprint analysis, comparison with verified YAC addresses of adjacent STSs, or PCR of all possible clones in the degenerate set of addresses. After preliminary contig assembly, most of the clones identified as well as the putative adjacent YACs were individually tested with each STS in the contig.

In order to resolve confusion caused by possible cross-contamination among microtiter plate wells we adopted two approaches. The first approach compared the CEPH/Genethon

fingerprints, where available, of the putative YAC positives with the fingerprints of other YACs known by STS content to overlap the YAC to be resolved. Shared fingerprint bands among these YACs identified with a high degree of confidence the true positive YAC address among several neighboring candidates in several cases. The second approach was based on a calculation of the actual distances between wells of two YAC addresses sharing STSs, divided by the number of STS hits in common; when this measure fell below a certain threshhold for any pair of addresses, they were consolidated into a single address. This heuristic in all cases corresponded well to human judgements about likely cross-contamination, and was shown to be justified in cases that were checked experimentally. Level 1 data from the CEPH/Genethon genome mapping project were confirmed and included in the table.

YAC Contig Assembly

To date, we have used 235 markers to identify 575 YACs. Although the number of YACs we identified indicates nearly 5X coverage of the chromosome, the depth of coverage is uneven: all somatic cell hybrid bins contain YACs, but the 22q11.23 to q12.31 region (bins 12-15; see below) has much deeper coverage than elsewhere. The YACs and STSs fall into 15 islands, defined as sets of STSs and sets of YACs all of which can be reached from each other by following a path of connectivity alternating between STSs and YACs. Singleton YACs detected by one STS each, numbering 25, are omitted from this total. We had difficulty obtaining unequivocal clone and STS order within the largest of these islands, and a clear clone tiling path, even with deep clone coverage of the area and many STSs. In the central portion of the chromosome YAC connectivity has been achieved over a distance exceeding 10 Mb, yet an unbroken clone tiling path remains elusive despite extensive testing of YACs versus STSs in that region. This may be due in part to false positive and negative YAC/STS results (although results have been carefully confirmed), internal deletions within YAC clones, and sequences present at more than one location on the chromosome. Given these problems, the objective becomes to find

an ordering of STSs that minimizes gaps. In ideal data, there should be an order of STSs, corresponding to a true YAC contig, such there are no such gaps. However, in our data all postulated orders of STSs in an island result in some number of "gaps" within YACs in the island, defined as cases where a YAC is negative for some STS but positive for STSs located to both the left and right in the ordering.

For very large islands, finding the STS order with a the absolute minimum number of gaps is computationally intractable, but several approaches have been developed to finding approximate solutions. A simulated annealing (39,40) program we developed employs a random search strategy that seeks local energy minima in the space of all possible orderings, where energy is defined in terms of numbers and sizes of gaps (see Materials and Methods). This approach can be expected to yield somewhat different results for multiple runs, both because there may be more than one valid ordering even for ideal data, and because for "noisy" data the search may find different local energy minima which are near the actual optimum. In practice, the results of multiple runs of simulated annealing are generally similar, although not identical. We refer to these orderings of STSs and YACs as contigs, though it should be emphasized that the larger islands should be viewed as putative contigs at present.

A schematic representation of the coverage of the chromosome in contigs is shown in Figure 1. The chromosome is shown divided into 25 intervals derived from the somatic cell hybrid map of Budarf et al (27). Bin 1 formally includes the short arm of the chromosome but, since very few single copy sequences have been detected there (42), we consider bin 1 effectively to begin on the 22q side of the centromere. The contigs, based on the bin assignment of the STSs that detected the YACs in each, are shown as dark blocks. Since STS content mapping provides only limited information on contig size, the true extent of coverage and the sizes of the gaps separating the contigs are unknown. The stippled block represents a contig of cosmids in a region that proved difficult to clone in YACs. YACs detected by STSs in this part of the chromosome were unstable, and were underrepresented in the libraries screened (M. Budarf, unpublished observations). The cosmid map of this region will be described in a separate

publication. Figure 1 makes clear the low coverage of the distal portion of the chromosome, which arises from the lower density of markers and underrepresentation of the region in the mega-YAC library.

Figure 2 shows simulated annealing results for the largest contig, using a novel method of representing such data to which we have given the name 'Searls plot', after the author of the program. As noted, results of simulated annealing tend toward local minima of the objective function that may differ among runs. The relative merits of these STS orderings and implied YAC contigs cannot be judged with confidence on the basis of the STS data alone. On the other hand, a number of such orderings independently arrived at may be expected to represent a reasonable sampling of the contours of the search space of possible STS orderings. If the predicted orderings do not resemble each other, then little can be said about which is closest to the true optimum, but if they are all similar, one may be more confident in their consensus. Figure 2 shows the degree and nature of the consensus for multiple simulated annealings. The minimum energy ordering among all runs is indicated by the list of STSs running down the left hand side. The gray boxes in the diagram show the positions along the horizontal axis at which the indicated STS occurs in a run, so that the major diagonal denotes complete agreement with the minimum energy run. Other gray boxes indicate other positions at which that STS occurred in other runs, and the shading of a box reflects the number of times a particular STS occurred at the same position in a run. If the predictions for an STS tend to cluster at more than one position in multiple runs, one may infer that the evidence is not strong enough to greatly favor one position over another, though it may be possible to narrow the possibilities to a few regions.

As noted above, even with ideal data it may be possible to have more than one ordering, particularly over subregions of the contig. Obviously, a given ordering of STSs may be reversed in its entirety, without changing the apparent fit to the YAC data in isolation, and for that reason each simulated annealing run is reversed if necessary to more closely approach the consensus. However, there may also be subregions over which the STSs can be reversed without affecting the energy materially, and in this case the Searls plot will display a characteristic 'X' pattern

across the diagonal, representing the alternative orderings. Another characteristic pattern is a displacement of a subregion laterally on the plot, with either a forward or reversed directionality, indicating parts of the contig that display local integrity but which can be moved elsewhere in the larger scheme of things, with little or no penalty. Finally, there are subregions where STSs tend to be in proximity to each other, but where there is little support for ordering them with respect to each other. This may occur, for example, where there are multiple YACs with the same STS hits, but no YACs with only partial overlap to split the STSs and provide order information. These appear as "clouds" of points at or near the diagonal; it can be seen that with a sufficient sample size such regions would approach a uniform distribution of points within a diffuse "superblock". Figure 2 shows a major "X" indicating that the ordering in the distal half of the contig was inverted in a significant number of the simulated annealing runs. We interpret this to mean that the link between D22S591 and D22S47 should be viewed with caution and we have yet to confirm by other means whether actual continuity of YAC coverage exists in this region.

Figure 2 shows that our data suggest, with some confidence, a general ordering of STSs in most sections of this region of the chromosome, but in some areas there is significant scatter. Some of this deviation is systematic in nature, as described in the previous paragraph, and some in all likelihood merely reflects regions where the data is error-prone. An external test of the accuracy of this method is provided not only by bin information but by the meiotic and pulsed-field gel maps (43-46) of the region; the orders of the subsets of markers in both of these maps are similar in the converged order arrived at by simulated annealing, which in this case was done without regard to information from any of these other methods. Figure 3 shows a single simulated annealing solution to the largest contig.

DISCUSSION

We used physical, meiotic, and breakpoint maps of human chromosome 22 to localize contigs of overlapping YAC clones that provide extensive coverage of the long arm of the chromosome. The physical map is developing rapidly due to considerable new data obtained by screening YAC libraries with STSs. The contigs, most of which are anchored by landmarks that have been ordered by meiotic or hybrid mapping, provide extensive coverage of the long arm of the chromosome. Although long range continuity of the clones is not yet complete, the present information is of immediate use to the gene mapping community for identifying disease genes and chromosomal breakpoints. The current state of the physical map reported here reflects the fundamental characteristics of the reagents and methods used, as well as the inherent nature of chromosome 22 itself.

STSs that were developed for chromosome 22 are not randomly distributed along the chromosome. The contig(s) spanning interval 22q11.2-22q13.1 is(are) the most evolved as the result of the high density of markers in this region and greater than average representation of the region in the YAC libraries. The distribution of markers shows a bias towards the center of the long arm of the chromosome (27,47). This is partly because many STRP markers were used as STSs, and these are known to be concentrated in the 22q12 G-dark chromosomal band (52). However, it is not known why other randomly chosen STSs generated from flow-sorted material should also be biased in this way. The distal third of the long arm is correspondingly poor in STSs, and appears to be underrepresented in the YAC libraries, and as a consequence, contains only 2 small contigs and 7 singleton YACs. Interestingly, the distal portion of the long arm appears to be resistant to cloning in both plasmid and YAC libraries, and the consequent paucity of mapping information indicates the need for alternative strategies for covering this region. Currently, we are targetting the region by generating STSs from inter-Alu plasmid libraries made from radiation hybrid cell lines that retain only the distal portion of the chromosome. Success in developing new STSs in this way has shown that YACs, not markers, are likely to be limiting for

STS-YAC contig mapping, and that complete coverage of this region will probably depend on a different cloning vehicle. Current candidate systems are bacterial artificial chromosomes (BACs) (48), P1 phage clones (49), P1 artificial chromosomes (PACs) (50), and cosmids.

Screening multi-dimensional pools of YACs was the only practical way to test all 25,000 Mega-YACs for the presence or absence of a given STS, but created several types of problems. Contamination of adjacent wells during preparation of the pools, absence of amplification in one dimension, or the presence of more than one YAC in the same pool were examples of difficulties that are inherent to pooling schemes which can result in false positive, false negative, and ambiguous YAC addresses. Most of the results obtained from the pool screenings have been resolved by a variety of methods, including analysis of YACs seen with adjacent STSs, fingerprint analysis of selected YACs, and ultimately, the verification of the PCR on the individual YAC. To decrease the errors caused by false negatives on STS order, most STSs were screened on adjacent YACs as well.

The CEPH mega-YACs, which have an average insert size of 0.9 Mb (30), provided the best tool for linking STSs and assembling contigs, and were screened with all available markers. By requiring double linkage before declaring contiguity among STSs in the largest contig, large clones were required, and YACs from the other libraries, while contributing to deep coverage in most regions, did not, in general, contribute to contig assembly. However, in some notable cases contig construction was dependent upon the smaller clones, and as the map matures, they will be useful in resolving the order of closely spaced STSs, and as tools for isolating cosmids or other smaller clones as the map moves towards a higher level of resolution for eventual sequencing.

In addition to the known families of chromosome 22 specific repeats on long arm, such as the BCR, immunoglobulin and GGT gene families, we observed several markers which appear to behave as low copy repeats. In such cases, the PCR assay amplifies two identical or related sequences with products of similar molecular weights. Examples of this were D22S33 and D22S275, which gave several bands of similar size, and detected 15 and 14 YACs respectively. Repetitive STSs created inconsistencies in the data, manifested as large apparent gaps in YAC

clones, since contig assembly software tries to assign them single contig locations. In fact, they may be present at two or more locations. Repeats therefore artificially connect YACs at disparate locations. We arbitrarily decided that STSs detecting 14 or more megaYACs would be declared potentially repetitive and excluded them form contig construction.

The CEPH-Genethon tiling paths (38), provided relatively little additional information because the areas covered by tiling paths coincided with the region where the STS physical map was already well covered. We independently screened the mega-YAC library for the same Genethon genetic markers (55,59), and confirmed the YAC addresses and the level-1 tiling paths present in the November 1993 CEPH-Genethon data release (38). We extracted a few YAC addresses derived by ALU-PCR hybridizations in 22q11.2- q13.1 region that were missed during YAC pool screening. Unfortunately, the areas where the STS content map was poor was also not represented in the tiling paths, or present only in higher level paths that could not be confirmed. Fingerprint analysis on the megaYACs generated by CEPH (51,38), was utilized to resolve ambiguous addresses derived from screening pools of YACs. This method, successful in one third of ambiguous addresses tested, reduced the number of alternate addresses that need to be verified for YAC determination. We did try to assemble the 22q11.2-q13.1 by fingerprint analysis alone using only the fingerprints of YACs that were previously identified to this region. The results had only limited success, yielding small contigs with less than 10 YACs that were already shown to have extensive overlap in STS content.

We chose to represent the data for the large contig in two ways: a single simulated annealing solution , and the Searls plot, derived from multiple runs of simulated annealing. These representations, combined with the YAC-STS results shown in Table 1, provide an objective and useful means of using these data. Previous localization of markers by recombination or breakpoints greatly facilitated the evaluation of the STS content map. The marker order in region 22q11.2-q13.1 spanning more than 11 cM, was broadly consistent with the orders of subsets of markers arrived at by meiotic and pulsed-field gel mapping (43-46). The smaller contigs contain at most two genetically ordered markers, which does not allow real comparisons of marker order

with the meiotic map. In essence, we have made the assumption that the framework linkage map (52) is correct, and used it to anchor and orient the smaller contigs. The best validation of the smaller contigs came from concordance with the somatic cell hybrid binning results.

It is clear that due to biological problems with YACs, the STS-content mapping results from the large contig did not allow us to obtain a fine structure order of the region. This may well be true for many other regions in the genome. The need for additional methodologies to obtain a finer scaffold map of STSs is evident. Radiation hybrids, which allow the study of multiple, larger DNA fragments at a higher redundancy will provide more confidence in generating a high resolution STS order. They will also allow contiguity of the STS map in regions where YAC clones are few or absent.

In conclusion, the physical map of human chromosome 22 has advanced considerably, due to the large scale screening of the CEPH Mega-YAC library with chromosome 22 specific STSs. Current efforts to achieve a complete set of overlapping clones for the long arm of the chromosome are directed at the generation of additional STSs for clone screening, as well as targeted strategies for the distal third of the chromosome using ALU-PCR hybridization strategies.

MATERIALS AND METHODS

Pooling of YAC libraries

For use in the Philadelphia genome center, two dimensional pools of the CEPH/Genethon YAC libraries were constructed as described (53). A Biomek 1000 robotic workstation (Beckman Instruments) was used for yeast DNA isolation and pooling. In brief, yeast clones were grown to saturation in ura- trp- dropout medium in microtiter plates at 30C. 50-75 ul of each clone was pooled into a 1 ml deep-well plate (Beckman Instruments) in which spheroplast preparation and lysis was performed as described elsewhere (54). The lysate was extracted twice

with Strataclean resin (Stratagene) according to manufacturers recommendations. The DNA was then precipitated with isopropanol and the pellet was allowed to dry. After resuspension in TE (10 mM Tris-HCl pH 8.0, 1 mM EDTA, pH 8.0) and treatment with DNAase-free RNAase, the DNA was precipitated with isopropanol and the pellet was dried and resuspended in water. Limited use was also made of commercially purchased DNA pools constructed in three dimensional blocks equivalent to eight microtiter plates each (Research Genetics, Huntsville, Albama).

PCR was performed in 20 ul reactions using approximately 20 ng of pooled yeast DNA in standard PCR buffer (1X buffer (Boehringher-Mannheim) : 10mM Tris-HCl, 1.5mM Mg2+, 50mM KCl, pH 8.3) with 20 nM (final concentration) primers and 0.5U Taq polymerase (Perkin Elmer Cetus or Boehringher Mannheim). PCR conditions were: a five minute denaturation step at 92°· C followed by forty five cycles of 94°C/20 seconds, annealing for 20 seconds, 72°C/80 seconds and a 7 minute extension at 72°C. Suitable annealing temperatures were determined for each STS. The majority of the PCR assays were performed on MJ Research PTC-100 thermal cyclers. Products were analyzed by gel electrophoresis using 1.5 % agarose.

STSs screened at the Whitehead Institute/MIT Center for Genome Research were analyzed using a semi-automated system. The STSs were screened on plates 709 to 972 of the CEPH megaYAC library, generously provided by Daniel Cohen. The YAC library was screened by a two-level pooling scheme. At the first level, there are 32 superpools consisting of DNA from the 768 YACs in a block of eight 96 well plates. Corresponding to each block, there are 8 row, 12 column, and 8 plate subpools. STSs positive at the superpool screen were then screened on the corresponding subpools to identify YAC addresses.

PCRs were prepared by a robotic station built by ROSYS and modified by IAS (Intelligent Automation Systems, Inc., Cambridge, MA). PCR was performed in 20 ul volumes containing 10 ng target DNA, 1X PCR Buffer (10mM Tris/HCl, 50 mM KCl, 1.5 mM Mg++, and 0.001% gelatin), 4 nmol dNTP, 5 pmol each primer, and 0.5 units of Taq. PCRs were completed on custom built thermocyclers (locally called waffle irons, by IAS) each having a

capacity of 16 192 well plates (Costar, Cambridge MA). PCR conditions were: an initial four minute denaturation at 94°C followed by 30 cycles of 50 sec at 94°C, 1.5 minutes at 58°C, 1 minutes at 72°C, and a final extension period of 10 minutes at 72°C.

STSs were screened by either standard agarose gel stained with ethidium bromide or by high throughput chemi-luminescence dot-blot analysis: The PCR products were transferred from the 192 well plates to nylon membranes using a custom built 96 pin pipettor (IAS) and a 6,144 reaction capacity dot-blotting apparatus (96 X 16 X 4X well density, IAS). Subsequent hybridization and detection of the Hybond N+ membrane (Amersham) membranes was done using the ECL kit (Amersham). Hybridization occurred overnight using non-radioactive probes designed from PCR products. STSs known to contain an internal repeat sequence such as CA or AGAT were probed with a molecule containing the repeat structure which had also been labeled with horseradish peroxidase (HRP). All blots were stringently washed with Urea, 2X SSC and SDS at 42° and detected using the standard ECL reagents. Computer images of each autoradiography were obtained using a CCD camera. The VIEW software (Carl Rosenberg, Whitehead Institute) can locate and identify the positive dots, as well as generate an intensity reading.

Fingerprint Resolution Of Degenenerate Addresses.

The STS screening on YAC pools yielded many degenerate YAC addresses, which occur as a result of having more than one positive YAC per block of 8 microtiter plates, from having one dimension in a two or three dimensional screen consistently fail to amplify, and from false positive results. These degenerate addresses represent a small set of addresses (2 to 12) of which usually one or two addresses are contain the specific STS. We used fingerprint data to establish overlaps between the set of ambiguous YACs and the set of definite YACs. We applied a simple band-matching test to the CEPH-genethon fingerprint dataset; we declared pairs of clones with a statistically significant number of matching bands as overlapping. We promoted ambiguous

YACs overlapping one or more definite YACs to "disambiguated" status, denoted in table 2 with a superscript "4". Parameters for declaring overlap are stringent, allowing resolution of only 1/3 of degenerate addresses. However, empirical testing of over 500 fingerprint resolved addresses from random STSs has demonstrated that greater than 95 % can be confirmed by testing the individual YAC DNAs.

Most YAC addresses obtained by screening the YAC pools , fingerprint analysis, and those derived from adjacent STSs during contig building were verified by testing DNA prepared from individual YACs in the library.

Construction of a chromosome 22 specific YAC library. DNA from hybrid cell line GM10888 (chromosome 22 in a Chinese hamster background) was used to create a chromosome 22 specific YAC library essentially as described (56). In brief, high molecular weight DNA from this cell line was partially digested with EcoRI and after ligation to pYAC4 was size selected on a 1% FMC Seaplaque GTG low melting agarose gel in a CHEF-DRII apparatus (BioRad). YACs containing human chromosome 22 DNA were identified by colony hybridization using total human DNA or human CotI DNA as probes.

Contig assembly was performed using a new software package written for use on SPARCstation Unix workstations (Sun Microsystems, Mountain View CA) in a combination of 'C', the logic programming language Prolog (SICStus Prolog, Swedish Institute of Computer Science, PO Box 1263, S-164 28 KISTA, Sweden), and the graphical user interface language Tcl/Tk [57]. The algorithm is based on the technique of simulated annealing, used by a number of others for contig assembly [39,58]; our implementation in particular is similar in broad outline to one developed by CEPH for this purpose [40]. Briefly, in this technique a search space of probe (STS) order permutations, which would be intractable to explore exhaustively, is randomly reordered by selecting from a set of operations such as movement of single probes, swapping of probes, moving of clusters, and inversion of clusters. Any ordering is assigned a notional "energy" that reflects its fit to the STS-YAC data; our energy function involves examining the number and size of apparent gaps required in YACs to account for an ordering of STSs, i.e.

positions where an expected STS hit is not observed, as well as arbitrary other objectives reflecting additional sources of information about probe order. The objective is to minimize this energy by accepting moves that reduce the overall energy. In order to avoid being trapped in a local energy minimum, the process takes place in the context of an abstract "temperature;" a good energy minimum is sought by gradually "cooling" the random search, so that the entire search space is accessible and poor local minima can be escaped, yet there is a gradual convergence (though it cannot be guaranteed that any one solution is optimal). The graphical user interface was designed for maximum interaction with the user, who has the option of reordering probes manually by any of the operations described above, or of asking the program to do so via simulated annealing, for the entire working probe set or any subregion. Islands of connected probe sets can be accumulated in a controlled fashion and with varying stringency as to degree of connectedness. These sets may then be winnowed based on a variety of heuristics to eliminate non-informative or doubtful probes, clones, or points. For example, adjacent or nearby wells with similar reactivities, likely to be due to cross-contamination, may be automatically combined, or YACs that appear to span noncontinuous bins may be removed, etc.

Acknowledgments

The work undertaken in the human genome center for chromosome 22 was supported by grant numbers P50-HG00425 (NCHGR) and CA39926 (NCI) from the NIH. Studies in the Whitehead Institute/MIT Center for genome research were supported by National Institute of Health Center for Genome Research Grant P50-HG00098. We wish to thank Eric Green and Glen Evans for screening for Washington University YACs, Eckart Meese and Marco Giovannini for providing STSs prior to publication, Daniel Cohen, Ilya Chumakov and Jean Weissenbach for the CEPH YAC libraries and the Alu-PCR generated chromosome 22 subset, and Willem Van Loon for biomek routines. Thomas Hudson is a recipient of a Clinician-Scientist Award from the Medical Research Council of Canada.

REFERENCES

1. Morton,N.E. (1991) Parameters of the human genome. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 7474-7476.

2. Kaplan,J.C., Aurias,A., Julier,C., Prieur,M. and Szajnert,M.F. (1987) Human chromosome 22. *J. Med. Genet.* **24**, 65-78.

3. Nowell,P.C. and Hungerford,D.A. (1960) A minute chromosome in human chronic granulocytic leukemia. *Science* **132**, 1497-1499.

4. Rowley,J.D. (1973) A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature* **243**, 290-293.

5. Berger,R., Bernheim,A., Weh,H.J., Flandrin,G., Daniel,M.T., Brouet,J.C. and Colbert,N. (1979) A new translocation in Burkitt's tumor cells. *Hum. Genet.* **53**, 111-112.

6. Aurias,A., Rimbaut,C., Buffe,D., Dubousset,J. and Mazabraud,A. (1983) Chromosomal translocations in Ewing's sarcoma. *N. Engl. J. Med.* **309**, 496-497.

7. Turc-Carel,C., Philip,I., Berger,M.P., Philip,T. and Lenoir,G.M. (1983) Chromosomal translocations in Ewing's sarcoma. *N. Engl. J. Med.* **309**, 497-498.

8. Zang,K.D. (1982) Cytological and cytogenetical studies on human meningioma. *Cancer Genet. Cytogenet.* **6**, 249-274.

9. Dumanski,J.P., Carlbom,E., Collins,V.P. and Nordenskjold,M. (1987) Deletion mapping of a locus on human chromosome 22 involved in the oncogenesis of meningioma. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 9275-9279.

10. Seizinger,B.R., Martuza,R.L. and Gusella,J.F. (1986) Loss of genes on chromosome 22 in tumorigenesis of human acoustic neuroma. *Nature* **322**, 644-647.

11. Seizinger,B.R., Rouleau,G., Ozelius,L.J., Lane,A.H., ST. George- Hyslop,P., Huson,S., Gusella,J.F. and Martuza,R.L. (1987) Common pathogenetic mechanism for three tumor types in bilateral acoustic neurofibromatosis. *Science* **236**, 317-319.

12. Trofatter,J.A., MacCollin,M.M., Rutter,J.L., Murell,J.R., Duyao,M.P., Parry,D.M., Eldridge,R., Kley,N., Menon,A.G., Pulaski,K., Haase,V.H., Ambrose,C.M., Munroe,D., Bove,C., Haines,J.L., Martuza,R.L., MacDonald,M.E., Seizinger,B.R., Short,M.P., Buckler,A.J. and Gusella,J.F. (1993) A novel moesin-, ezrin-, radixin-like gene is a candidate for the neurofibromatosis 2 tumor suppressor. *Cell* **72**, 791-800.

13. Rouleau,G.A., Merel,P., Lutchman,M., Sanson,M., Zucman,J., Marineau,C., Hoang-Xuan,K., Demczuk,S., Desmaze,C., Plougastel,B., Pulst,S.M., Lenoir,G., Bijlsma,E., Fashold,R., Dumanski,J., de Jong,P., Parry,D., Eldridge,R., Aurias,A., Delattre,O. and Thomas,G. (1993) Alteration in a new gene encoding a putative membrane-organizing protein causes neuro-fibromatosis type 2. *Nature* **363**, 515-521.

14. Biegel,J.A., Rorke,L.B., Packer,R.J. and Emanuel,B.S. (1990) Monosomy 22 in rhabdoid or atypical tumors of the brain. *J Neurosurg.* **73**, 710-714.

15. Biegel,J.A., Burk,C.D., Parmiter,A.H. and Emanuel,B.S. (1992) Molecular analysis of a partial deletion of 22q in a central nervous system rhabdoid tumor. *Genes Chromosom.Cancer* **5**, 104-108.

16. Zackai,E.H. and Emanuel,B.S. (1980) Site-specific reciprocal translocation, t(11;22) (q23;q11), in several unrelated families with 3:1 meiotic disjunction. *Am. J. Med. Genet.* **7**, 507-521.

17. Fraccaro,M., Lindsten,J., Ford,C.E. and Iselius,L. (1980) The 11q;22q translocation: a European collaborative analysis of 43 cases. *Human Genet.* **56**, 21-51.

18. De La Chapelle,A., Herva,R., Koivisto,M. and Aula,P. (1981) A deletion in chromosome 22 can cause DiGeorge syndrome. *Hum. Genet.* **57**, 253-256.

19. Kelley,R.I., Zackai,E.H., Emanuel,B.S., Kistenmacher,M., Greenberg,F. and Punnett,H.H. (1982) The association of the DiGeorge anomalad with partial monosomy of chromosome 22. J. *Pediatr.* **101**, 197-200.

20. Carey,A.H., Roach,S., Williamson,R., Dumanski,J.P., Nordenskjold,M., Collins,V.P., Rouleau,G., Blin,N., Jalbert,P. and Scambler,P. (1990) Localization of 27 DNA markers to

the region of human chromosome 22q11-pter deleted in patients with the DiGeorge syndrome and duplicated in the der22 syndrome. *Genomics* **7**, 299-306.

21. Fibison,W.J.,Budarf,M., McDermid,H., Greenberg,F. and Emanuel,B.S. (1990) Molecular studies of DiGeorge syndrome. *Am. J. Hum. Genet.* **46**, 888-895.
22. Emanuel,B.S., Driscoll,D., Goldmuntz,E., Baldwin,S., Beigel,J., Zackai,E.H., McDonald-McGinn,D., Sellinger,B., Gorman,N., Williams,S and Budarf,M. (1993) The Phenotypic Mapping of Down Syndrome ansd Other Aneuploid Conditions, Wiley-Liss: 207-224.
23. McDermid,H.E., Duncan,A.M.V., Brasch,K.R., Holden,J.J.A., Magenis,E., Sheehy,R., Burn,J., Kardon,N., Noel,B., Schinzel,A., Teshima,I. and White,B.N. (1986) Characterization of the supernumery chromosome in cat eye syndrome. *Science* **232**, 646-648.
24. Olson,M., Hood,L., Cantor,C. and Botstein,D. (1989) A common language for physical mapping of the human genome. *Science* **245**, 1434-1435.
25. Green,E.D. and Olson,M. (1990) Chromosomal region of the cystic fibrosis gene in yeast artificial chromosomes: a model for human genome mapping. *Science* **250**, 94-98.
26. Green,E.D. and Green,P. (1991) Sequence-tagged site (STS) content mapping of human chromosomes: theoretical considerations and early experiences. *PCR Methods Applic.* **1**, 77-90.
27. Budarf, M.L., Eckman, B., Michaud, D., Buetow, K.H., Williams, S., McDermid, H., Goldmuntz, E., Gavigan, S., Meese, E., Biegel, J., Dumanski, J., Bell, C.J. and Emanuel, B.S. (1994) Regional localization of over 300 loci on human chromosome 22 with an extended regional mapping panel. *Genomics*, submitted.
28. Burke,D.T., Carle,G.F. and Olson,M.V. (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* **236**, 806-812.
29. Albertsen,H.M., Abderrahim,H., Cann,H.M., Dausset,J., Le Paslier,D. and Cohen,D. (1990) Construction and characterization of a yeast artificial chromosome library containing seven haploid human genome equivalents. *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4256-4260.

30. Chumakov,I., Rigault,P., Guillou,S., Ougen,P., Billaut,A., Guasconi,G., Gervy,P., LeGall,I., Soularue,P., Grinas,L. et al. (1992) Continuum of overlapping clones spanning the entire human chromosome 21q. *Nature* **359**, 380-387.

31. Brownstein,B.H., Silverman,G.A., Little,R.D., Burke,D.T., Korsmeyer,S.J., Schlessinger,D. and Olson,M.V. (1989) Isolation of single-copy human genes from a library of yeast artificial chromosome clones. *Science* **244**, 1348-1351.

32. Weber,J.L. and May,P.E. (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum.Genet.* **44**, 388-396.

33. Wilcox, A.S., Khan,A.S., Hopkins,J.A. and Sikela,J.M. (1991) Use of 3' untranslated sequences of human cDNAs for rapid chromosome assignment and conversion to STSs: implications for an expression map of the genome. *Nucleic Acids Res.* **19**, 1837-1843.

34. Adams,M.D., Kelley,J.M., Gocayne,J.D., Dubnick,M., Polymeropoulos,M.H., Xiao,H., Merril,C.R., Wu,A., Olde,B. and Moreno,R.F. (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* **252**, 1651-1656.

35. Riley,J., Butler,R., Ogilvie,D.J., Finniear,R., Jenner,D., Anand,R., Smith,J.C. and Markham,A.F. (1990) A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. Nucleic Acids Res. **18**, 2887-2890.

36. Nelson,D.L., Ledbetter,S.A., Corbo,L., Victoria,M.F., Ramirez-Solis,R., Webster,T.D., Ledbetter,D.H. and Caskey,C.T. (1989) Alu polymerase chain reaction: a method for rapid isolation of human-specific sequences from complex DNA sources. *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6686-6690.

37. Lehrach,H. et al. (1990) In Davies,K.E. and Tilghman,S.M. (eds.), Genome Analysis Volume 1: Genetic and Physical Mapping. Cold Spring Harbor Laboratory Press, Cold Spring Harbor: 39-81.

38. Cohen,D., Chumakov,I. and Weissenbach,J. (1993) A first-generation physical map of the human genome. *Nature* **366**, 698-701.

39. Cuticchia,A.J., Arnold,J. and Timberlake,W.E. (1992) The use of simulated annealing in chromosome reconstruction experiments based on binary scoring. *Genetics* **132**, 591-601.

40. Rigault, P. (1993) In Lim,H.A., Fickett,J., Cantor,C.R. and Robbins,R.J. (eds.) Clone ordering by simulated annealing: Application to the STS-content map of chromosome 21. Proceedings of the Second International Conference on Bioinformatics, Supercomputing, and Complex Genome Analysis. World Scientific Publishing: 169-183.

41. Arratia,R., Lander,E., Tavare,S. and Waterman,M. (1992) Genomic mapping by anchoring random clones: a mathematical analysis. *Genomics* **11**, 806-827

42. Schinzel,A.A., Basaran,S., Bernasconi,F., Karaman,B., Yuksel-Apak,M. and Robinson,W.P. (1994) Maternal uniparental disomy 22 has no impact on the phenotype. Am. J. Hum. Genet. **54**, 21-24.

43. McDermid,H.E., Budarf,M.L. and Emanuel,B.S. (1993) Long-range restriction map of human chromosome 22q11-22q12 between the lambda immunoglobulin locus and the Ewing sarcoma breakpoint. *Genomics* **18**, 308-318.

44. Bucan,M., Gatalica,B., Nolan,P., Chung,A., Leroux,A., Grossman,M.H., Nadeau,J.H., Emanuel,B.S. and Budarf,M. (1993) Comparative mapping of 9 human chromosome 22q loci in the laboratory mouse. *Hum. Mol. Genet.* **2**, 1245-1252.

45. Dumanski,J.P., Carlbom,E., Collins,V.P., Nordenskjold,M., Emanuel,B.S., Budarf,M.L., McDermid,H.E., Wolff,R., O'Connell,P. and White,R. (1991) A map of 22 loci on human chromosome 22. *Genomics* **11**, 709-719.

46. Delattre,O., Azambuja,C.J., Aurias,A., Zucman,J., Peter,M., Zhang,F., Hors-Cayla,M.C., Rouleau,G. and Thomas,G. (1991) Mapping of human chromosome 22 with a panel of somatic cell hybrids. *Genomics* **9**, 721-727.

47. Hudson, T.J., Colbert, A.M.E., Reeve, M.P., Bae, J.S., Lee, M.K., Nussbaum, R.L., Budarf, M.L., Emanuel, B.S. and Foote, S. (1994) Isolation and regional mapping of 110 chromosome 22 STSs.*Genomics*, in press.

48. Shizuya,H., Birren,B., Kim,U.J., Mancino,V., Slepak,T., Tachiiri,Y. and Simon,M. (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 8794-8797.

49. Sternberg,N. (1990) Bacteriophage P1 cloning system for the isolation, amplification, and recovery of DNA fragments as large as 100 kilobase pairs. *Proc. Natl. Acad. Sci. U.S.A.* **87**, 103-107.

50. Ioannou,P.A., Amemiya,C.T., Garnes,J., Kroisel,P.M., Shizuya,H., Chen,C., Batzer,M.A. and de Jong,P.J. (1994) A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nat. Genet.* **6**, 84-89.

51. Barillot,E., Lacroix,B. and Cohen,D. (1991) Theoretical analysis of library screening using a N-dimensional pooling strategy. *Nucleic Acids Res.* **19**, 6241-6247.

52. Buetow,K.H., Duggan,D., Yang,B., Ludwigsen,S., Puck,J., Porter,J., Budarf,M., Spielman,R. and Emanuel,B.S. (1993) A microsatellite-based multipoint index map of human chromosome 22. *Genomics* **18**, 329-339.

53. Amemiya,C.T., Alegria-Hartman,M.J., Aslanidis,C., Chen,C., Nikolic,J ., Gingrich,J.C. and de Jong,P.J. (1992) A two-dimensional YAC pooling strategy for library screening via STS and Alu-PCR methods. *Nucleic Acids Res.* **20**, 2559-2563.

54. Green,E.D and Olson,M.V. (1990) Systematic screening of yeast artificial-chromosome libraries by use of the polymerase chain reaction. *Proc. Natl. Acad. Sci. U.S.A.* **87**, 1213-1217.

55. Weissenbach, J., Gyapay, G., Dib. C., Vignal. A., Morissette. J., Millasseau. P., Vaysseix. G. and Lathrop M. (1992) A second-generation linkage map of the human genome. *Nature* **359**, 777-8

56. Lee,J.T., Murgia,A., Sosnoski,D.M., Olivos,I.M. and Nussbaum,R.L. (1992) Construction and characterization of a yeast artificial chromosome library for Xpter-Xq27.3: a systematic determination of cocloning rate and X-chromosome representation. *Genomics* **12**, 526-533.

57. Ousterhout,J.K. (1994) "Tcl and the Tk Toolkit." Addison-Wesley, Reading, MA.

58. Mott,R., Grigoriev,A., Maier,E., Hoheisel,J. and Lehrach,H. (1993) Algorithms and software tools for ordering clone libraries: application to the mapping of the genome of Schizosaccharomyces pombe. *Nucleic Acids Res.* **21**, 1965-1974.

59. Gyapay, G., Morrisette, J., Vignal, A., Dib, C., Fizames, C., Millaseau, P., Marc, S., Bernardi, G., Lathrop, M. and Weissenbach, J. (1994) The 1993-94 genethon human genetic linkage map. *Nature Genet.* **7**, 246-339

60. Figlewicz, D.A., Delattre, O., Guellaen, G., Krizus, A., Thomas, G., Zucman,J., and Rouleau, G.A. (1993) Mapping of human γ-glutamyl transpeptidase genes on chromosome 22 and other autosomes. *Genomics* **17**, 299-305

61. Budarf, M.L., Canaani, E. and Emanuel, B.S. (1988) Linear order of the four BCR-related loci in 22q11. *Genomics* **3**, 168-172

Footnote to Table 2

1, verified by PCR on purified DNA or a single yeast colony

2, clear positive by PCR on DNAs pooled in two or three dimensions

3, found positive by colony or Southern hybridization

4, disambiguated by restriction fingerprint data.

e, YACs kindly provided by Dr. Glen Evans

r, YACs taken from ref. 13

Figure legends

Figure 1. Estimated coverage of the chromosome in contigs. The horizontal lines are the boundaries separating 22 bins. Contigs are shown as blocks. The stippled block shows the location of a cosmid contig encompassing the DiGeorge critical region (DGCR).

Figure 2. Searls Plot of simulated annealing data for the largest contig accumulated from multiple runs of the program. The list of loci down the left of the figure is the "minimum energy" ordering of markers (see the text for detailed explanation). Gray boxes indicate the position on the horizontal axis at which the indicated STSs occurred during individual runs. Darker boxes indicate that an STS was positioned in the same location in multiple runs. Boxes falling repeatedly on the diagonal indicate high confidence in the minimum energy ordering. Horizontal dotted lines indicate the chromosomal bin location of each STS. The bin intervals are shown at the top of the figure. Circles indicate the consensus positions of markers that are present on the meiotic map.

Figure 3. A single solution for the largest contig in the central region of chromosome 22q. The contig was contructed as follows: YACs and STSs were selected by connectivity to D22S1, obeying the double linkage rule. Singletons (YACs detected by one STS only) were then eliminated, as were markers that detected more the 14 YACs. Singletons were eliminated a second time, and the resulting set of markers and YAcs were subjected to simulated annealing. Marker order is shown along the top of the figure. Above each marker name is the bin interval that the marker was mapped to, e.g. 15/16 indicates the marker is in bin 15-16. YACs are shown as heavy horizontal black lines.

Table titles:

Table1. Loci used for YAC identification

Table 2. YACs localized on chromosome 22

Locus	Probe	Source	Type
none	22-5	Bell	sts
ACR	ACR	Trofatter-Maccollin	sts
ADORA1	ADORA1	Budarf	sts
ARSA	CP8	Gieselmann	clone
	ARSA	Gieselmann	sts
ATP6E	ATPaseP31	Bell	sts
	ATPaseP31-2	Bell	sts
BCR	5'BCR	Canaani	clone
	BCR	Dunham	sts
	GB21	Hudson	sts
BCRL2	BCRL2	Budarf	clone
BCRL3	BCRL3	Budarf	clone
BCRL4	BCRL4	Budarf	clone
BZRP	pPBS11	Strauss	clone
	BZRP-2	Bell	sts
none	CB10	Bell	sts
none	COS7-1	Bell	sts
	COS7-1-2	Bell	sts
CRKL	CRKL3pr	Bell	sts
CRYBB2	CRYB2	Haines	sts
	CRYB2A	Dunham	sts
CYP2D8P	CYP2D8P	Buetow	sts
D22S1	D22S1	Denny	sts
D22S9	D22S9	Gusella	sts
D22S15	D22S15	Denny	sts
D22S23	D22S23	Gusella	sts
D22S24	D22S24	Unknown	sts
D22S28	W23C	Bell	sts
D22S29	D22S29	Gusella	sts
D22S33	PHI4	Hudson	sts
D22S37	pH13	Budarf	clone
	pH13	Budarf	sts
D22S38	pH15	Budarf	clone
	PHI15	Hudson	sts
D22S40	PH61	Budarf	sts
D22S42	D22S42*1	Gusella	sts
D22S43	PHI32	Hudson	sts
D22S44	PHI35	Hudson	sts
D22S45	pH41a	Rappaport	sts
D22S47	pH59	Budarf	clone
	pH59	Budarf	sts
D22S50	pH74-2	Budarf	sts
D22S51	pE6	Rappaport	sts
D22S55	pH91	Budarf	sts
D22S56	pH97b	Budarf	clone
	D22S56	Gusella	sts
D22S57	PHI86	Hudson	sts
D22S58	PHI102	Hudson	sts
D22S60	pH109a	Budarf	sts
D22S61	pH109b	Budarf	sts
D22S63	pH120a	Budarf	sts
D22S64	pH130	Budarf	sts
D22S72	LN15	Budarf	sts
	LN15-2	Budarf	sts
D22S91	KI-211	Bell	sts
D22S102	D22S102*1	Gusella	sts
D22S111	KI-197	Bell	sts

Locus	Probe	Source	Type
D22S117	KI-153/4.9KB	Bell	sts
D22S119	KI-189	Bell	sts
D22S137	KI-222	Bell	sts
D22S156	Mfd33	Weber	sts
D22S184	NB85	Budarf	sts
D22S186	NB14	Budarf	sts
D22S190	NB62	Budarf	sts
D22S192	NB97	Budarf	sts
D22S193	D22S193	Dunham	sts
	NB129	Budarf	sts
D22S258	Mfd162	Weber	sts
D22S264	COS39	Rouleau	sts
D22S268	D22S268	Rouleau	sts
D22S270	Mfd204	Weber	sts
D22S272	AFM024xc9	Weissenbach	sts
D22S273	AFM106xd2	Weissenbach	sts
D22S274	AFM164th8	Weissenbach	sts
D22S277	AFM168xa1	Weissenbach	sts
D22S278	AFM182xd12	Weissenbach	sts
D22S279	AFM205yc11	Weissenbach	sts
D22S280	AFM225xf6	Weissenbach	sts
D22S281	AFM238wc11	Weissenbach	sts
D22S282	AFM261ye5	Weissenbach	sts
D22S283	AFM262vh5	Weissenbach	sts
D22S292E	D22S292E	Polymeropoulos	sts
D22S294	4.11	Puck	sts
D22S295	9.11	Puck	sts
	9.11-2	Puck	sts
D22S297	27.7	Puck	sts
D22S299	35.12	Puck	sts
D22S300	42.13	Puck	sts
D22S301	45.4	Puck	sts
D22S302	51.3	Puck	sts
D22S303	68.12	Puck	sts
D22S304	76.1	Puck	sts
D22S306	80.1	Puck	sts
D22S307	82.1	Puck	sts
D22S308	99.1	Puck	sts
D22S310	114.1	Puck	sts
D22S315	AFM183xe9	Weissenbach	sts
D22S351	22TG1	Mcphail	sts
D22S412E	D22S412E	Maglott	sts
D22S417	D22S417	Gerken	sts
D22S418	AFM031yb10	Weissenbach	sts
D22S419	AFM211yf10	Weissenbach	sts
D22S420	AFM217xf4	Weissenbach	sts
D22S425	AFM265yf5	Weissenbach	sts
D22S427	AFM288we5	Weissenbach	sts
D22S431	UT582	White	sts
D22S447	59.1	Puck	sts
D22S540	PB14	Hudson	sts
D22S541	PH31	Hudson	sts
D22S543	PH863	Hudson	sts
D22S544	PB257	Hudson	sts
D22S546	PB1185	Hudson	sts
D22S552	PH1362	Hudson	sts
D22S553	PC39	Hudson	sts
D22S554	PH1364	Hudson	sts

Locus	Probe	Source	Type
D22S556	PH1367	Hudson	sts
D22S557	PH1379	Hudson	sts
D22S559	PC106	Hudson	sts
D22S560	PB216	Hudson	sts
D22S561	PB282	Hudson	sts
D22S562	PB287	Hudson	sts
D22S563	PH476	Hudson	sts
D22S564	PB379	Hudson	sts
D22S565	PB422	Hudson	sts
D22S566	PB426	Hudson	sts
D22S567	PB22	Hudson	sts
D22S568	PB606	Hudson	sts
D22S569	PB728	Hudson	sts
D22S570	PH570	Hudson	sts
D22S571	PB1024	Hudson	sts
D22S572	PB1033	Hudson	sts
D22S574	PB1045	Hudson	sts
D22S576	PB1068	Hudson	sts
D22S577	PB1069	Hudson	sts
D22S579	PB1073	Hudson	sts
D22S582	PB1130	Hudson	sts
D22S584	PB1144	Hudson	sts
D22S588	PB987	Hudson	sts
D22S589	PH138	Hudson	sts
D22S591	PB876	Hudson	sts
D22S594	PB918	Hudson	sts
D22S595	PB931	Hudson	sts
D22S596	PB552	Hudson	sts
D22S604	PH518	Hudson	sts
D22S607	PH621	Hudson	sts
D22S609	PH672	Hudson	sts
D22S611	PH176	Hudson	sts
D22S615	PH808	Hudson	sts
D22S617	PH917	Hudson	sts
D22S618	PH927	Hudson	sts
D22S620	PH937	Hudson	sts
D22S623	PH951	Hudson	sts
D22S624	PH955	Hudson	sts
D22S626	PH1006	Hudson	sts
D22S627	PH1027	Hudson	sts
D22S629	PH1051	Hudson	sts
D22S630	PH604	Hudson	sts
D22S631	PH608	Hudson	sts
D22S633	PC295	Hudson	sts
D22S635	PH831	Hudson	sts
D22S638	PH964	Hudson	sts
D22S639	PH965	Hudson	sts
D22S642	PH843	Hudson	sts
D22S644	PC273	Hudson	sts
D22S650	PH1247	Hudson	sts
D22S652	PH1104	Hudson	sts
D22S653	PH1131	Hudson	sts
D22S655	IGJ2	Dunham	sts
D22S656	PB266	Hudson	sts
D22S659	PB739	Hudson	sts
D22S663	PH710	Hudson	sts
D22S666	PH1036	Hudson	sts
D22S669	GB31	Hudson	sts
D22S715	WI-1905	Hudson	sts
D22S718	WI-2547	Hudson	sts

Locus	Probe	Source	Type
D22S739	25.4	Hudson	sts
D22S745	GBX56	Hudson	sts
D22S746	GBX61	Hudson	sts
D22S776	LN86	Budarf	clone
D22S778	LN89-2	Bell	sts
D22S781	LN98	Budarf	sts
D22S789	LN44	Budarf	sts
D22S792	LN50	Budarf	clone
D22S793	LN53	Budarf	clone
D22S794	LN55	Budarf	clone
D22S795	LN63	Budarf	sts
D22S799	LN77	Budarf	clone
DIA1	5'DIA	Leroux	clone
	DIA1	Dunham	sts
none	EN38	Budarf	sts
none	EWS3'	Giovannini	sts
none	EWSex5	Giovannini	sts
EWSR1	EWSR1	Delattre	sts
F8VWFP	F8VWFP	Buetow	sts
FIBB	FIBB	Bell	sts
G22P1	G22P1	Dunham	sts
	G22P1	Hudson	sts
GGT	GGTX	Budarf	clone
	GGTY	Budarf	clone
	GGTZ	Budarf	clone
GGT1	GGT1/2	Dunham	sts
	GGT1	Hudson	sts
GNAZ	GNAZ	Budarf	sts
HCF2	HCF2	Dunham	sts
	HCF2	Hudson	sts
IGKVP3	IGKVP3.2	Budarf	sts
IGL@	IGLC7		sts
	IGLLbb1	Bauer	sts
IGLC2	IGLC2	Dunham	sts
	IGLC2	Naylor	sts
IL2RB	IL2RB	Buetow	sts
none	KI-1547	Bell	sts
	KI-1547-2	Bell	sts
LIF	p3.1-5'	Lowe	clone
	LIF	Denny	sts
none	MEST14	Meese	sts
none	MEST39	Meese	sts
MMP11	STROM-F2	Emanuel	sts
NAGA	GB26	Hudson	sts
	NAGA	Dunham	sts
NEFH	HW10	Lees	clone
	NEFH	Gusella	sts
OSM	OSM1	Bruce	sts
PDGFB	PDGFB	Dunham	sts
PVALB	PVALB	Berchtold	sts
	GB27	Hudson	sts
SGLT1	SGLT1	Hudson	sts
	SGLT1-2	Bell	sts
TCN2	TCN2	Quadros	clone
	TCN2	Dunham	sts
TIMP3	TIMP3	Budarf	sts
TOP1P2	TOP1P2	Haines	sts
YESP	GB32	Hudson	sts
	YESP	Dunham	sts
none	Z7	Taub	sts

Bin	Locus	Positive YACs
n.d.	COS7-1	361_D_9[1], 744_B_11[1], 873_C_4[1], 911_B_11[1], 957_B_1[1]
n.d.	D22S272	211_D_4[2], 336_H_8[2], 522_A_11[2], l:602_C_3[1], l:603_C_4[1], 849_E_1[1], 93_G_5[2]
n.d.	D22S427	884_E_1[1], 908_H_9[1]
n.d.	D22S543	723_B_6[1], 734_B_10[1], 745_G_7[1], 765_E_2[1], 771_H_4[1], 776_H_2[1], 792_H_10[1], 800_A_4[4], 829_D_11[1], 829_E_11[1], 878_D_3[1], 921_F_9[1]
n.d.	D22S553	966_A_8[1]
n.d.	D22S559	752_C_9[1], 781_D_5[1], 789_D_9[1], 791_B_12[1], 798_D_4[1], 804_F_2[1], 905_F_4[1], 921_H_5[1], 922_H_5[1], 942_C_6[1], 942_C_7[1], 950_D_11[1], 964_B_7[1], 967_C_12[1]
n.d.	D22S562	741_B_3[1], 744_F_6[1], 750_B_4[1], 763_A_3[1], 793_E_9[1], 856_C_1[1]
n.d.	D22S618	763_A_3[4], 856_C_1[4]
n.d.	D22S627	803_G_9[1]
n.d.	D22S656	746_B_2[1], 814_A_11[1], 917_G_12[1], 959_A_7[1]
n.d.	D22S659	798_A_1[4], 939_G_9[4]
n.d.	D22S663	719_C_7[1], 744_B_8[1], 771_H_4[1], 788_C_5[1], 792_F_9[1], 806_E_8[1], 823_D_7[1], 855_B_5[1], 893_D_8[4], 906_H_12[1], 911_A_2[4], 911_A_7[4], 911_H_8[4], 914_A_5[4]
n.d.	D22S666	776_A_2[1], 796_C_10[1]
n.d.	D22S715	803_D_11[4], 902_E_1[4]
n.d.	D22S718	882_A_6[4]
1A	ATP6E	422_G_4[1], 504_H_2[1], 758_H_11[1], 770_C_4[1], 803_G_9[1], 880_F_10[1], 924_C_12[1], 924_C_2[1]
1A	D22S9	172_G_7[2], 264_F_4[2], 42_E_12[2], 734_B_10[1], 765_E_2[1], 803_G_9[1], 829_D_11[1], 829_E_11[1], 925_G_12[1]
1A	D22S789	925_G_12[2]
1A	D22S795	734_B_10[1], 781_E_3[1], 813_A_3[1], 816_A_3[1], 973_A_6[1]
1A	F8VWFP	385_B_12[1], 453_G_6[1], l:603_H_9[1]
1A-1B	D22S24	100_G_7[2], 204_A_6[2], 204_A_9[2], 734_B_10[1], 829_D_11[1], 891_F_12[1], 925_G_12[2]
1A-1B	D22S50	745_G_7[2], 776_H_2[2], 829_D_11[2]
1A-1B	IGKVP3	210_E_12[2], 487_H_6[2], 891_C_8[1]
1A-9	D22S556	784_C_10[1], 884_E_1[1], 908_H_9[1]
1A-9	D22S609	966_A_8[1]
1A-9	D22S626	788_C_5[1], 908_H_9[1]
1B	D22S43	765_E_2[1], 924_C_2[1], 925_G_12[1]
2	D22S57	765_E_2[1], 803_G_9[1]
2	D22S137	685_E_2[2], 749_H_6[1], 803_G_9[1]
2	D22S420	730_H_4[1], 749_H_6[1], 765_E_2[1], 803_G_9[1], 925_G_12[1]
2	D22S111	791_F_9[2]
6	D22S184	759_E_10[1], 766_B_6[1], 792_F_9[2], 814_F_11[1], 861_D_9[1], 875_B_3[1], 893_H_6[1], 953_F_3[1], 977_G_5[1]
6	KI-1547	684_G_6[2], 729_F_9[1], 787_A_12[2], 849_E_9[2], 861_D_9[1]
7	BCRL2	266_A_4[3]
7,9,12	GGT1	741_G_7[1], 749_H_10[1], 759_F_8[1], 784_C_10[1], 785_E_4[1], 792_F_9[1], 831_B_6[1], 849_E_9[1], 858_F_11[1], 861_D_9[1], 874_E_2[1], 884_E_1[1], 891_F_12[1], 908_H_9[1], 916_A_3[1], 924_E_12[1], 951_F_11[1]
7	GGTX	266_A_4[3]
8A	CRKL	859_A_4[2]
8A	D22S117	623_B_1[1]
8A	D22S264	199_E_6[2], 295_G_9[2], 549_D_4[2], 67_A_3[2]

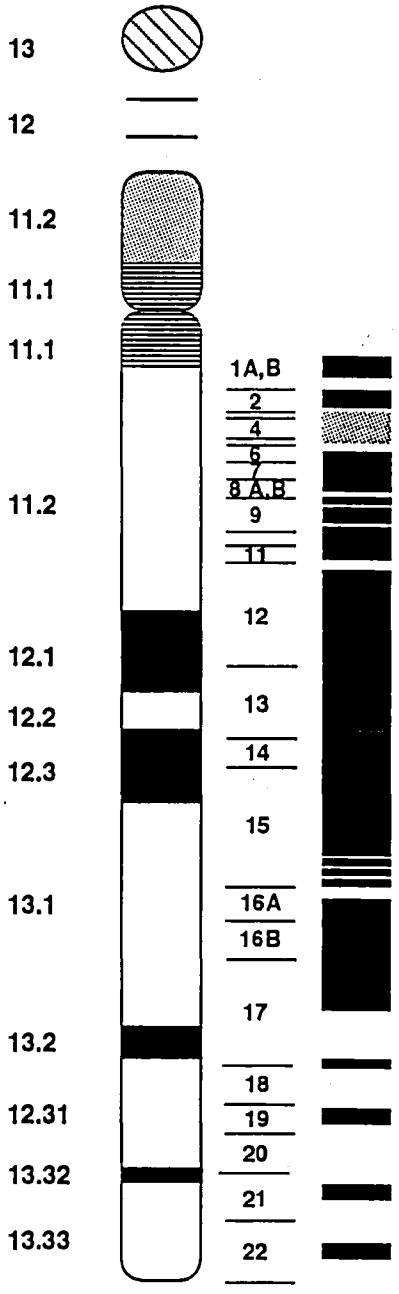
Bin	Locus	Positive YACs
8A	D22S306	11_F_10[1], 27_D_1[1], 366_F_5[1], 509_G_5[1], 872_F_9[1]
8A	D22S308	11_F_10[1], 27_D_1[1], 283_H_7[1], 398_G_5[1], 872_F_8[2], 872_F_9[1], 974_F_11[2]
8A	HCF2	118_D_3[2], 248_E_11[2], 301_G_8[2], 412_D_5[2], l:601_G_9[1], 742_B_5[1], 792_F_9[1], 944_C_4[1], 944_C_7[1], 952_F_5[1], 966_A_8[1]
8A	MEST39	742_B_5[1], 792_F_9[1], 944_C_4[1], 944_C_7[1]
9	BCRL4	147_D_3[3], 188_A_5[3], 191_A_11[3]
9	D22S303	267_D_3[1], 272_A_12[1], 417_H_7[1], 974_F_11[2]
9	D22S425	884_E_1[1], 908_H_9[1]
9	D22S563	884_E_1[1]
9	D22S655	191_A_11[1], 272_A_12[1], 417_H_7[1], l:603_F_10[1], 825_G_8[2], 836_H_11[2], 974_F_11[2]
9	GGTY	147_D_3[3], 188_A_5[3], 191_A_11[3]
9	GNAZ	118_B_11[2], 361_D_9[2], 54_C_8[2], 771_G_1[2]
9	IGL@	272_A_12[2], 417_H_7[2]
9	IGLC2	191_A_11[1], 272_A_12[2], 784_C_10[1], 874_A_4[1]
9-13	D22S119	873_C_4[1]
10	BCR	361_D_9[1], 446_B_5[1], 449_E_6[1], 874_A_4[1], 874_C_4[1]
10-11	D22S567	741_F_3[1], 778_H_7[1], 789_E_6[1], 820_H_4[1], 861_E_8[1], 869_G_8[1], 873_D_12[1], 895_F_12[1], 904_G_6[1], 907_C_7[1], 927_D_8[1], 938_F_9[1], 958_D_1[1], 958_F_1[1]
11	ADORA1	182_B_9[1], 24_H_10[1], 24_H_3[1], 339_A_2[2], 339_C_9[2], 818_E_12[1], 823_H_2[1], 838_E_4[2], 877_B_4[2], 877_E_2[2]
11	D22S156	765_E_3[1]
11	D22S794	l:601_G_11[3], 765_E_3[3], 767_B_2[3]
11,12	IGLLbb1	33_H_12[1], 34_H_1[1], 355_H_8[1], 357_A_6[1], 546_G_11[1], 546_G_12[1], 546_H_11[1], l:601_F_10[1], l:604_D_3[1], 784_C_10[1], 874_A_4[1], 883_H_11[1]
11	MMP11	346_H_10[1], 768_H_3[2]
11-12	D22S301	131_F_4[1], 131_F_5[1], 19_H_3[1], l:603_G_4[1], 784_C_10[1], 829_G_9[1], 874_A_4[1]
12	BCRL3	200_C_9[3], 220_G_3[3], 273_E_4[3]
12	CRYBB2	200_G_4[2], 200_G_6[2], 495_A_10[1], 78_F_2[1], 78_G_2[2], 904_B_10[2]
12	D22S1	119_D_6[2], 263_G_7[2], 445_H_8[2], 786_F_4[1], 797_A_2[1], 825_A_2[1], 825_A_9[2], 935_H_8[1]
12	D22S33	119_C_6[2], 190_D_6[2], 229_B_6[2], 445_H_8[2], l:601_H_1[1], l:603_H_1[1], 739_B_9[1], 786_C_12[2], 786_F_4[1], 788_B_12[1], 788_C_12[2], 797_A_2[1], 801_B_3[1], 825_A_2[1], 923_A_11[1], 935_H_8[1]
12	D22S42	269_G_7[2], 302_G_2[2], 302_G_3[2], 444_B_2[2], l:603_G_8[1], 789_E_6[1], 79_C_4[2], 808_G_1[1], 812_C_11[1], 820_H_4[1], 824_E_8[1], 861_E_8[1], 869_G_8[1], 873_D_12[1], 895_F_12[1], 904_G_6[1], 907_C_7[1], 927_D_8[1], 938_F_9[1], 958_F_1[1]
12	D22S56	179_H_10[r], 365_D_3[2], 449_E_11[2], 460_A_7[2], 715_B_2[1], 789_E_6[1,3], 817_F_7[1], 820_H_4[1], 927_D_8[1], 938_F_1[1], 938_F_9[1], 958_D_1[1], 958_F_1[1], A40_C_9[r], A86_G_1[1], D60_D_4[1], D64_F_8[1], D98_H_2[1]
12	D22S72	153_B_11[2], 200_G_6[2], 324_D_3[2], 778_E_6[2], 808_G_1[2], 812_C_11[2], 817_F_7[1], 861_E_8[2], 873_C_2[2], 927_D_8[2], 938_F_1[2], 938_F_9[2]
12	D22S186	220_G_2[2], 786_C_11[1], 786_F_4[1], 797_A_2[1], 825_A_2[1], 923_A_11[1], 935_H_8[1], 949_E_6[1]
12	D22S190	218_D_10[2], 239_G_2[2], 398_E_2[2], 773_G_1[2], 778_H_7[1], 786_C_11[1], 786_C_12[1], 797_A_2[1], 799_B_9[1], 806_A_1[1], 839_G_4[1], 887_B_8[1], 887_D_4[1], 904_B_10[2], 928_H_10[1], 953_D_10[1]
12	D22S192	229_B_6[2], 330_H_5[2], 43_G_1[2], 445_B_9[2], 739_B_9[1], 786_F_4[1], 797_A_2[1], 825_A_2[1], 935_H_8[1], 99_A_11[2], 99_C_6[2]
12	D22S193	815_E_7[1], 820_H_4[1], 927_D_8[1], 949_E_6[1]
12	D22S310	205_A_7[1], 239_G_2[1], 398_E_2[1], 739_B_9[1], 786_C_11[1], 797_A_2[1], 806_A_1[1], 839_G_4[1], 887_B_8[1], 887_D_4[1], 904_B_10[1], 928_H_10[1], 935_H_8[1], 953_D_10[1]
12	D22S315	395_H_9[2], 537_C_8[2], 777_D_1[1], 817_F_7[1], 904_B_10[1], 953_D_10[1], 965_E_10[1]

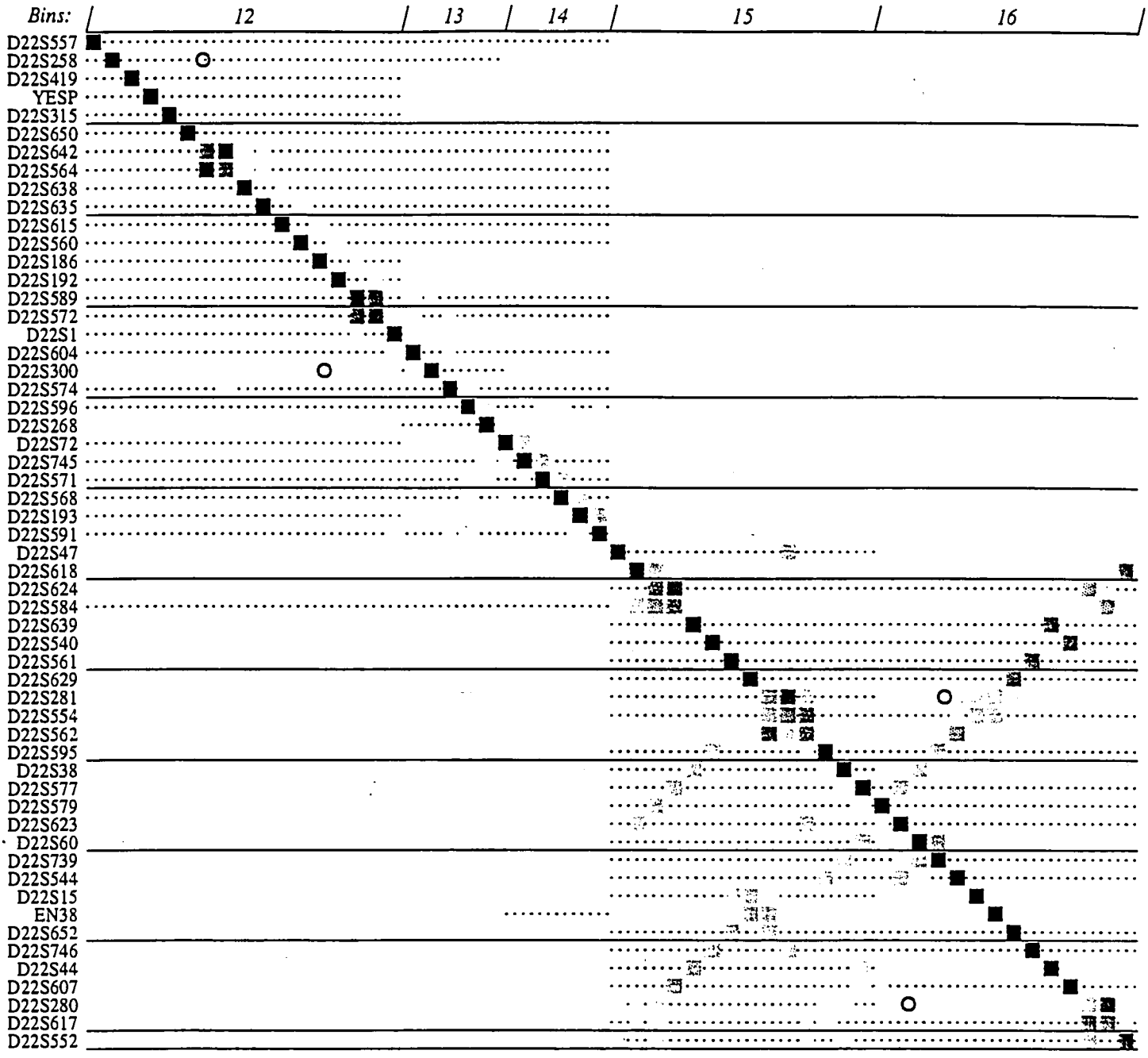
Bin	Locus	Positive YACs
12	D22S351	205_A_7[2], 239_G_2[2], 398_E_2[2], 773_G_1[2], 786_C_11[1], 786_G_1[2], 797_A_2[1], 806_A_1[1], 839_G_4[1], 887_B_8[1], 887_D_4[1], 904_B_10[2], 928_H_10[1], 935_H_8[1], 953_D_10[1]
12	D22S419	777_D_1[1], 806_A_1[1], 824_E_8[1], 839_G_4[1], 953_D_10[1], 953_E_6[1], 965_E_10[1]
12	D22S431	102_H_6[2], 501_C_8[2], 832_A_11[2], 832_H_3[2]
12	EWSex5	418_A_2[e]
12	-GGTZ	200_C_9[3], 220_G_3[3], 273_E_4[3]
12	TOP1P2	148_H_6[2], 251_F_2[2], 354_B_7[2], 512_B_10[2], l:603_F_8[1]
12	YESP	159_E_10[2], 182_G_1[2], 188_B_7[2], 395_H_9[2], 777_D_1[1], 953_D_10[1], 965_E_10[1]
12-13	D22S258	129_C_10[2], 222_C_8[2], 341_E_2[2], 402_G_7[2], 777_D_1[1], 953_E_6[1], 965_E_10[1]
12-13	EWSR1	210_B_7[2]
12-14	D22S541	873_C_2[1], 961_B_2[4]
12-14	D22S557	953_E_6[1], 965_E_10[1]
12-14	D22S560	786_C_11[1], 797_A_2[1], 825_A_2[1], 839_G_4[1], 887_B_8[1], 923_A_11[1], 928_H_10[1], 935_H_8[1], 949_E_6[1]
12-14	D22S564	786_C_11[1], 797_A_2[1], 806_A_1[1], 887_B_8[1], 904_B_10[1], 953_D_10[1], 965_E_10[1]
12-14	D22S566	778_H_7[1], 789_E_6[1], 808_G_1[1], 812_C_11[1], 820_H_4[1], 824_E_8[1], 828_E_8[1], 839_G_4[1], 861_E_8[1], 869_G_8[1], 873_D_11[1], 873_D_12[1], 887_D_4[1], 895_F_12[1], 904_G_6[1], 907_C_7[1], 927_D_8[1], 927_D_9[1], 938_F_9[1], 958_D_1[1]
12-14	D22S568	715_B_2[1], 820_H_4[1], 927_D_8[1], 938_F_9[1], 958_D_1[1], 958_F_1[1]
12-14	D22S569	744_F_6[1], 750_B_4[1], 759_E_6[1], 763_A_3[1], 773_G_1[1], 774_G_1[1], 793_E_9[1], 856_C_1[1], 858_H_8[1], 873_C_2[1], 880_E_10[1], 882_A_6[1], 925_G_8[1], 935_H_8[1], 939_D_5[1]
12-14	D22S570	710_G_6[1], 728_B_4[1], 739_B_9[1], 750_B_4[1], 778_F_1[1], 782_H_7[1], 789_E_6[1], 808_G_1[1], 812_C_11[1], 820_H_4[1], 825_A_2[1], 830_H_8[1], 858_H_8[1], 861_E_8[1], 869_G_8[1], 873_C_2[1], 873_D_11[1], 873_D_12[1], 927_D_8[1], 938_F_9[1], 958_D_1[1], 958_F_1[1]
12-14	D22S571	778_H_7[1], 789_E_6[1], 812_C_11[1], 820_H_4[1], 873_D_12[1], 895_F_12[1], 904_G_6[1], 907_C_7[1], 927_D_8[1], 928_H_10[1], 938_F_9[1], 958_F_1[1]
12-14	D22S572	739_B_9[1], 786_F_4[1], 788_B_12[1], 923_A_11[1], 935_H_8[1], 941_C_5[1]
12-14	D22S574	786_C_11[1], 797_A_2[1], 806_A_1[1], 927_D_8[1], 938_F_9[1]
12-14	D22S576	741_F_3[1], 778_H_7[1], 789_E_6[1], 808_G_1[1], 812_C_11[1], 820_H_4[1], 861_E_8[1], 869_G_8[1], 873_D_11[1], 873_D_12[1], 895_F_12[1], 899_G_4[1], 904_G_6[1], 907_C_7[1], 927_D_8[1], 928_H_10[1], 938_F_9[1], 958_D_1[1], 958_F_1[1]
12-14	D22S582	769_B_11[1]
12-14	D22S584	750_B_4[1], 763_A_3[1], 769_B_11[1]
12-14	D22S588	803_D_3[1]
12-14	D22S589	739_B_9[1], 786_F_4[1], 788_B_12[1], 923_A_11[1], 935_H_8[1]
12-14	D22S591	788_B_12[1], 801_B_3[1], 949_E_6[1]
12-14	D22S594	789_E_6[1], 808_G_1[1], 812_C_11[1], 820_H_4[1], 824_E_8[1], 861_E_8[1], 869_G_8[1], 895_F_12[1], 904_G_6[1], 907_C_7[1], 927_D_8[1], 938_F_9[1], 958_D_1[1], 958_F_1[1]
12-14	D22S596	817_F_7[1], 927_D_8[1], 938_F_9[1], 954_A_1[1]
12-14	D22S604	786_F_4[1], 797_A_2[1], 825_A_2[1], 958_F_1[1]
12-14	D22S615	786_C_11[1], 786_C_12[1], 797_A_2[1], 825_A_2[1], 839_G_4[1], 928_H_10[1], 935_H_8[1], 949_E_6[1]
12-14	D22S631	949_E_6[1]
12-14	D22S635	786_C_11[1], 797_A_2[1], 806_A_1[1], 825_A_2[1], 839_G_4[1], 887_B_8[1], 928_H_10[1], 935_H_8[1], 953_D_10[1]
12-14	D22S638	786_C_11[1], 786_C_12[1], 788_B_12[1], 797_A_2[1], 806_A_1[1], 839_G_4[1], 887_B_8[1], 887_D_4[1], 904_B_10[1], 928_H_10[1], 953_D_10[1]
12-14	D22S642	739_B_9[1], 786_C_11[1], 786_C_12[1], 797_A_2[1], 806_A_1[1], 887_B_8[1], 904_B_10[1], 953_D_10[1], 965_E_10[1]

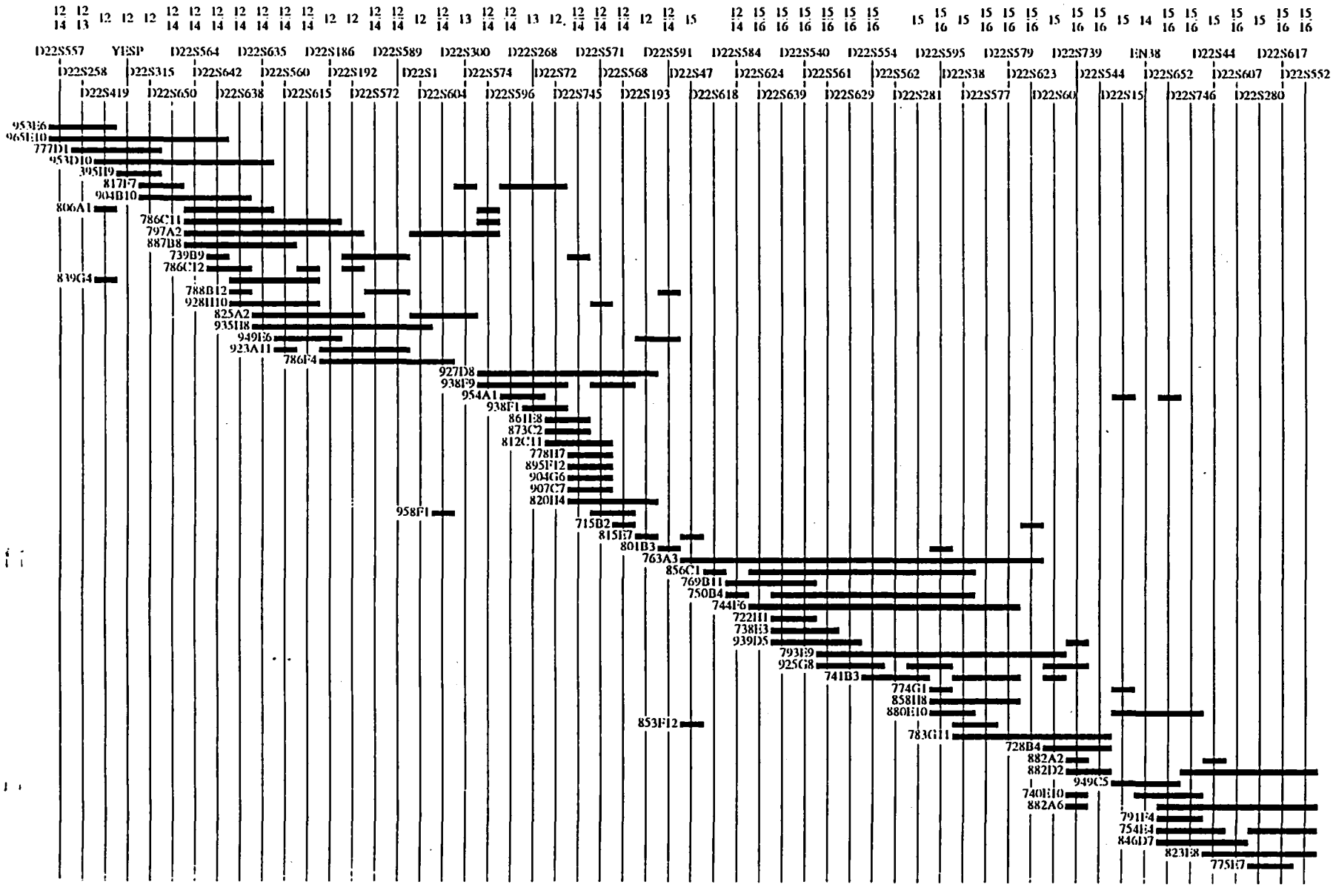
Bin	Locus	Positive YACs
12-14	D22S650	817_F_7[1], 904_B_10[1], 953_D_10[1], 965_E_10[1]
12-14	D22S653	788_B_12[1], 905_H_8[4], 907_C_7[1]
12-14	D22S669	715_B_2[1], 778_F_1[1], 782_H_7[1], 789_E_6[1], 808_G_1[1], 812_C_11[1], 820_H_4[1], 830_H_8[1], 858_H_8[1], 861_E_8[1], 869_G_8[1], 873_C_2[1], 958_F_1[1], 962_D_4[1]
12-14	D22S745	731_G_9[4], 739_B_9[1], 778_H_7[1], 812_C_11[1], 820_H_4[1], 861_E_8[1], 869_G_8[1], 873_C_2[1], 895_F_12[1], 904_G_6[1], 907_C_7[1], 927_D_8[1]
13	D22S268	179_H_10[2], 285_B_1[1], 414_F_9[2], 710_G_6[1], 817_F_7[1], 873_G_8[1], 927_D_8[1], 927_D_9[1], 938_F_1[1], 938_F_9[1], 954_A_1[1], A226_C_4[e]
13	D22S300	185_A_8[1], 387_F_1[1], 429_F_10[1], 797_A_2[1], 817_F_7[1], 825_A_2[1], 945_C_1[1], 974_H_8[2], 9_G_2[1]
13	D22S447	306_E_3[1], 768_C_6[1], 779_C_7[1]
13	EWS3'	B84_D_4[e]
13	LIF	911_F_12[2], A226_C_4[1], B125_A_9[1], D45_B_10[1]
13	MEST14	504_B_10[2]
13	NEFH	104_H_6[2], 799_B_9[3], 799_B_9[2], 817_F_7[1], 922_D_8[2], B186_A_7[e], B63_F_6[e]
13	OSM	A226_C_4[e]
13	TCN2	351_D_8[2], 419_E_6[2], l:601_F_9[1], 768_C_6[1,3]
14	D22S37	284_B_11[2], 358_G_6[2], 393_E_7[2], l:604_B_6[3], 769_B_1[1], 776_A_2[1,3]
14	D22S776	l:604_B_11[3], l:604_B_6[3], 776_A_2[3]
14	EN38	740_E_10[1], 778_E_1[1], 880_E_10[1], 880_E_12[1], 949_C_5[1]
15	22-5	744_E_1[2], 769_B_5[2]
15	D22S15	315_F_10[2], 447_D_4[2], 72_F_8[2], 774_G_1[1], 880_E_10[1], 949_C_5[1], 954_A_1[1]
15	D22S28	809_C_6[1], 844_H_7[1], 908_C_6[1], 912_F_5[1]
15	D22S29	738_E_3[1], 744_F_6[1], 750_B_4[1], 759_E_6[1], 763_A_3[1], 793_E_9[1], 799_B_9[1], 801_B_3[1], 839_G_4[1], 856_C_1[1], 858_H_8[1], 873_D_12[1], 880_E_10[1], 925_G_8[1], 939_D_5[1]
15	D22S38	l:601_C_2[3], 741_B_3[1,3], 744_F_6[1], 750_B_4[1], 763_A_3[1,3], 783_G_11[1,3], 793_E_9[1,3], 828_E_8[1], 853_F_12[1], 856_C_1[1], 858_H_8[1], 880_E_10[1], 935_H_8[1]
15	D22S44	233_B_10[1], 264_E_4[1], 425_C_8[1], 434_E_11[1], 438_H_10[1], 462_E_11[1], 462_E_12[1], 462_F_5[1], 754_E_4[1], 823_E_8[1], 846_D_7[1], 882_A_2[1], 882_A_6[1], 882_D_2[1]
15	D22S47	l:603_B_4[3], l:603_C_4[3], 744_A_7[2], 763_A_3[3], 794_H_3[2], 815_E_7[2], 853_F_12[2], 934_E_1[2], 943_E_1[2]
15	D22S51	768_B_7[1], 768_C_6[1], 769_A_9[1], 848_D_3[1]
15	D22S58	826_D_12[2], 891_D_12[2]
15	D22S60	180_C_6[2], 416_F_8[1], 439_A_10[1], 465_G_10[1], 487_D_10[1], 523_F_5[1], 541_G_2[1], 728_B_4[1], 741_B_3[1], 783_G_11[1], 786_G_1[1], 793_E_9[1], 925_G_8[1]
15	D22S61	366_B_4[2], 488_B_7[2], 769_B_11[2]
15	D22S91	981_A_1[2]
15	D22S102	151_C_5[2], 525_B_11[2], 75_G_4[2]
15	D22S273	949_E_6[1]
15	D22S277	444_F_3[2], 529_A_10[2], l:602_B_10[1], l:602_B_5[1]
15	D22S278	107_F_2[2], 114_E_5[2], 124_F_2[2], 506_C_6[2], 59_B_12[2]
15	D22S280	754_E_4[1], 775_E_7[1], 823_E_8[1], 882_A_6[1], 882_D_2[1]
15	D22S281	741_B_3[1], 744_F_6[1], 750_B_4[1], 763_A_3[1], 793_E_9[1], 802_B_11[4], 856_C_1[1], 925_G_8[4]
15	D22S283	712_A_3[1], 881_G_4[1]
15	D22S292E	849_E_1[1], 854_E_4[2]
15	D22S304	157_D_4[1], 204_F_8[1], 366_B_4[1], 402_E_5[1], 488_B_6[1], 741_B_3[1], 769_B_11[1]
15	D22S412E	849_E_1[2]
15	D22S778	897_A_10[2]

Bin	Locus	Positive YACs
15	D22S792	849_E_1[1,3]
15	D22S793	l:603_D_9[3], l:603_G_8[3], l:603_H_11[3]
15	D22S799	l:602_A_8[3]
15	IL2RB	206_E_7[1]
15	PVALB	858_H_8[1], 882_C_9[2]
15	SGLT1	134_G_3[1], 32_H_11[1], 335_A_3[1], 416_E_11[1], 740_E_10[1], 754_E_4[1], 791_F_4[1], 846_D_7[1], 880_E_10[1], 882_A_6[1], 902_A_1[1], 909_C_8[1], 949_C_5[1], 954_A_1[1]
15	TIMP3	233_B_10[1]
15-16B	D22S540	722_H_1[4], 738_E_3[1], 744_F_6[1], 750_B_4[1], 763_A_3[1], 763_A_4[2], 769_B_11[1], 856_C_1[1], 939_D_5[1]
15-16B	D22S544	728_B_4[1], 778_G_9[4], 783_G_11[1], 882_D_2[4]
15-16B	D22S552	754_E_4[1], 823_E_8[1], 882_A_6[1], 882_D_2[1]
15-16B	D22S554	741_B_3[1], 744_F_6[1], 750_B_4[1], 759_E_6[1], 763_A_3[1], 793_E_9[1], 856_C_1[1], 925_G_8[1]
15-16B	D22S561	163_E_9[2], 738_E_3[1], 744_F_6[1], 750_B_4[1], 763_A_3[1], 793_E_9[1], 856_C_1[1], 91_A_4[2], 925_G_8[1], 939_D_5[1]
15-16B	D22S577	741_B_3[1], 744_F_6[1], 763_A_3[1], 783_G_11[1], 793_E_9[1], 853_F_12[1], 858_H_8[1]
15-16B	D22S579	741_B_3[1], 744_F_6[1], 763_A_3[1], 783_G_11[1], 793_E_9[1], 858_H_8[1]
15-16B	D22S595	744_F_6[1], 750_B_4[1], 755_E_12[1], 763_A_3[1], 773_G_1[1], 774_G_1[1], 793_E_9[1], 801_B_3[1], 856_C_1[1], 858_H_8[1], 880_E_10[1], 925_G_8[1]
15-16B	D22S607	823_E_8[1], 846_D_7[1], 882_A_6[1], 882_D_2[1]
15-16B	D22S617	754_E_4[1], 775_E_7[1], 823_E_8[1], 882_A_6[1], 882_D_2[1], 882_D_6[1]
15-16B	D22S620	882_G_4[1]
15-16B	D22S623	715_B_2[1], 763_A_3[1], 783_G_11[1], 793_E_9[1]
15-16B	D22S624	744_F_6[1], 763_A_3[1], 769_B_11[1], 776_E_10[1], 856_C_1[1]
15-16B	D22S629	744_F_6[1], 750_B_4[1], 763_A_3[4], 793_E_9[1], 856_C_1[1], 925_G_8[1], 939_D_5[1]
15-16B	D22S630	803_D_3[1], 948_B_2[1]
15-16B	D22S633	159_D_4[2], 366_B_4[2], 488_B_6[2], 526_A_9[2], 744_E_1[2], 769_B_1[2], 769_B_11[1]
15-16B	D22S639	722_H_1[4], 738_E_3[1], 744_F_6[1], 750_B_4[1], 763_A_3[1], 769_B_11[1], 856_C_1[1], 939_D_5[1]
15-16B	D22S644	882_G_4[1]
15-16B	D22S652	740_E_10[1], 754_E_4[1], 791_F_4[1], 846_D_7[1], 880_E_10[1], 882_A_6[1], 909_C_8[1], 949_C_5[1], 954_A_1[1]
15-16B	D22S739	724_A_9[4], 728_B_4[1], 740_E_10[1], 781_F_1[1], 783_G_11[1], 882_A_2[1], 882_A_6[1], 882_D_2[1], 925_G_8[1], 939_D_5[1]
15-16B	D22S746	740_E_10[1], 754_E_4[1], 791_F_4[1], 846_D_7[1], 880_E_10[1], 882_A_6[1], 882_D_2[1]
16A	D22S299	159_F_1[1], 251_E_4[1], 409_E_5[1], 414_F_1[1], 419_E_9[2], 496_B_7[2], 514_H_10[2], 532_H_5[2], l:603_B_2[1], 803_D_3[1], 924_C_2[1], 948_B_2[1]
16A	D22S302	409_E_5[1], 419_E_9[1], l:603_B_2[1], 803_D_3[1], 924_C_2[1], 948_B_2[1]
16A	PDGFB	207_B_1[2]
16B	D22S279	140_E_12[2], 238_B_2[2], 310_E_1[2], 428_F_2[2], 494_D_9[2], 793_F_6[2], 803_D_3[1], 807_F_1[2], 893_D_6[2]
17	CYP2D8P	148_H_11[1], 151_C_6[1], 177_A_8[1], l:603_B_9[3], 730_B_7[1], 803_D_3[1,3]
17	D22S307	61_E_6[1], 755_E_12[1], 759_E_6[1], 803_D_3[1], 84_C_7[1]
17	G22P1	375_B_3[2], 375_G_4[2], 435_C_9[2], 435_D_3[2], l:603_C_9[1], l:604_C_4[1], 752_C_9[4], 781_D_5[1], 789_D_9[1], 791_B_12[1], 850_D_12[2], 884_H_2[4], 905_F_4[2], 912_G_11[4], 950_D_11[4], 967_C_12[4], 967_C_12[2]
17	NAGA	181_C_9[2], 190_C_9[2], 263_B_7[2], 730_B_7[1], 803_D_3[1], 84_C_7[1]
17-19	D22S565	774_E_5[1], 930_A_11[1], 961_D_4[1]
17-22	D22S546	766_G_6[4]
18	D22S417	894_F_9[2]

Bin	Locus	Positive YACs
18	DIA1	391_C_6[1], l:604_C_2[3], 666_F_9[2], 675_G_9[2]
19	ACR	124_C_2[2], 220_D_2[2], 508_D_4[2], 508_D_5[2], 80_B_6[2], 896_A_8[2], 918_E_8[2]
19	D22S270	825_A_2[1], 836_E_11[2]
19	D22S418	871_D_10[1], 930_A_11[1]
20	BZRP	127_C_4[1], 243_G_1[1], 273_G_10[1], l:601_B_1[3], 954_D_7[2]
20	D22S64	314_E_10[2], 316_E_10[2], 844_E_10[2]
20	D22S282	736_A_3[1]
20	D22S297	13_A_3[1], 149_A_2[1], 16_F_2[1], 170_A_11[1], l:603_G_1[1]
20-22	D22S611	715_C_1[1], 715_C_2[1]
21	D22S40	523_D_6[2], 523_G_5[2], 778_F_10[2], 778_F_11[2], 778_F_9[2]
21	D22S274	131_A_5[2], 471_C_10[2], 507_C_11[2], 53_C_7[2], 776_B_9[2], 895_E_12[2]
21	D22S294	120_E_2[2], 253_C_12[2], 335_A_10[2], 463_C_10[2]
21	D22S781	262_F_9[1]
21	FIBB	740_B_12[2], 891_D_12[2]
21-22	Z7	120_E_4[1], 140_F_4[1], 253_C_12[2], 293_A_10[1], 790_F_9[2], 854_D_7[1], 915_B_9[2], 915_D_9[1]
22	ARSA	l:603_F_3[1,3]
22	CB10	370_B_6[2]
22	D22S23	765_F_6[1]
22	D22S45	156_A_12[2], 318_E_10[2], 318_E_4[2], 546_C_12[2]
22	D22S55	17_D_8[1], 412_H_3[1]
22	D22S63	228_E_12[1], 337_G_1[1], 361_G_1[1], 361_G_12[1], 368_C_5[1], 421_G_1[1], 61_B_11[1]
22	D22S295	715_C_1[1], 715_C_2[1], 716_C_1[1], 724_B_2[1], 741_C_1[1], 774_E_5[1], 786_G_10[1], 809_B_5[1], 930_A_11[1]


13
12
11.2
11.1
11.1
11.2
12.1
12.2
12.3
13.1
13.2
12.31
13.32
13.33
1A,B
2
4
5
8 A,B
9
11
12
13
14
15
16A
16B
17
18
19
20
21
22


Bins:
12
13
14
15
16
D22S557
D22S258
D22S419
YESP
D22S315
D22S650
D22S642
D22S564
D22S638
D22S635
D22S615
D22S560
D22S186
D22S192
D22S589
D22S572
D22S1
D22S604
D22S300
D22S574
D22S596
D22S268
D22S72
D22S745
D22S571
D22S568
D22S193
D22S591
D22S47
D22S618
D22S624
D22S584
D22S639
D22S540
D22S561
D22S629
D22S281
D22S554
D22S562
D22S595
D22S38
D22S577
D22S579
D22S623
D22S60
D22S739
D22S544
D22S15
EN38
D22S652
D22S746
D22S44
D22S607
D22S280
D22S617
D22S552


D22S557
YESP
D22S564
D22S635
D22S186
D22S589
D22S300
D22S268
D22S571
D22S591
D22S584
D22S540
D22S554
D22S595
D22S579
D22S739
EN38
D22S44
D22S617
D22S258
D22S315
D22S642
D22S560
D22S192
D22S1
D22S574
D22S72
D22S568
D22S47
D22S624
D22S561
D22S562
D22S38
D22S623
D22S544
D22S652
D22S607
D22S552
D22S419
D22S650
D22S638
D22S615
D22S572
D22S604
D22S596
D22S745
D22S193
D22S618
D22S639
D22S629
D22S281
D22S577
D22S60
D22S15
D22S746
D22S280
953E6
965E10
777D1
953D10
395H9
817F7
904B10
806A1
786C11
797A2
887D8
739B9
786C12
839G4
788B12
928H10
825A2
935H8
949E6
923A11
786F4
927D8
938F9
954A1
938F1
861E8
873C2
812C11
778H7
895F12
904G6
907C7
820H4
958F1
715B2
815F7
801B3
763A3
856C1
769B11
750B4
744F6
722H1
738H3
939D5
793E9
925G8
741B3
774G1
858H8
880E10
853F12
783G11
728B4
882A2
882D2
949C5
740E10
882A6
791F4
754E4
846D7
823E8
775H7

DEPARTMENT OF HEALTH & HUMAN SERVICES

Public Health Service

National Institutes of Health
National Center for Human
Genome Research
Bethesda, Maryland 20892

Building 38A, Room 610
(301) 496-7531
(301) 480-2770 (Fax)

JUN 9 1993

MEMO

TO: Members of the Scientific Community

FROM: NCHGR Staff

RE: Summary of goals for the Whitehead Institute (Eric Lander, PI) and University of Iowa (Jeff Murray, PI) Centers grants

In response to requests from the scientific community, NCHGR is providing the attached summaries, as prepared by the grantees themselves, of the goals and organization of the centers recently established at the Whitehead Institute (Eric Lander, P.I.) and the University of Iowa (Jeff Murray, P.I.).

Consistent with their policies of releasing data to the genomics community and to facilitate a fast start and rapid integration of the maps, these two mapping groups have already discussed and identified many existing markers and resources that will be exchanged and mapped in both systems. The two centers will exchange data (which will also be publically available) to ensure that they efficiently use new markers as they are developed (for example, the low heterozygosity STRPs that the Iowa group will develop, but not map, will be placed on the physical map by the Whitehead group). The use of similar PCR conditions and consistency in the manner in which oligonucleotide primers are distributed should also facilitate the dissemination of useful reagents to each other and to the community.

The names, addresses, and telephone and fax numbers of the P.I.s are provided and we suggest that you contact them if you have specific questions. We hope you will find this information useful.

Whitehead/MIT Center for Genome Research

This report briefly summarizes the goals and organization of the Whitehead/MIT Center for Genome Research (CGR), with the aim of assisting colleagues interested in obtaining information or materials generated by the Center.

Organization. The CGR formally commenced under a grant starting on March 1, 1993. A renewal of a previous genome center grant focusing solely on the mouse genome, the new center has expanded its scope to include both the mouse and human. Laboratory space is currently under renovation with expected completion by June 1, 1993. CGR will begin full-scale operations at this time, with smaller scale efforts underway until then.

The Center involves five institutions: Whitehead Institute for Biomedical Research, Massachusetts Institute of Technology, Princeton University, The Jackson Laboratory, and the Centre d'Etude de Polymorphisme Humaine.

Personnel. CGR is overseen by Eric Lander (Director), David Page (Associate Director) and Nat Goodman (Associate Director). The other members are Daniel Cohen (CEPH), Nic Dracopoli (MIT), Rudolf Jaenisch (Whitehead/MIT), Paul Matsudaira (Whitehead/MIT), Joseph Nadeau (JAX), James Orlin (MIT), and Shirley Tilghman (Princeton).

Overall Goals. CGR's principal goals are (1) to construct a genetic and physical map of the mouse genome and physical map of the human genome, and (2) to make these maps readily accessible to the scientific community.

Specific Plans: Mouse Genome. The Mouse Genome Mapping project is funded for five years and involves both genetic and physical mapping.

The five-year goal for genetic mapping is (i) to create a genetic map with 6,000 simple sequence length polymorphisms (SSLPs) based on a low resolution cross with crossovers at an average spacing of 1 cM; (ii) to integrate our SSLP map with the gene-based map, by typing one-quarter of the SSLPs in a subset of the interspecific backcross of Copeland and Jenkins; and (iii) to carry out finer resolution mapping in cross with crossovers at an average spacing of 0.1 cM.

The five-year goal for physical mapping is (i) to produce a YAC library providing >8-fold coverage of the mouse genome in clones of average size 700 kb (our current library provides roughly 4-fold coverage); (ii) to construct an STS content map of the mouse genome consisting of 10,000 STSs (consisting of the 6,000 SSLPs and 4,000 random STSs); and (iii) to achieve sufficient closure that the physical map consists of average contigs of approximately 10-20 Mb.

We will initially focus our attention on increasing the density of the genetic map and expanding the YAC library. These priorities are based on extensive conversations with the mouse community. (In short, for positional cloning, it is more valuable to have a closer genetic marker that has not yet been screened against the YAC library than a more distant genetic marker that has been screened.)

We have set the following tentative goals:

Year	Genetic Goal	Physical Goal
Year 1	2800 SLLPs total	0 STS total
Year 2	4600 SLLPs total	1500 STS total
Year 3	6000 SLLPs total	4500 STS total
Year 4	integration/ fine structure	8500 STS total
Year 5	integration/ fine structure	10,000 STS total/ closure

Thus, we will initially serve the mouse community by disseminating genetic markers and YAC libraries. We will start to screen SSLPs in the YAC library in year 2 and should be able to provide the corresponding addresses beginning in the middle of year 2. We will integrate and disseminate this information through the public databases (e.g., the Jackson Lab databases). If there is community interest, we are also prepared to set up an on-line local database as well as to serve as a clearinghouse for other information about STS-YAC-contig data.

Specific Plans: Human Genome. The Human Genome Mapping project is funded for three years and involves physical mapping.

The three-year goal for physical mapping is to construct an STS content map of the human genome consisting of at least 8500 STSs. These STSs will consist of some 4250 SSLP genetic markers (from among those being developed by Jean Weissenback's group at Genethon, Jeff Murray's Genome Center, and the general community) and some 4250 random STSs (which we are generating and assigning to chromosomes by means of somatic cell hybrids). The STSs will be initially typed in the CEPH megabase YAC library. (In addition, we are currently attempting to develop a large-insert YAC library with significantly reduced chimerism. If we are successful, we expect to screen this library as well.)

We have set the following tentative goals:

Year	Physical Goal
Year 1	1500 STS total
Year 2	4500 STS total
Year 3	8500 STS total

If the project is successful, we hope to continue for an additional two years to achieve essentially complete closure.

We will initially serve the community by screening known SSLPs and random STSs against the YAC library and providing addresses. We will clearly not have large contigs based on STS content mapping until a high density of STSs are screened (certainly not before the middle of year (2). In addition, we will have no information about subchromosomal localization of the random STSs until then.

Because the same YAC library has been fingerprinted by Daniel Cohen's group at Genethon, however, it should be possible to use the SSLPs as anchors for fingerprint-based contigs. Together with Cohen, we will integrate and disseminate this information through the public databases (e.g., GDB). If there is community interest, we are also prepared to set up an on-line local database as well as to serve as a clearinghouse of other information about STS-YAC-contig data. Discussions are currently underway concerning setting up such a database. We would hope to have it available by the end of summer 1993.

Data Release. CGR is committed to prompt release of map data. Data will be released in batches as soon as it has been confirmed. We expect to make data releases for the end of each calendar quarter--with the first release being June 30, 1993. We invite suggestions about how to make data release and distribution most useful to the community.

Access to Materials. We will continue to devise arrangements to ensure accessibility of materials to the community. All STSs (including SSLPs) are made available under an arrangement with Research Genetics, Inc. designed to ensure inexpensive access to PCR primer pairs. We have also arranged for distribution of our mouse YAC library and the CEPH mega-YAC to interested companies willing to provide any of the following services: library replication; sale of DNA pools for PCR screening; and contract screening services. To date, several companies have expressed interest. In addition, Shirley Tilghman's lab at Princeton maintains a core facility providing DNA pools for PCR screening of the mouse YAC library.

Commercialization Policy. CGR has adopted a strict policy governing the commercialization of the genomic maps (including all clones, genetic markers, primers and sequences) developed under its auspices. The policy states that: (a) The maps will be made promptly available to the scientific community, will be placed in the public domain, and will not be patented; and (b) No advance access to the maps will be granted to any commercial entity in advance of public access.

Commercial arrangements to ensure distribution of STSs and YACs are entered into on a non-exclusive basis and involve no license fee or other financial compensation to CGR or its personnel.

For Additional Information: Please write to: Eric Lander, Center for Genome Research, Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142. Fax: (617) 258-6505. Phone: (617) 258-5192.

■ COOPERATIVE HUMAN LINKAGE CENTER ■

CHLC REPORT

VOLUME 1, NUMBER 1, MAY, 1993

INTRODUCTION

Jeffrey C. Murray, M.D.
Principal Investigator

This newsletter represents the first report from the Cooperative Human Linkage Center (CHLC) established by the NCHGR in Fall, 1992. We have included short descriptions of each of the involved projects, which are located at The University of Iowa, Fox Chase Cancer Center, Marshfield Medical Research Foundation and Harvard Medical School. In addition to short project descriptions, we have included the first round of genetic maps developed by the center.

The long-range goal of the center is to develop high heterozygosity genetic maps that are greatly enriched for the presence of easy-to-use PCR-formatted microsatellite markers, with a particular emphasis on tri- and tetranucleotide repeats that are easy to genotype. The grant will synthesize published genotypic data developed on the CEPH families by outside investigators, as well as genotypic information generated from marker development in CHLC core laboraotires. The center is also open to assisting outside investigators who would like incorporation of their own genotypic information into these maps, as well.

The maps presented here are a preliminary synthesis of publicly available genotypic information existing in the CEPH database and are seeded with the first sets of markers developed through our own efforts. We provide information for online access to a CHLC database of these markers and maps which will be revised collectively at approximately six month intervals. In addition, information and access to markers will be provided, both as an online service and through direct reagent access facilitated through primer availability at Research Genetics.

We will continue to work with others to bring genetic maps to a high degreee of resolution and to facilitate disease gene mapping using a variety of strategies that benefit from the availability of highly polymorphic markers. Such strategies include not only linkage analysis, but also studies of non-traditional inheritance such as imprinting, locus expansion, and loss of heterozygosity studies. In addition, the markers developed in this center will also provide STSs for physical mapping efforts currently underway. All markers developed will be assigned chromosomal localizations, and although only those markers with heterozygosities above 0.7 will initially be genotyped and entered into the linkage maps, all markers with chromosomal assignments will be made available for efforts by other laboratorties for genetic or physical mapping.

We welcome comments and suggestions pertaining to the newsletter and our plans and these can be communicated directly by e-mail, phone or fax to any of the relevant co-investigators or contacts listed at right.

Jeffrey C. Murray, M.D.
Assistant Professor of Pediatrics
The University of Iowa
Iowa City, IA 52242
TEL: (319) 356-3508
FAX: (319) 335-6970
E-mail: jeff-murray@umaxc.weeg.uiowa.edu

Geoffrey M. Duyk, M.D., Ph.D.
Department of Genetics, EQRF Room 447
Harvard Medical School, 200 Longwood Ave.
Boston, MA 02115
TEL: (617) 432-6072
FAX: (617) 432-7663
E-mail: duyk@rascal.med.harvard.edu

Val C. Sheffield, M.D., Ph.D.
Assistant Professor of Pediatrics
The University of Iowa
Iowa City, IA 52242
(319) 356-2674
FAX: (319) 356-3347.
E-mail: sheffield@vaxa.weeg.uiowa.edu

James L. Weber, Ph.D.
Senior Scientist, Human Genetics
Marshfield Medical Research Foundation
Marshfield, WI 54449
TEL: (715) 387-9179
FAX: (715) 389-3808
email: weberj@dgabby.mfldclin.edu

Kenneth H. Buetow, Ph.D.
Fox Chase Cancer Center
7701 Burholme Avenue
Philadelphia, PA 19111
TEL: (215) 728-3152
FAX: (215) 728-3574
E-mail: kh_buetow@fccc.edu

Robert F. Weir, Ph.D.
Professor of Pediatrics
University of Iowa
Iowa City, IA 52242
TEL: (319) 335-6705
FAX: (319) 335-8318

Nancy Newkirk
CHLC Administration
The University of Iowa
TEL: (319) 335-6899
FAX: (319) 335-6970

PROJECT 1

Geoff Duyk, M.D., Ph.D.

Our marker selection approach has been to develop technology which enables us to rapidly accumulate small insert clones from all classes of trinucleotide and tetranucleotide STRs. The basic strategy, termed marker selection, requires the construction of high complexity, small insert libraries essentially free of chimeras or clones without inserts.

This choice reflects the prior existence of large efforts to develop dinucleotide repeat markers, the general perception that these classes of markers result in more readable amplification products and the possibility that the availability of STRPs from multiple repeat classes will permit hybridization- based multiplex genotyping. In addition, with the increasing recognition that trinucleotide repeat expansion may be an important mechanism underlying human genetic disease, the availability of a large number of trinucleotide STRPs may provide an important resource for disease gene identification.

Other activities of Project 1 include studies devoted to increasing genotyping throughput as well as the development of efficient methods for recovery of STRs from large insert clones. Such methods will be essential for gap filling. As the project matures, the availability of a large set of STRPs will permit the investigation of the basis for repeat variability and explosion, help establish a set of cDNAs maintaining STRP sequences and further exploration of the role of repeat expansion in mutation. Investigators interested in additional information, detailed protocols, vectors or bacterial strains should contact:

Geoffrey M. Duyk, M.D., Ph.D.
Department of Genetics
East Quad Research Facility, Room 447
Harvard Medical School
200 Longwood Avenue
Boston, MA 02115
TEL: (617) 432-6072
FAX: (617) 432-7663
E-mail: Duyk@rascal.med.harvard.edu

PROJECT 2

Val C. Sheffield, M.D., Ph.D.

Project 2 of the Cooperative Human Linkage Center has as its primary goal the development of a minimum of 2,000 new, highly polymorphic (>0.70 heterozygosity) short tandem repeat polymorphisms (STRPs) with an emphasis on developing tri- and tetranucleotide repeat markers. The strategy for marker development consists of sequencing marker-selected clones obtained from Dr. Duyk's laboratory (Project 1), selecting PCR primers flanking the repeat and testing the PCR product for polymorphic information content. All markers are assigned to a specific chromosome using monochromosomal somatic cell hybrids, and all highly polymorphic markers are sent to Dr. Jeffrey Murray's (Project 3) and Dr. James Weber's (Project 4) laboratories for high resolution genetic mapping.

In the past few months, Project 2 has developed over 300 tetranucleotide markers. These markers are highly polymorphic, assayable using a standardized PCR condition, and have readily interpretable alleles. In addition to the goal of developing new STRPs, SSCP and DGGE are being used to identify polymorphisms in the 3' untranslated region of cDNA sequence. The identification of polymoprphisms in cDNA sequence allows placement of cDNAs on the genetic map.

Another goal of Project 2 is to develop a set of approximately 200-300 uniformly distributed STRPs which can be used for primary disease linkage studies. To this end, a primary linkage set of approximately 200 markers was developed, which were assayable using a single PCR condition. These markers, most of which are dinucleotide repeats, have proven extremely useful for disease linkage studies. For example, in collaboration with others, Project 2 has used the primary linkage set of markers to identify five hereditary eye disease loci. In order to improve the efficiency of primary linkage studies, the dinucleotide repeat markers are gradually being replaced with tetranucleotide repeat markers.

An underlying theme of the CHLC is the distribution of its resources to the user community. To this end, the CHLC will distribute STRP primers through Research Genetics and other interested companies. In addition, arrangements can be made for investigators working on disease families to bring their family resources to the University of Iowa to perform linkage studies on a collaborative basis.

Val C. Sheffield, M.D., Ph.D.
Assistant Professor of Pediatrics
University of Iowa
Iowa City, IA 52242
(319) 356-2674
FAX: (319) 356-3347
email: sheffield@vaxa.weeg.uiowa.edu

PROJECT 3

Jeffrey C. Murray, M.D.

The primary goal of Project 3 is to generate genotypes for the STRPs developed through Projects 1 and 2. These genotypes are then fed to Project 4 for incorporation into the developing linkage maps. Project 3 focuses around generating high quality, reliable genotypes using a variety of robotic assists, on a subset of the 60 CEPH families . Genotypes are currently generated by body-labelling PCR products using ^{35}S, and analysis of fragments on sequencing gels.

Genotypes are set up from formatted 96-well titre plates that include vacant wells at intervals to allow for controls and gel alignment. Multiplexing is currently done at the level of gel loading. Genotypes are scored and entered by hand in duplicate, with a subset of those generated also typed in duplicate through Project 4 to allow for data validity checks.

The project also has a limited ability to assist outside investigators in their own genotyping efforts. This would include hosting two-day to two-month visits for investigators who wish to carry out

genotyping on their own samples, genotyping of newly-generated anonymous markers or shotgun linkage searches in familial disorders.

Jeffrey C. Murray, M.D.
Cooperative Human Linkage Center
The University of Iowa, #440 EMRB
Iowa City, IA 52242
TEL: (319) 356-3508
FAX: (310 335-6970
E-mail: jeff-murray@umaxc.weeg.uiowa.edu

PROJECT 4

James L. Weber, Ph.D.

The major goals of Project 4 are to type newly developed STRPs through the CEPH families, to improve STRP genotyping technology, to collaboratively map disease genes, and to analyze several human meiotic parameters such as interference and sexual differences in recombination.

Typing of new STRPs will initially involve use of about 210 individuals from 14 of the largest CEPH families. Emphasis will be placed upon reduction of typing errors through the use of standard arrays of DNA templates within microtiter plates and 12 channel pipetting devices. Alleles will be assigned consistently among different families leading to useful estimates of allele frequencies.

Improving STRP genotyping technology will initially involve efforts to maximize the numbers of genotypes obtained per sequencing gel. Routinely three to six markers will be amplified simultaneously and electrophoresed together on 144 lane gels. In this way, up to 850 genotypes will be obtained per gel. Image analysis software specifically designed for STRPs will be used to speed the scoring of the markers and to avoid inconsistencies in allele assignment among families. Hardware and software for fluoresence-based sizing of alleles will gradually be developed to decrease the amount of labor required for genotyping.

Collaborative disease gene mapping efforts which have already resulted in the localization of a dozen genes will be continued through the CHLC. Visitors will come to Marshfield for periods of up to two months to engage in concentrated genotyping efforts. Because of limited amounts of available equipment, generally only one visitor will be accepted at one time. Visitors are responsible for all travel costs and living expenses in Marshfield, but all supplies will be provided by the CHLC. Interested individuals should contact Jim Weber at the address below. Groups working on disorders prevalent in minority groups or disorders that primarily affect women are especially encouraged to apply.

As many as 10^6 new genotypes will be determined by the CHLC over the next few years. These data represent an enormous new resource of human meiotic information. Distributions of crossovers along the chromosomes, crossover interference, sex-specificity in recombination rates, recombination hotspots, and relationships between genetic and physical distances are among the meiotic parameters that will be analyzed.

James L. Weber, Ph.D.
Senior Scientist, Human Genetics
Marshfield Medical Research Foundation
Marshfield, WI 54449
TEL: (715) 387-9179
FAX: (715) 389-3808
email: weberj@dgabby.mfldclin.edu

PROJECT 5

Kenneth H. Buetow, Ph.D.

It is the primary goal of this project to use the marker and genotype data generated in Projects 1-4 to construct a high integrity, fine structure, meiotic map of each human chromosome. Map construction will be conducted in a two-tiered manner. First, a high heterozygosity 10 cM resolution index map of PCR-detectable markers will be constructed. Next, likelihood and crossover minimization techniques will be used to integrate additional points to achieve a 2.5 cM resolution index map. These techniques will also be applied to obtain likely locations for previous RFLP typing from the CEPH panel and lower heterozygosity gene loci. It is recognized the map construction here will parallel efforts in progress in other gene mapping laboratories. The centralized effort conducted in this investigation will be complementary to these investigations.

As the first step toward accomplishing the above goals, a collection of maps have been generated that combine publicly available data with new genotype data generated by CHLC investigators. These maps integrate the genetic maps generated by the NCHGR Index Map Consortium and Genethon. They are augmented by data on additional markers provided by CHLC and CEPH investigators. The datasets are available through anonymous FTP (see below).

To generate the maps, the CHLC is using a new, semi-automated, map construction algorithm. The mapping algorithm is a stepwise construction procedure that utilizes the program CRIMAP as its analytic engine. The dataset is intially diagnosed for pairwise observations that show heterogeneity in pairwise recombination estimates by family. Such loci are excluded from primary construction. Loci are initially added to the map in order of information content. As each locus is added, support for the map and map expansion is re-evaluated. Loci that expand the map and/or are not supported by lod 3 criteria are removed. Loci demonstrating map expansion are moved to the end of the list for consideration in locus placement. The process is repeated until no loci can be added to the map at lod 3 support. The maps built by this alogrithm are somewhat more sparse than maps built by more traditional mapping algorithms (average marker density is 6.7 cM). However, they have very high confidence, and low error rates. These maps, called skeletal maps, and their corresponding error profiles are available through anonymous FTP.

The CHLC group has also generated a more highly annotated collection of maps. These maps were constructed using the STRP-based skeletal maps as starting points and expanded using the

CRIMAP-BUILD procedure with framework selection criteria for locus inclusion. These framework maps, their diagnostics, and likely locations for points that do not meet framework criteria, are also available through anonymous FTP. The sex-averaged version of these framework maps is included with this newsletter.

The map construction in this project will proceed simultaneously with development of statistical tools that allow the assessment of map quality and integrity. The primary focus of these efforts will be the development of statistical diagnostic methods for the evaluation of mapping outcomes. It is the goal of such diagnostics to identify error typings and biologically interesting observations.

Two concurrent approaches to the development of these tools will be taken. The first will use computational methods to assess the relative contributions to the final outcome of individual observations. These tests will be conducted at the level of individual typing, gamete, locus and family levels. As these methods are computer intensive, parallel/distributed algorithms for analysis/re-analysis of multipoint data are under development. In addition to these methods, explicit tests which are extensions of the statistical methods used in regression diagnostics will be explored.

Finally, means of applying goodness-of-fit tests will be evaluated. These will include the contrast of outcomes based on pairwise analysis (multiple pairwise likelihood analysis and seriation) as well as the use of empirical Bayes methods for assessing fit. The efficacy of using empirical Bayes methods to update linkage maps will also be examined.

Kenneth H. Buetow, Ph.D.
Fox Chase Cancer Center
Division of Population Science
7701 Burholme Ave.
Philadelphia, PA 19111
TEL: (215) 728-3152
FAX: (215) 728-3574
E-mail: kh_buetow@fccc.edu

INFORMATICS Core

Robert K. Stodola
Kenneth H. Buetow, Ph.D

The objective of the Informatics Core is provision of computer based tools that facilitate scientific aims of the Center. Its responsiblities include the storage, retrieval, and interpretation of the map reagents and data generated in the proposed research. The Informatics Core is charged with the management of Center-generated mapping reagents (sequence information, primers, genotypes, etc.), distribution and storage of protocols, and management and distribution of mapping outcomes (chromosome maps, meiotic breakpoint locations, etc.).

The primary purpose of this core is to generate and maintain a "production," database. This database will provide access to common resources and information within the Center. As CHLC efforts are proceeding at four geographically disparate locations (Harvard, University of Iowa, Marshfield, and Fox Chase) the current strategy is to build client-server based applications using the internet as a medium of communication between the four sites. Work is currently proceeding in the areas of database construction, distributed applications, and Graphical User Interface (GUI) tools.

The preliminary database has been constructed and a number of graphical interfaces to the database have been developed. We have selected Sybase as the database system and are currently using it with DECStation 5000 series computers. Several DEC AXP systems running OSF/1 have been purchased, and we plan to port the database when Sybase becomes available on this platform. To avoid dependence on Sybase, we have isolated the applications from the database with a database-independent interface, and used code generation techniques to reduce the complexity of building this interface library.

We have created several interesting distributed applications. One such is a distributed Primer PipeLine. Marker generation is currently underway at Harvard and the University of Iowa. Raw sequences are produced using ABI sequencers with Macintosh interfaces. The raw sequence files are copied directly onto a CHLC DECStation at these sites, and transferred to Fox Chase for processing. The PipeLine then assembles, strips cloning vector, identifies repeat regions, selects primers using PRIMER, verifies uniqueness, applies user selection criteria, and generates primer synthesis orders. At each stage data and user selections are stored in the production database for further information and use.

We are also developing a distributed linkage analysis program. Using the DCE (Distributed Computing Environment) component of OSF/1, we are partitioning the linkage analysis into a number of pieces which can be submitted to any available processor in the project. We anticipate making use of spare CPU cycles on all of the CHLC computer systems, including those at the remote sites by running linkage servers as a background process.

The CHLC Informatics Core is also responsible for the development and maintenance of a public access information system. This system will provide tools that facilitate the communication of the Center's mapping resourses to the outside genetics communities. Primary assistance in gaining access to information or services beyond those described here can be requested via electronic mail at help@chlc.org. It is anticipated that the CHLC public access database server will not become operational until Fall/Winter of 1993. In the interim, CHLC data will be available via anonymous FTP to ftp.chlc.org and through a CHLC Gopher Server addressed gopher.chlc.org. Described below is the information currently available.

README

A file describing the current contents. Each of the folders below also may include a README file describing the contents

chlc/newsletters

The CHLC newsletters in plain text and postscript

chlc/genotypes/tables
Tabular descriptions of marker systems in the chromosome specific datasets

chlc/genotypes/typing
Chromosome-specific genotype sets in CRIMAP file format chlc/maps/framework

chlc/maps/framework
Framework maps of all markers currently mapped by CHLC (including markers from other sources)

chlc/maps/skeletal
Maps generated using the stringent map build algorithm described above

Each maps folder contains three folders:

./diagnostics	Diagnostic data on maps.
./figures	Postscript figures.
./tables	Map information in text form.

chlc/markers/chlc
CHLC-produced marker data

chlc/markers/marshfield
Marshfield- produced marker data

A collection of public analytic services will also be supported by the Informatics Core. These services will be a subset of the analysis and evaluation tools used within the project which do not require exceptional computational resources. This will be provided free of charge and without any implied commitment to any level or service, accuracy or usefulness.

These servers will be provided via automated electronic mail servers, and we can take no responsibility for the privacy or confidentiality of these channels. The services provided will often include procedures developed by people outside the CHLC group. When these have not been placed in the public domain, we have asked permission to use these programs and procedures and kindly thank these indivduals and groups for their use. In all cases, each automated response will include attribution supplied by the author for his or her work. Instructions for each automated service can be found by sending any electronic mail message to the server address.

An information server has been placed in service that provides descriptive information about the CHLC project and data. It can be reached by sending e-mail to:

info-server@chlc.org

Mail to servers other than the info-server will reply with instructions on how to correctly structure messages to receive service and describe the services provided. It is anticipated that as of June 1, 1993 a server to perform linkage mapping will be in place. Initially, this server will take an individual marker system's genotype data and return markers from the CHLC data sets that show linkage. This information will include recombination fraction and lod scores. Later versions will provide map position information. To check the status of the linkage server send e-mail to:

linkage-server@chlc.org

Questions about CHLC services may be directed to ***help@chlc.org***. Since there are people on the other end of this address, please be patient. There aren't a lot of people on the other end, and all have lots to do!

In order to make it convenient to have CHLC announcements delivered via either USENET News or via electronic mail, and to avoid adding to the confusion of how to subscribe to yet another mail service, all CHLC postings will be presented via an appropriate BIOSCI newsgroup (currently, via BIOSCI/GENETIC-LINKAGE). If you have access to USENET news, this is the newsgroup:

bionet.molbio.gene-linkage

If you don't have access to USENET news or prefer to subscribe via electronical mail, the following instructions taken from Dave Kristofferson's "BIOSCI/bionet Frequently Asked Questions" posted to bionet.announce on May 1, 1993):

"For those who need e-mail subscriptions or who want to cancel current e-mail subscriptions, please send a request to one of the following addresses. Please choose the site that serves your location. Simply pick the newsgroup(s) from the list above that you wish to subscribe to and request that your address be added to the chosen mailing lists. Please use plain English; no special message syntax is required in your subscription or cancellation request.

Address	Serving
biosci@net.bio.net	The Americas and Pacific Rim
biosci@daresbury.ac.	Europe, Africa, and Central Asia

If you are changing e-mail addresses, please be sure to send a message to your request that your subscriptions be changed or canceled!!"

Dave also strongly recommends that all participants subscribe to the BIOSCI/ANNOUNCE group (USENET bionet.-announce).

Robert K. Stodola
Kenneth H. Buetow
Fox Chase Cancer Center
7701 Burholme Ave.
Philadelphia, PA 19111
TEL: (215) 728-3660
FAX: (215) 728-2513
E-mail: rk_stodola@fccc.edu

ELSI Core
Robert F. Weir, Ph.D.
James W. Hanson, M.D.

The ELSI (ethical, legal, and social implications) core is funded to carry out two projects: an IRB-type committee on genetics research and a postdoctoral fellowship program. The ELSI Committee Chair and Core Director, Dr. Robert Weir is the Director of the Program in Biomedical Ethics at the University of Iowa. Current committee members are listed below.

ELSI COMMITTEE MEMBERS

Robert Weir, Ph.D.	*ELSI Core Chair* *Biomedical Ethicist*
Jeff Murray, M.D.	*P.I., CHLC*
James Hanson, M.D.	*Medical Geneticist*
Kathy Mathews, M.D.	*Pediatric Neurologist* *Genetics Researcher*
Susan Johnson M.D.	*OB-Gynecologist* *U of I IRB Chair*
Laura Hart, R.N., Ph.D.	*College of Nursing* *IRB Member*
Stanely Grant, R.N.	*OB-GYN* *Prenatal Diagnosis*

Still to be added to the committee are a consumer of genetics services and a health-law attorney.

The ELSI committee has undertaken an analysis of the consent documents currently being used in genetics research. A written request for examples of these documents has been mailed to 150 genetics researchers nationwide, who were selected at random from the American Society of Human Genetics (ASHG) membership directory. Part of the committee's long-range plan is to develop one or more consent form models for genetics research that will prove helpful to both scientific investigators and to persons who participate as subjects in genetics-related research. The ELSI committee also plans to provide educational materials to be used by IRBs when they consider proposals for genetics research. We will coordinate our work with some of the work already done by the ASHG, the Alliance of Genetic Support Groups, and the Poynter Center at Indiana University.

The ELSI Core's postdoctoral fellowship program will be advertised nationally in the near future. This program will be directed at professionals outside the biological sciences who teach courses, give presentations, publish articles or books, or do other work pertaining to the ethical and legal issues of modern genetics. Such individuals would include persons in the fields of philosophy, history, law, journalism or religion. They will be at the University of Iowa for 2-4 months. During that time they will have a variety of work-related experiences in a molecular genetics lab, one or more other genetics labs, and several clinical genetics settings. On completion of this fellowship program, participants will have achieved a broader understanding of the challenges, technical vocabulary and problems regularly confronted by persons who work in molecular genetics and/or clinical genetics settings.

Robert F. Weir, Ph.D.
Professor of Pediatrics
University of Iowa
Iowa City, IA 52242
TEL: (319) 335-6705
FAX: (319) 335-8318

James W. Hanson, M.D.
Professor of Pediatrics
University of Iowa
Iowa City, IA 52242
(319) 356-2674
FAX: (319) 356-3347

ADMINISTRATIVE Core

Jeffrey C. Murray, M.D.

The Administrative Core serves as a focus for the overall center activities and also includes within it an educational component designed to establish outreach to the lay public.

Secondary School Educational Outreach

The Administrative Core is currently exploring several mechanisms to improve the knowledge base of secondary school students in relationship to the Human Genome Project. Funding is available for mini-sabbaticals by secondary school teachers to spend 1-2 months in the laboratory in a combination of didactic involvement related to human genetics and hands-on laboratory experience in genetic linkage analysis. In addition, collaborations are being developed with a number of external organizations, both in the development of textural materials related to teaching of secondary school students about the Human Genome Project in both its scientific and ethical implications, and also in direct outreach to such schools. The CHLC also participates in programs to have high school and undergraduate college students spend time in the laboratory, as well, again in a combination of didactic and hands-on laboratory experiences.

Jeffrey C. Murray, M.D.
Cooperative Human Linkage Center
The University of Iowa

If you would like to receive future issues of the CHLC Report in hard copy, please complete and send in the following form:

On completion, return to: **CHLC Administration, #440 EMRB, The University of Iowa, Iowa City, IA 52242**

Name

Institution

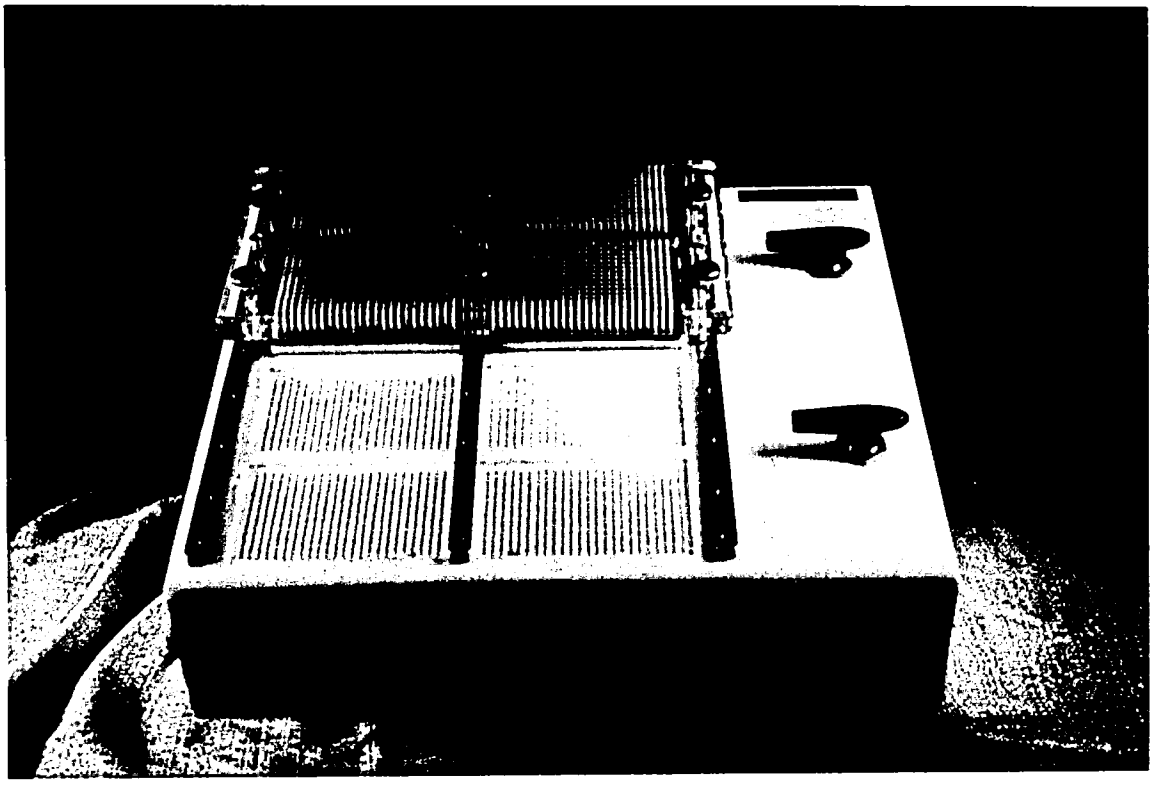
Department

Street/Building

City, State, Zip (COUNTRY)

VS3000™

High Throughput Vacuum Blotter



- ❑ Uses Eight Microplate-Sized Membranes with 384 Dots per Membrane, Total of 3072 Dots
- ❑ Uses 96-Tipped ***SPLATT™*** Pipetter for Fast, Accurate Blotting for pipetting 0.8 to 50 µliters Per Dot
- ❑ High Throughput Without Expensive Robotic Systems

SPLATT™ Pipetter Also Available from IAS Products, Inc.

IAS Products, Inc.

142 Rogers Street, Cambridge, MA 02142
Phone: 617-354-3830 Fax: 617-547-9727

Pls. file Lander

SPLATT™

SIMULTANEOUS PLATE LOADING AND TRANSFER TOOL

Patent Pending

SPLATT™ is a high precision 96-tip pipetter

SPLATT™ means high throughput without expensive robotic hardware

SPLATT™ pipettes to and from any combination of 96, 192, and 384-well microplates or a vacuum blotter

---

IAS Products, Inc.

142 Rogers Street, Cambridge, MA 02142

Phone: 617-354-3830 Fax: 617-547-9727

SPLATT™

SIMULTANEOUS PLATE LOADING AND TRANSFER TOOL

System Specifications

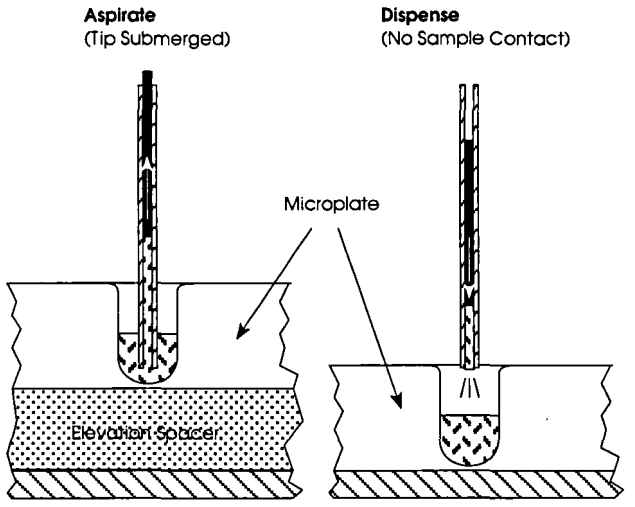
SPLATT™ Features

- ❑ 96 washable pipette tips - no expensive disposables
- ❑ Interchangeable, autoclavable tip cartridges
- ❑ Precision greater than 2% at all volumes
- ❑ Pipetting volumes from 0.8 to 50µl per tip

Example SPLATT™ Assisted Tasks

- Fill multiple plates from a master
- Copy a library of plates
- Transfer samples from four 96-well plates to one 384-well or 192-well plate
- Transfer oil to the top of all samples in a plate
- Add enzyme to hot plates in a thermal cycler
- Blot samples to a membrane using IAS Product's VC3000 Vacuum Blotting System

SPLATT™ Operation



The patent pending SPLATT™ design allows all 96 samples to be transferred without the need for physical tip touch-off to the plate. A high fluid velocity combined with an abrupt deceleration of the plungers within the tips "flings" even the smallest drops from the tips into the wells of the microplate below.

SPLATT™ Specifications

Main Unit (Without Cartridge)	
Air/Gas Supply	60-80 psi (4-5.5 Bar) (Optional External Pump Available)
Weight	2 lbs. (0.9 Kg)
Dimensions	7.5H X 6.5L X 3.6W inches (190H X 165L X 92W mm)

SPLATT™ Plate Tray Options

Model Number	Microplate Type	Elevation Spacer Model
PT-010	96 Well	ES-010
PT-020	192 Well	ES-020
PT-040	384 Well	ES-040

SPLATT™ Cartridge Options

Model Number	Volume Range	Tip Outer Dia. (mm)	Tip Len. (mm)
SC-020	0.8-2µl	0.6	22
SC-050	2-5µl	0.8	22
SC-100	4-10µl	1	22
SC-250	10-25µl	1.3	22
SC-500	20-50µl	1.8	22

IAS Products, Inc.

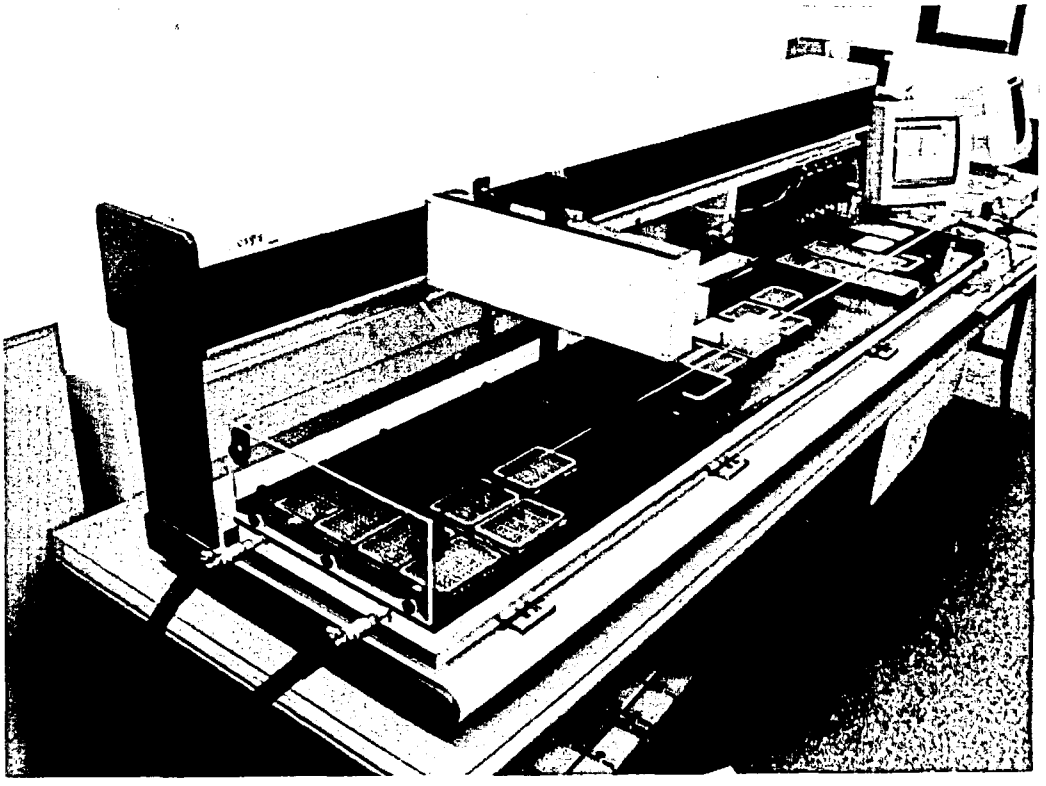
142 Rogers Street, Cambridge, MA 02142

Phone: 617-354-3830 Fax: 617-547-9727

Intelligent Automation Systems, Inc.

Developers of Automated Instrumentation for Genetic Research

- ❑ Automated process development
- ❑ Custom instrumentation design
- ❑ Laboratory robotic systems
- ❑ Machine vision system development

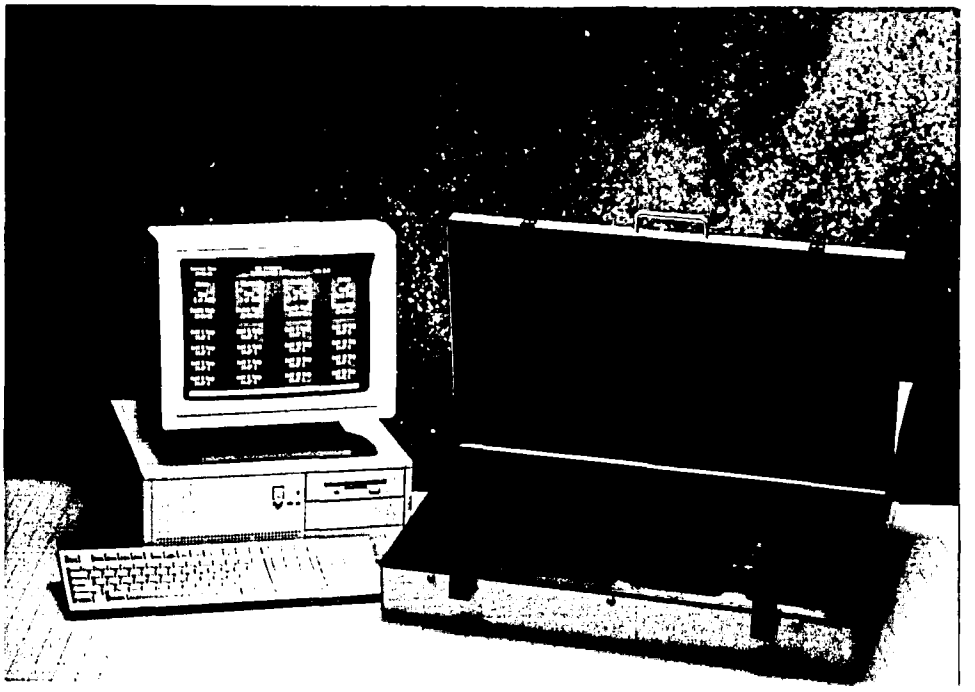


Intelligent Automation Systems, Inc. is a group of MIT-trained engineers committed to helping scientists and manufacturers develop advanced instrumentation.

142 Rogers Street, Cambridge, MA 02142 USA Phone: 617-354-3830 Fax: 617-547-9727

TC 1600 THERMOCYCLER*

High-speed thermal cycling for large numbers of biological and chemical samples



- Uses 16 standard PVC or polycarbonate microtitration plates (96 or 192 wells)
- Runs four quadrants (4 plates each) with different heating and cooling profiles
- Stores and recalls predefined profiles
- Allows up to ten set point temperatures from 30°C to 100°C with looping and pausing
- Records log file of actual temperature profiles
- Requires 208, 220 or 240 volts single phase at 70 amps - 60 or 50Hz and cold water hookup
- Specifications :

  Heating Rate: 1.3°C/sec Uniformity: +/- .5°C at 60°C
  Cooling Rate: 3.7°C/sec +/- 1°C at 95°C

* Patent Pending

IAS Products, Inc.

142 Rogers Street Cambridge, MA 02142 USA Phone: (617) 354-3830 Fax: (617) 547-9727

CH	LOCUS	ASSAY	HET	PIC	SIZE RANGE MIN	MAX	REFERENCE
01	ACTN2	CA4	0.50	0.47	91	107	GDB
01	AMY2B	NA	0.70	0.63	76	86	DRACOPOLI,N.& MEISLER,M.(1990) GENOMICS7,97-102.
01	ANGIOTEN	NA	0.75		113	133	KOTELEVTSEV,Y.V. ET AL.(1991) NAR19,6978.
01	APOA2	MFD 3	0.74	0.65	131	145	GENOMICS 15:251-258, 1993
01	AT3	NA	0.78		277	316	PERRY,D.J. (1993) HMG 2,618.
01	ATP1A2	NA	0.72		397	405	LIM,L.& GILL,M.(1993) HMG 2,616.
01	C4BPA/B	NA	0.74		130	142	VELASCO.E. ET AL.(1992) HMG 1,552.
01	CLN1	HY-TM1	0.87		140	209	GDB
01	CRP	MFD 57	0.60	0.53	127	145	GLATT,K. ET AL.(1992) HMG 1,348.
01	CRTM	NA	0.67		102	110	GDB
01	D1S102	MFD 52	0.63	0.50	186	204	GLATT,K. ET AL.(1992) HMG 1,348.
01	D1S103	MFD 64	0.88	0.78	82	102	GLATT,K. ET AL.(1992) HMG 1,348.
01	D1S104	MFD 67	0.76	0.66	152	168	GLATT,K. ET AL.(1992) HMG 1,348.
01	D1S116	NA	0.00	0.62	89	101	SHARMA,V. ET AL.(1991) NAR 19,1169.
01	D1S1167	NIB1152	0.86		163	187	GDB
01	D1S117	NA	0.77	0.77	100	132	SHARMA V.& LITT M. (1991) NAR19,1168.
01	D1S158	NA	0.89	0.88	137	163	OVERBECK,LD.ET AL(1992)HMG 1,141.
01	D1S1586	ACT1B03	0.56		91	118	GDB
01	D1S1587	ATA1D01	0.44		144	168	GDB
01	D1S1588	ATA2E04	0.63		118	139	GDB
01	D1S1589	ATA4E02	0.73		199	220	GDB
01	D1S159	MIT-MX4	0.67		147	147	HUDSON,T. ET AL.(1992) GENOMICS 13,622-29.
01	D1S1590	ATA5E03	0.67		148	169	GDB
01	D1S1591	GAAT2B03	0.73		86	86	GDB
01	D1S1592	GAAT4D10	0.67		232	244	GDB
01	D1S1593	GATA13G07	0.87		209	209	GDB
01	D1S1594	GATA22D12	0.60		104	128	GDB
01	N/A	GATA25	0.57		N/A	N/A	GDB
01	D1S1595	GATA25B02	0.77		265	297	GDB
01	D1S1596	GATA26G09	0.73		105	125	GDB
01	D1S1597	GATA27E01	0.70		155	179	GDB
01	D1S1598	GATA27F07	0.56		111	139	GDB
01	D1S1599	GATA31H02	0.79		230	230	GDB
01	D1S160	MIT-MS48	0.72		150	150	HUMAN GENET 87:401, 1991
01	D1S1600	GATA3B11	0.81		148	164	GDB
01	D1S1601	GATA3D01	0.75		220	220	GDB
01	D1S1602	GATA42A04	0.73		295	295	GDB
01	D1S1603	GATA42F05	0.77		196	196	GDB
01	D1S1604	GATA43D10	0.82		203	203	GDB
01	D1S1605	GATA45A06	0.67		390	390	GDB
01	D1S1606	GATA46C02	0.92		287	287	GDB
01	D1S1607	GATA48C11	0.69		269	269	GDB
01	D1S1608	GATA49A06	0.83		269	269	GDB
01	D1S1609	GATA50F11	0.91		196	196	GDB
01	D1S161	MIT-E112	0.84		159	159	HUMAN GENET 87:401, 1991
01	D1S1610	GATA50H07	0.50		169	169	GDB
01	D1S1611	GGAA10G11	0.81		217	217	GDB
01	D1S1612	GGAA3A07	0.50		121	121	GDB
01	D1S1613	GGAA7C04	0.80		304	304	GDB
01	D1S1614	GGAA8F12	0.56		210	246	GDB
01	D1S1615	GGAT3G04	0.63		265	265	GDB
01	D1S1616	GGAT4C11	0.44		137	145	GDB
01	D1S162	MIT-MS154	0.91		134	134	HUMAN GENET 87:401, 1991
01	D1S163	MIT-MS217	0.68		200	200	HUMAN GENET 87:401, 1991
01	D1S164	MIT-MS165	0.83		229	229	HUMAN GENET 87:401, 1991
01	D1S165	MIT-A115	0.71		156	177	GENOMICS 8:400- , 1990
01	D1S167	NA	0.74		159	183	BOWCOCK,A. ET AL.(1992) HMG 1,138.
01	D1S170	COS370	0.79		217	217	GDB
01	D1S175	MFD 96	0.80		145	165	GENOMICS 8:400- , 1990
01	D1S176	MFD197			107	107	GLATT,K. ET AL.(1992) HMG 1,348.
01	D1S177	MFD160	0.55		92	92	GENOMICS 8:400- , 1990
01	D1S178	MFD 89	0.50		142	148	GLATT,K. ET AL.(1992) HMG 1,348.
01	D1S179	MFD174	0.70		163	193	GLATT,K. ET AL.(1992) HMG 1,348.
01	D1S180	MFD126	0.90		163	189	GLATT,K. ET AL.(1992) HMG 1,348.
01	D1S184	MFD211	0.53	0.47	77	85	GLATT,K. ET AL.(1992) HMG 1,348.
01	D1S185	MFD215	0.50	0.47	114	136	GLATT,K. ET AL.(1992) HMG 1,348.
01	D1S186	MFD217	0.84	0.82	82	106	GLATT,K. ET AL.(1992) HMG 1,348.
01	D1S187	MFD227	0.71	0.68	83	103	GLATT,K. ET AL.(1992) HMG 1,348.
01	D1S188	MFD246	0.86	0.85	149	173	GLATT,K. ET AL.(1992) HMG 1,348.
01	D1S189	AFM036xc5	0.78		124	136	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801

CH	LOCUS	ASSAY	HET	PIC	SIZE RANGE MIN	MAX	REFERENCE
01	D1S190	AFM046xc11	0.94		293	331	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S191	AFM046xh10	0.75		153	169	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S192	AFM051xh8	0.67		203	211	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S193	AFM057xf4	0.78		94	106	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S194	AFM057xf8	0.67		233	239	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S195	AFM063xb6	0.31		183	189	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S196	AFM063xg9	0.74		267	279	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S198	AFM074za5	0.81		308	322	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S199	AFM078yg5	0.84		94	116	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S200	AFM093xf3	0.80		154	178	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S201	AFM094tb7	0.73		186	204	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S202	AFM095ta5	0.77		77	91	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S203	AFM095wc9	0.64		123	129	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S204	AFM102xe3	0.46		248	252	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S206	AFM113xf6	0.82		206	218	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S207	AFM116xb2	0.85		142	170	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S208	AFM120xa5	0.78		134	152	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S209	AFM120xd4	0.81		69	169	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S210	AFM122xa3	0.64		117	125	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S211	AFM122xe1	0.86		172	198	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S212	AFM142xc9	0.80		105	125	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S213	AFM147xf8	0.87		104	124	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S214	AFM147yf8	0.79		120	142	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S215	AFM150xh4	0.73		189	207	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S216	AFM154xc7	0.90		228	260	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S217	AFM156xg7	0.66		130	142	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S218	AFM157xe7	0.84		266	286	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S219	AFM161xb2	0.83		154	176	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S220	AFM162xg3	0.83		231	251	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S221	AFM164xe1	0.63		215	225	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S222	AFM164yg1	0.72		258	276	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S223	AFM168yb2	0.77		252	264	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S224	AFM179yg3	0.66		120	130	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S225	AFM184xa9	0.80		111	131	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S226	AFM184xe11	0.84		90	106	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S227	AFM184yf6	0.68		111	125	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S228	AFM196xb4	0.78		117	129	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S229	AFM196xh4	0.78		191	207	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S230	AFM197xb6	0.79		177	189	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S231	AFM198wa3	0.85		158	168	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S232	AFM198ye9	0.53		184	202	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S233	AFM199zd2	0.85		102	132	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S234	AFM200yf12	0.83		226	238	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S235	AFM203yg9	0.69		175	195	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S236	AFM205ta11	0.80		190	218	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S237	AFM205xd8	0.77		172	192	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S238	AFM205xg1	0.87		272	302	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S239	AFM205yg3	0.70		242	258	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S240	AFM207vh8	0.63		236	242	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S241	AFM211xa1	0.52		218	226	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S242	AFM212xb10	0.85		213	227	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S243	AFM214yg7	0.87		142	170	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S244	AFM220yf4	0.82		285	296	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S245	AFM224xc1	0.83		235	253	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S246	AFM225zg7	0.72		200	224	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S247	AFM234tb6	0.87		243	263	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S248	AFM234vb4	0.82		191	211	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S249	AFM234wf6	0.88		155	185	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S250	AFM240yg1	0.78		133	147	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S251	AFM248ya5	0.83		249	271	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S252	AFM249zg9	0.83		99	119	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S253	AFM254wb9	0.48		164	172	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S254	AFM260xf1	0.66		198	208	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S255	AFM260zg5	0.77		74	88	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S303	AFM081zc5	0.53		181	191	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S304	AFM116xf8	0.61		168	174	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S305	AFM220xf8	0.83		156	176	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S306	AFM238xd10	0.62		261	281	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
01	D1S318	MFD147	0.74	0.71	154	180	GLATT,K. ET AL.(1992) HMG 1,348.
01	D1S319	MFD252	0.70	0.65	151	161	GLATT,K. ET AL.(1992) HMG 1,348.

CH	LOCUS	ASSAY	HET	PIC	SIZE RANGE MIN	MAX	REFERENCE
01	D1S320	MFD255	0.70	0.65	191	205	GLATT,K. ET AL.(1992) HMG 1,348.
01	D1S322	MFD 60	0.70		90	110	GENOMICS 8:400- , 1990
01	D1S333	MFD275	0.84	0.82	155	199	GENOMICS 8:400- , 1990
01	D1S334	MFD293	0.55	0.49	305	333	GENOMICS 8:400- , 1990
01	D1S399	MFD186	0.60		110	126	GENOMICS 8:400- , 1990
01	D1S412	AFM031xd12	0.70		185	207	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S413	AFM165xc9	0.79		246	262	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S414	AFM179xg5	0.62		185	205	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S415	AFM183yg7	0.42		197	201	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S416	AFM185xd6	0.68		146	162	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S417	AFM192xg3	0.73		189	195	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S418	AFM197yg1	0.80		176	190	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S419	AFM199wh2	0.73		162	194	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S420	AFM199xb6	0.68		199	213	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S421	AFM199zb12	0.70		146	152	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S422	AFM200ve3	0.23		157	167	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S423	AFM042xe3	0.56		162	168	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S424	AFM203vd4	0.75		197	229	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S425	AFM203zb6	0.66		92	108	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S426	AFM205xh2	0.61		132	158	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S427	AFM205zc11	0.71		275	297	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S428	AFM206xd6	0.83		181	195	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S430	AFM210vc9	0.67		167	189	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S431	AFM212xf6	0.60		207	221	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S433	AFM217xe11	0.77		140	150	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S434	AFM217yh8	0.73		240	252	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S435	AFM217zb2	0.75		157	177	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S436	AFM217zc3	0.36		200	240	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S437	AFM220xe1	0.65		133	165	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S438	AFM220yc7	0.66		156	174	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S439	AFM225xe11	0.53		243	265	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S440	AFM234wc9	0.65		135	141	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S441	AFM234xb2	0.71		164	178	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S442	AFM234zd12	0.62		230	236	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S443	AFM238vg3	0.63		310	328	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S444	AFM078xh3	0.71		172	180	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S445	AFM238xf10	0.52		146	156	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S446	AFM238ya7	0.76		89	132	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S447	AFM024xd2	0.72		123	141	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S448	AFM240wb4	0.42		209	215	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S449	AFM240za9	0.54		221	245	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S450	AFM247te9	0.71		243	267	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S451	AFM248tf9	0.85		174	188	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S452	AFM248wg5	0.63		217	227	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S453	AFM248zh9	0.76		217	227	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S454	AFM080xd3	0.73		155	163	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S456	AFM024xf8	0.76		197	211	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S457	AFM084xb3	0.68		199	214	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S458	AFM113xc7	0.75		154	164	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S459	AFM119yh10	0.66		177	193	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S460	AFM123yc5	0.67		145	159	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S461	AFM123yc7	0.43		234	246	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S462	AFM126xa1	0.79		238	244	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S463	AFM154xg11	0.63		223	233	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S464	AFM270vc9	0.71		101	121	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S465	AFM270zf5	0.84		207	227	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S466	AFM275xe1	0.65		155	175	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S467	AFM277yf9	0.76		167	177	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S468	AFM280we5	0.54		173	191	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S469	AFM280za5	0.88		289	333	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S470	AFM281xg9	0.58		155	165	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S471	AFM281yg1	0.64		187	197	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S472	AFM282zd1	0.66		228	244	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S473	AFM286xd9	0.75		235	259	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S474	AFM286yc1	0.62		106	114	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S476	AFM289yc1	0.74		159	171	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S477	AFM289ye9	0.69		216	230	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S478	AFM290vb9	0.49		155	175	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S479	AFM290wd1	0.62		102	126	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S480	AFM290yf9	0.77		198	206	WEISSENBACH J: NATURE GENETIC, JUNE 1994

CH	LOCUS	ASSAY	HET	PIC	SIZE RANGE MIN	SIZE RANGE MAX	REFERENCE
01	D1S481	AFM294wg1	0.58		235	255	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S482	AFM294zd1	0.69		177	189	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S483	AFM296zc9	0.71		222	228	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S484	AFM297wb9	0.47		136	142	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S485	AFM297xf9	0.82		157	181	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S486	AFM297zg1	0.71		214	222	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S487	AFM298vc5	0.81		260	266	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S488	AFM299ze9	0.52		181	205	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S489	AFM309ve9	0.76		141	153	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S490	AFM309yd1	0.71		195	217	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S491	AFM310vb1	0.85		171	183	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S492	AFM310xh9	0.76		94	98	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S493	AFM311ve1	0.61		132	144	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S494	AFM319zh9	0.69		177	187	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S495	AFM323ya5	0.96		138	164	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S496	AFM329xd5	0.77		213	239	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S497	AFM331vb1	0.71		250	276	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S498	AFM336xb1	0.74		183	205	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S499	AFM338wb5	0.49		194	206	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S500	AFM343vf9	0.77		167	181	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S501	AFM359tb5	0.83		172	178	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S502	AFM361td9	0.66		252	286	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S503	AFMa123ya9	0.74		203	213	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S504	AFMa123yf1	0.73		124	138	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S505	AFMa127wb5	0.58		145	167	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S506	AFMa127wh9	0.72		123	141	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S507	AFMa127zc9	0.81		183	203	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S508	AFMa128ye9	0.56		73	85	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S509	AFMa132wa9	0.47		69	77	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S510	AFMa132yc9	0.79		173	195	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S511	AFMa133xc5	0.81		218	230	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S512	AFMa134vb1	0.60		102	124	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S513	AFMa134xf9	0.84		179	197	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S514	AFMa151za5	0.74		145	157	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S515	AFMa152yg9	0.63		198	222	WEISSENBACH J: NATURE GENETIC, JUNE 1994
01	D1S518	GATA7C01	0.88		191	223	GDB
01	D1S519	GATA2B02	0.71		123	453	GDB
01	D1S520	GATA5A06	0.57		172	172	GDB
01	D1S526	wg2c9	0.83		187	247	ARMOUR, J. ET AL. (1994) HMG 3, 599-605.
01	D1S53	pL673	0.91		171	201	NISHIMURA,D. ET AL.(1992) NAR 20,1167.
01	D1S532	GAAT1D9	0.50		117	133	GDB
01	D1S533	GATA10C02	0.86		193	225	GDB
01	D1S534	GATA12A07	0.94		205	207	GDB
01	D1S535	GATA2H05	0.56		174	174	GDB
01	D1S540	GATA4H05	0.17		140	156	GDB
01	D1S547	GATA4A09	0.68		282	308	GDB
01	D1S548	GATA4H04	0.76		148	172	GDB
01	D1S549	GATA4H09	0.82		157	193	GDB
01	D1S550	GATA5G07	0.82		169	189	GDB
01	D1S551	GATA6A05	0.73		166	186	GDB
01	D1S552	GGAT2A07	0.81		244	256	GDB
01	D1S556	NA	0.83		109	139	WESTON, M. ET AL. (1994) HMG 3, 1211.
01	F13B	NA	0.65		169	185	NISHIMURA,D.& MURRAY,J.(1992) NAR 20,1167.
01	FGR	NA		0.63	135	143	PATEL,M. ET AL.(1992)HMG 1,65.
01	FLER16	NA	0.70		143	165	BRINI,A. ET AL.(1993) HMG 2,619.
01	HLX1	NA	0.82		150	172	GDB
01	HSD3B2	PCR1	0.58		281	372	GDB
01	LAMB2	NA	0.89		134	150	GDB
01	MYCL1	HYTM1	0.87	0.87	140	209	GDB
01	PKLR	NA	0.72		316	331	GDB
01	SPTA1	NA	0.81	0.77	128	146	GDB
02	APOB	GZ9/10	0.66		143	152	GDB
02	CD8A	NA	0.71	0.66	138	170	POLYMEROPOULOS,M. ET AL.(1991) NAR 19,1718.
02	CRYG1B	NA	0.69	0.61	117	126	POLYMEROPOULOS,M. ET AL.(1991) NAR 19,4571.
02	CTLA4	NA	0.92	0.91	92	130	POLYMEROPOULOS,M. ET AL.(1991) NAR 19,4307.
02	D2S100	MIT-MH105	0.52		143	143	HUMAN GENET 87:401, 1991
02	D2S101	MFD115	0.58	0.59	86	100	JONES,M. ET AL.(1992) HMG 1,131-33.
02	D2S102	MFD128	0.85	0.86	138	162	JONES,M. ET AL.(1992) HMG 1,131-33.
02	D2S103	MFD145	0.82	0.77	109	125	JONES,M. ET AL.(1992) HMG 1,131-33.
02	D2S104	MFD149	0.72	0.57	114	132	HAUGE,X. ET AL.(1991) NAR 19,4308.

CH	LOCUS	ASSAY	HET	PIC	SIZE RANGE MIN	MAX	REFERENCE
02	D2S105	MFD156	0.69	0.63	107	125	JONES,M. ET AL.(1992) HMG 1,131-33.
02	D2S106	MFD 70	0.70		102	102	GENOMICS 8:400- , 1990
02	D2S107	MFD 82	0.30		85	85	GENOMICS 8:400- , 1990
02	D2S108	2CA12AC	0.76		388	400	GDB
02	D2S110	AFM016yc5	0.73		153	167	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S111	AFM021tc1	0.81		126	140	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S112	AFM032xg5	0.72		136	150	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S113	AFM052xb4	0.78		206	230	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S114	AFM052xf8	0.87		214	236	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S115	AFM058ye3	0.72		106	126	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S116	AFM064xh7	0.79		134	150	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S117	AFM065yf11	0.83		186	212	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S118	AFM066xc1	0.79		169	191	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S119	AFM077yb7	0.81		214	232	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S120	AFM077yc5	0.77		187	211	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S121	AFM087xa1	0.83		156	184	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S122	AFM087xg9	0.91		92	110	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S123	AFM093xh3	0.77		197	227	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S124	AFM094zc9	0.69		157	163	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S125	AFM112yd4	0.83		88	100	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S126	AFM119xc7	0.82		141	161	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S127	AFM119xd2	0.86		283	303	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S128	AFM143xd12	0.80		146	164	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S129	AFM143xe11	0.79		162	180	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S130	AFM150xf2	0.76		196	214	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S131	AFM155yd2	0.86		229	247	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S132	AFM157xe3	0.79		189	213	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S1321	ATA1B01	0.57		225	225	GDB
02	D2S1322	ATA3G09	0.67		126	141	GDB
02	D2S1323	GAAT1B02	0.57		324	328	GDB
02	D2S1324	GAAT1F2	0.69		202	210	GDB
02	D2S1325	GATA22E06	0.63		107	131	GDB
02	D2S1326	GATA26B04	0.83		232	268	GDB
02	D2S1327	GATA26D05	0.83		143	171	GDB
02	D2S1328	GATA27A12	0.83		142	166	GDB
02	D2S1329	GATA28H06	0.63		191	207	GDB
02	D2S133	AFM165zh8	0.69		283	301	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S1330	GATA32E03	0.57		262	262	GDB
02	D2S1331	GATA41E01	0.92		302	302	GDB
02	D2S1332	GATA42B12	0.36		126	126	GDB
02	D2S1333	GATA49F04	0.62		301	301	GDB
02	D2S1334	GATA4D07	0.71		266	310	GDB
02	D2S1335	GATA50E04	0.60		148	148	GDB
02	D2S1336	GATA8G06	0.68		241	261	GDB
02	D2S1337	GATA9A11	0.66		143	143	GDB
02	D2S1338	GGAA3A09	0.92		168	168	GDB
02	D2S1339	GGAA9B02	0.74		263	263	GDB
02	D2S134	AFM168xg11	0.78		196	216	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S135	AFM172xc3	0.62		129	135	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S136	AFM172xe7	0.75		103	990	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S137	AFM172xg3	0.69		140	158	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S138	AFM176xd4	0.68		111	125	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S139	AFM177xh4	0.82		175	197	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S140	AFM182ya5	0.78		151	167	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S1400	GGAA20G10	0.50		188	208	GDB
02	D2S141	AFM183xc3	0.88		152	178	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S142	AFM191wg9	0.77		254	266	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S143	AFM191xb8	0.84		109	132	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S144	AFM193yb4	0.85		152	200	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S145	AFM198ta1	0.60		248	275	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S146	AFM198wc5	0.58		188	198	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S147	AFM199vb6	0.73		126	144	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S148	AFM200wa11	0.63		180	200	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S149	AFM200yg11	0.74		210	228	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S150	AFM205tg1	0.84		218	242	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S151	AFM207vf8	0.83		211	229	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S152	AFM207xg1	0.77		269	285	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S153	AFM207ye11	0.81		253	361	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S154	AFM210yc11	0.65		314	326	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S155	AFM210yf10	0.82		163	171	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801

CH	LOCUS	ASSAY	HET	PIC	SIZE ANGE MIN	MAX	REFERENCE
02	D2S156	AFM211yd6	0.86		168	198	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S157	AFM212ze9	0.76		104	122	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S158	AFM217xh8	0.67		266	278	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S159	AFM218zg3	0.77		165	177	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S160	AFM220ze3	0.79		204	218	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S161	AFM224zf4	0.61		220	232	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S162	AFM225zg5	0.75		120	144	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S163	AFM234wa9	0.80		213	231	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S164	AFM234xb8	0.84		265	303	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S165	AFM234ya9	0.86		81	111	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S166	AFM234zh2	0.58		236	246	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S168	AFM240vf6	0.83		196	216	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S169	AFM240vh12	0.57		194	202	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S170	AFM240yf8	0.89		203	225	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S171	AFM242yd8	0.87		253	281	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S172	AFM248wc5	0.94		258	296	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S173	AFM249wg9	0.70		117	125	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S174	AFM254vc9	0.80		203	221	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S175	AFM260xc5	0.60		133	145	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S176	AFM262xb5	0.70		240	250	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S177	AFM267zc9	0.85		276	302	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S205	AFM240yc3	0.56		148	166	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S206	AFM259yc9	0.81		123	151	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S207	AFM263zh9	0.72		144	156	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
02	D2S211	NA	0.83		126	140	BARBER,T. ET AL.(1993) HMG 2,88.
02	D2S218	MFD270	0.53	0.43	267	277	GENOMICS 8:400- , 1990
02	D2S219	MFD291	0.57	0.50	146	154	GENOMICS 8:400- , 1990
02	D2S220	MFD292	0.79	0.76	157	177	GENOMICS 8:400- , 1990
02	D2S221	MFD294	0.58	0.51	190	200	GENOMICS 8:400- , 1990
02	D2S222	MFD301	0.77	0.75	118	138	GENOMICS 8:400- , 1990
02	D2S223	MFD307	0.74	0.71	233	253	GENOMICS 8:400- , 1990
02	D2S224	MFD266	0.64	0.58	165	185	GENOMICS 8:400- , 1990
02	D2S230	NA	0.83		146	160	BYERLEY,W. ET AL.(1993) HMG 2,1329.
02	D2S280	AFM155ye1	0.73		133	143	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S281	AFM168xg3	0.72		243	255	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S282	AFM184yc1	0.59		200	212	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S283	AFM191vg11	0.74		244	294	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S284	AFM192yd2	0.80		228	244	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S285	AFM038xh8	0.60		189	211	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S286	AFM200xb2	0.62		134	150	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S287	AFM200zc11	0.53		353	367	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S288	AFM200ze3	0.65		276	284	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S289	AFM044xa1	0.60		183	195	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S290	AFM203wb6	0.77		197	223	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S291	AFM203xc3	0.83		180	202	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S292	AFM203yb6	0.53		180	192	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S293	AFM205xa1	0.62		165	191	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S294	AFM205xf12	0.69		184	216	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S295	AFM205yb4	0.73		203	213	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S296	AFM205yg7	0.68		156	170	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S297	AFM207vc7	0.63		98	112	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S298	AFM210xb2	0.57		133	147	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S299	AFM212zh10	0.60		196	204	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S300	AFM214xc3	0.71		86	90	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S301	AFM214ye1	0.69		224	240	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S302	AFM218xa1	0.72		96	102	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S303	AFM220xa11	0.70		132	144	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S304	AFM224xb10	0.72		177	189	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S305	AFM073ya5	0.73		269	283	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S306	AFM225zd8	0.69		219	243	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S307	AFM074xg9	0.52		205	221	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S308	AFM234te1	0.66		228	234	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S309	AFM234vg5	0.64		176	204	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S310	AFM234xe3	0.48		139	145	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S311	AFM234zh6	0.65		185	207	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S312	AFM242xf10	0.84		234	240	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S313	AFM079za5	0.82		149	155	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S314	AFM248za9	0.56		255	271	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S315	AFM081yg5	0.74		114	130	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S316	AFM092yd11	0.56		140	154	WEISSENBACH J: NATURE GENETIC, JUNE 1994

CH	LOCUS	ASSAY	HET	PIC	SIZE ANGE MIN	MAX	REFERENCE
02	D2S317	AFM094xf11	0.53		250	256	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S318	AFM105xc1	0.80		212	221	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S319	AFM108xh8	0.77		124	134	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S320	AFM137xg11	0.80		175	193	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S321	AFM144yf8	0.73		140	156	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S322	AFM262xe5	0.53		191	209	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S323	AFM263wb5	0.45		177	193	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S324	AFM263xe1	0.68		265	275	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S325	AFM266vc5	0.86		190	214	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S326	AFM266ve1	0.43		156	174	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S327	AFM267ve9	0.76		123	133	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S328	AFM268va5	0.70		209	217	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S329	AFM268xg5	0.65		232	250	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S330	AFM269xd9	0.60		227	237	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S331	AFM269yd9	0.69		252	266	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S332	AFM270xh9	0.51		148	162	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S333	AFM270ze9	0.65		184	202	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S334	AFM273va9	0.57		119	127	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S335	AFM275yd5	0.81		98	114	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S336	AFM275yf5	0.46		107	139	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S337	AFM275za9	0.71		233	255	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S338	AFM276zf5	0.78		273	291	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S339	AFM277vb9	0.66		123	133	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S340	AFM277wc9	0.73		149	175	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S342	AFM280wd5	0.76		231	247	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S343	AFM281yd5	0.76		187	201	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S344	AFM284vd9	0.43		296	302	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S345	AFM288vb1	0.76		249	259	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S346	AFM289vf5	0.76		141	159	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S347	AFM289xb1	0.58		264	292	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S348	AFM289xd9	0.60		249	255	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S349	AFM290ye9	0.67		129	135	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S350	AFM292wd1	0.78		152	180	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S351	AFM294yf5	0.73		166	172	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S352	AFM296vg9	0.54		186	206	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S353	AFM296vh9	0.82		287	301	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S354	AFM296xa5	0.74		175	183	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S355	AFM296xb9	0.52		102	112	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S356	AFM297wc1	0.79		184	190	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S357	AFM297we1	0.60		198	216	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S358	AFM297xh5	0.54		147	175	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S359	AFM298xb9	0.84		214	228	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S360	AFM301wg1	0.62		106	126	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S361	AFM301za5	0.68		240	250	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S362	AFM302vh9	0.79		102	116	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S363	AFM303wc5	0.59		252	264	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S364	AFM303ya9	0.86		226	248	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S365	AFM303yc1	0.64		164	206	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S367	AFM303ze1	0.88		137	165	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S368	AFM304ta9	0.81		218	244	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S369	AFM304tb5	0.63		153	167	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S370	AFM310xf5	0.53		214	226	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S371	AFM311vg9	0.71		148	156	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S372	AFM312vf1	0.61		296	308	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S373	AFM316tg5	0.41		218	238	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S374	AFM318wf1	0.64		114	136	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S375	AFM318za9	0.86		173	183	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S376	AFM319xg1	0.80		266	270	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S377	AFM319zf9	0.42		158	164	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S378	AFM320yb9	0.65		203	217	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S379	AFM320yd9	0.79		178	188	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S380	AFM321xd9	0.73		229	253	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S381	AFM321yg5	0.82		298	312	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S382	AFM321zf9	0.69		156	170	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S383	AFM323wc5	0.76		172	186	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S384	AFM323zd5	0.47		241	253	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S385	AFM326yf9	0.17		162	168	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S386	AFM326zh9	0.84		210	216	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S387	AFM331zg5	0.78		191	221	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S388	AFM333vh5	0.38		254	266	WEISSENBACH J: NATURE GENETIC, JUNE 1994

CH	LOCUS	ASSAY	HET	PIC	SIZE ANGE MIN	MAX	REFERENCE
02	D2S389	AFM333wf9	0.73		189	219	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S390	AFM337yf9	0.79		179	193	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S391	AFM337yh5	0.62		142	152	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S392	AFM347ya5	0.76		218	224	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S393	AFM348tf1	0.56		84	103	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S394	AFM350td1	0.57		119	141	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S395	AFM356te5	0.70		144	166	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S396	AFM361ta5	0.89		230	244	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S397	AFM362td9	0.46		198	210	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S398	AFMa127xb9	0.50		113	139	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S399	AFMa131wb9	0.86		205	225	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S400	AFMa132zc9	0.73		186	194	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S401	AFMa140yg9	0.71		105	121	WEISSENBACH J: NATURE GENETIC, JUNE 1994
02	D2S402	AD17A	0.75	0.72	131	147	GDB
02	D2S405	GATA8F07	0.64		240	256	GDB
02	D2S407	GATA5H02	0.73		194	194	GDB
02	D2S408	GATA2E04	0.64		287	291	GDB
02	D2S410	GATA4E11	0.81		163	163	GDB
02	D2S414	MFD325	0.77	0.73	206	222	GENOMICS 8:400- , 1990
02	D2S415	MFD328	0.21	0.21	198	204	GENOMICS 8:400- , 1990
02	D2S416	MFD330	0.59	0.55	135	143	GENOMICS 8:400- , 1990
02	D2S417	MFD337	0.76	0.73	191	211	GENOMICS 8:400- , 1990
02	D2S418	MFD350	0.52	0.48	216	226	GENOMICS 8:400- , 1990
02	D2S422	ATC3E01	0.44		145	154	GDB
02	D2S423	GAAT1A5	0.62		111	131	GDB
02	D2S424	GAAT1C10	0.62		158	194	GDB
02	D2S425	GATA11H04	0.56		294	306	GDB
02	D2S426	GATA12B05	0.50		152	164	GDB
02	D2S427	GATA12H10	0.56		244	254	GDB
02	D2S428	GATA14B12	0.94		145	159	GDB
02	D2S430	GGAT2F11	N/A		340	340	GDB
02	D2S433	GATA3F05	0.77		179	199	GDB
02	D2S434	GATA4G12	0.82		262	286	GDB
02	D2S435	GATA5B07	0.80		192	212	GDB
02	D2S436	GATA5G02	0.90		179	202	GDB
02	D2S437	GATA6A03	0.81		185	221	GDB
02	D2S438	GATA6C12	0.67		147	159	GDB
02	D2S439	GATA6E08	0.50		165	193	GDB
02	D2S440	GATA6F08	0.60		187	207	GDB
02	D2S441	GATA8F03	0.75		127	159	GDB
02	D2S442	GATA8H05	0.81		196	208	GDB
02	D2S443	GGAA4D07	0.81		223	255	GDB
02	D2S444	GGAT4C08	0.56		110	126	GDB
02	D2S71	MFD 19	0.59		138	154	NAR 18(8):2203, 1990
02	D2S72	MFD 36	0.83	0.71	159	173	NAR 18(8):2200, 1990
02	D2S73	MFD 54	0.70	0.52	140	150	JONES,M. ET AL.(1992) HMG 1,131-33.
02	D2S93	MIT-G105	0.83		146	146	HUMAN GENET 87:401, 1991
02	D2S94	MIT-MS153	0.75		150	150	HUMAN GENET 87:401, 1991
02	D2S95	MIT-A119	0.85		146	146	HUMAN GENET 87:401, 1991
02	D2S96	MIT-N118	0.78		178	178	HUMAN GENET 87:401, 1991
02	D2S97	MIT-MS211	0.81		105	105	HUMAN GENET 87:401, 1991
02	D2S98	MIT-MS222	0.71		131	131	HUMAN GENET 87:401, 1991
02	D2S99	MIT-F6	0.73		192	192	HUMAN GENET 87:401, 1991
02	GCG	NA	0.82		125	125	WU,S. ET AL.(1991) NAR 19,1163.
02	HOX4E	NA	0.88		104	130	ROSEN,D.& BROWN,JR.,R.(1993) HMG 2,617.
02	IL1A	MFD 68	0.75	0.67	131	145	JONES,M. ET AL.(1992) HMG 1,131-33.
02	PAX3	NA		0.89	310	336	WOOD,S.&SCHERTZER,M.(1992)GENOMICS13,232.
02	TPO	NA	0.67	0.61	106	130	GENOMICS 8:400- , 1990
03	ACPP	NA	0.69	0.65	260	280	POLYMEROPOULOS,M. ET AL.(1991) NAR 19,4792.
03	AGTR1	ATCA	0.73		140	146	DAVIES, E. ET AL. (1994) HMG 3, 838.
03	D3F122S1E	CTG-33	0.70		116	116	GDB
03	D3S1007	NA	0.70		81	81	JORDAN,S.A.ET AL.(1991)NAR 19,1171.
03	D3S1029	CI3-917	0.62		168	168	JONES,M. ET AL.(1992) HMG 1,131-33.
03	D3S1038	NA	0.80		115	115	JORDAN,S.A.ET AL.(1991)NAR 19,1171.
03	D3S1067	NA	0.86		95	95	JORDAN,S.A.ET AL.(1991)NAR 19,1171.
03	D3S1076	CI13-1126	0.59		119	119	JONES,M. ET AL.(1992) HMG 1,131-33.
03	D3S1100	3GTA8B	0.82		154	170	GDB
03	D3S1110	CI3-1169	0.67		66	66	JONES,M. ET AL.(1992) HMG 1,131-33.
03	D3S1209	MIT-MS24	0.75		156	156	HUMAN GENET 87:401, 1991
03	D3S1210	MIT-MS140	0.71		157	157	HUMAN GENET 87:401, 1991

CH	LOCUS	ASSAY	HET	PIC	SIZE ANGE MIN	MAX	REFERENCE
03	D3S1211	MIT-I106	0.88		181	181	HUMAN GENET 87:401, 1991
03	D3S1212	MIT-E109	0.75		193	193	HUMAN GENET 87:401, 1991
03	D3S1214	MIT-E144	0.60		151	151	HUMAN GENET 87:401, 1991
03	D3S1215	MIT-MS207	0.78		101	101	HUMAN GENET 87:401, 1991
03	D3S1216	MIT-K117	0.86		170	170	HUMAN GENET 87:401, 1991
03	D3S1217	MIT-F8	0.83		190	190	HUMAN GENET 87:401, 1991
03	D3S1227	MIT-MS238	0.61		190	190	HUMAN GENET 87:401, 1991
03	D3S1228	NA	0.77		84	84	JORDAN,S.A.ET AL.(1991)NAR 19,1171.
03	D3S1229	NA	0.84	0.83	109	127	XIAO,H. ET AL.(1992)HMG 1,290.
03	D3S1232	MFD 16	0.75	0.75	148	176	GOTO,J. ET AL.(1992) HMG 1,350.
03	D3S1233	MFD 53	0.45	0.40	143	151	GOTO,J. ET AL.(1992) HMG 1,350.
03	D3S1234	MFD 76	0.66	0.54	99	121	GOTO,J. ET AL.(1992) HMG 1,350.
03	D3S1235	MFD 93	0.44	0.38	133	137	GOTO,J. ET AL.(1992) HMG 1,350.
03	D3S1236	MFD 99	0.50	0.55	69	85	GENOMICS 8:400- , 1990
03	D3S1237	MFD124	0.76	0.72	176	198	GOTO,J. ET AL.(1992) HMG 1,350.
03	D3S1238	MFD125	0.73	0.68	118	146	GOTO,J. ET AL.(1992) HMG 1,350.
03	D3S1239	MFD173	0.27	0.33	189	193	GENOMICS 8:400- , 1990
03	D3S1241	LIB 3-15	0.84		191	207	SCHMIDT, L. ET AL. (1993) HMG 2, 817-818.
03	D3S1243	MFD182	0.66	0.50	177	185	GENOMICS 1993, SUBMITTED
03	D3S1246	NA	0.82		110	128	XIAO,H. ET AL.(1992) HMG 1,652.
03	D3S1247	NA	0.80		153	173	XIAO,H. ET AL.(1992)HMG 1,290.
03	D3S1251	MFD201	0.77	0.74	125	139	GOTO,J. ET AL.(1992) HMG 1,350.
03	D3S1252	MFD203	0.67	0.62	164	184	GOTO,J. ET AL.(1992) HMG 1,350.
03	D3S1253	MFD205	0.48	0.45	77	89	GOTO,J. ET AL.(1992) HMG 1,350.
03	D3S1254	MFD233	0.46	0.43	212	220	GOTO,J. ET AL.(1992) HMG 1,350.
03	D3S1255	R7K110	0.86		156	168	KLAUKS, S. ET AL. (1994) 3, 840.
03	D3S1258	AFM031yc5	0.50		165	177	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1259	AFM036yb8	0.84		184	206	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1260	AFM038xc1	0.66		254	276	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1261	AFM057xc5	0.85		185	217	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1262	AFM059xa9	0.81		112	126	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1263	AFM079yg5	0.87		231	249	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1264	AFM081yh3	0.80		253	263	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1265	AFM087yb7	0.86		212	236	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1266	AFM095xc1	0.74		289	299	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1267	AFM116xh2	0.88		129	167	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1268	AFM116xh4	0.87		123	145	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1269	AFM119yh2	0.84		198	216	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1270	AFM122xb6	0.75		164	186	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1271	AFM126zc5	0.75		146	158	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1272	AFM136xc1	0.79		262	276	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1273	AFM151yf2	0.81		100	122	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1274	AFM154xa7	0.61		128	136	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1275	AFM157xb4	0.74		257	281	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1276	AFM161xg11	0.72		190	202	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1277	AFM164we1	0.82		260	280	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1278	AFM164xc5	0.88		203	231	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1279	AFM164yg9	0.86		264	282	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1280	AFM177xe9	0.76		250	264	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1281	AFM177xh8	0.67		114	132	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1282	AFM179xh10	0.80		140	154	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1283	AFM183yc3	0.71		150	160	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1284	AFM184yb2	0.75		155	177	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1285	AFM191yg5	0.74		232	242	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1286	AFM197xg11	0.89		119	153	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1287	AFM198te7	0.69		255	263	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1288	AFM198ya3	0.75		233	249	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1289	AFM198yf2	0.82		197	215	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1290	AFM198yh6	0.82		210	230	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1292	AFM199xd6	0.85		142	166	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1293	AFM200za1	0.80		116	144	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1294	AFM207yd8	0.74		181	213	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1295	AFM210yc5	0.63		126	146	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1296	AFM211we3	0.74		176	186	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1297	AFM217xd2	0.83		116	144	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1298	AFM220yb8	0.88		194	220	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1299	AFM220yh12	0.62		218	228	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1300	AFM220yh4	0.83		217	241	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1301	AFM224xf10	0.78		187	197	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1302	AFM225xb12	0.65		127	143	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801

CH	LOCUS	ASSAY	HET	PIC	SIZE MIN	ANGE MAX	REFERENCE
03	D3S1303	AFM225yd6	0.78		196	220	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1304	AFM234tf4	0.81		253	269	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1305	AFM234tg3	0.74		189	198	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1306	AFM234wa1	0.74		156	170	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1307	AFM238wb12	0.81		237	251	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1308	AFM240ve1	0.70		104	112	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1309	AFM240ya11	0.75		132	150	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1310	AFM242xh2	0.68		137	147	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1311	AFM254ve1	0.84		134	152	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1312	AFM256ya9	0.77		215	225	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1313	AFM259zg5	0.69		228	238	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1314	AFM260yb1	0.87		144	170	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1315	AFM263zc9	0.55		238	250	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1316	AFM268vc9	0.71		273	293	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
03	D3S1317	NA	0.68		159	171	LIM,L.& GILL,M.(1993) HMG 2,616.
03	D3S1339	LIB 45-17	0.70		128	140	SCHMIDT, L. ET AL. (1993) HMG 2, 817-818.
03	D3S1340	LIB 23-42	0.89		113	145	SCHMIDT, L. ET AL. (1993) HMG 2, 817-818.
03	D3S1349	NA	0.79		118	142	GDB
03	D3S1350	NA	0.70		208	232	GDB
03	D3S1352	NA	0.76		108	132	LINARES-RUIZ,A.(1993) HMG 2,1508.
03	D3S1358	NA	0.78		97	121	LINARES-RUIZ,A.(1993) HMG 2,1508.
03	D3S1359	NA	0.89		213	257	LI,H. ET AL.(1993) HMG 2,1327.
03	D3S1417	GTO1	0.65		208	220	THISELTON,D. ET AL.(1993)HMG 2,613.
03	D3S1418	GT06	0.84		222	242	TALBOT,C. ET AL.(1993) HMG 2,1325.
03	D3S1432	NA	0.67		86	94	LI, H. ET AL. (1994) HMG 3, 837.
03	D3S1434	NA	0.75		90	106	LI, H. ET AL. (1994) HMG 3, 837.
03	D3S1435	NA	0.79		154	164	LI,H. ET AL.(1993) HMG 2,1326.
03	D3S1439	MFD303	0.55	0.51	170	180	GENOMICS 8:400- , 1990
03	D3S1447	NA	0.77		121	137	GDB
03	D3S1448	NA	0.64		192	204	GDB
03	D3S1449	NA	0.63		173	185	GDB
03	D3S1450	NA	0.55		244	258	GDB
03	D3S1480	3051	0.78		174	174	GDB
03	D3S1481	1116	0.83		104	104	GDB
03	D3S1483	NA	0.89		270	290	LI, H. ET AL. (1994) HMG 3, 837.
03	D3S1484	NA	0.77		180	188	LI, H. ET AL. (1994) HMG 3, 837.
03	D3S1485	NA	0.81		180	192	LI, H. ET AL. (1994) HMG 3, 837.
03	D3S1489	NA	0.76		126	136	LI, H. ET AL. (1994) HMG 3, 837.
03	D3S1490	NA	0.77		132	144	LI, H. ET AL. (1994) HMG 3, 837.
03	D3S1497	NA	0.82		128	142	LI, H. ET AL. (1994) HMG 3, 837.
03	D3S1498	NA	0.75		136	150	LI, H. ET AL. (1994) HMG 3, 837.
03	D3S1547	AFM162xc9	0.63		93	100	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1548	AFM164tc7	0.76		165	171	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1549	AFM182yc5	0.61		212	232	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1550	AFM184yf4	0.82		142	150	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1551	AFM198yc1	0.57		241	259	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1552	AFM200xg5	0.55		130	138	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1553	AFM200ze11	0.62		163	175	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1554	AFM205va9	0.81		133	141	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1555	AFM056xf12	0.79		221	237	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1556	AFM210va7	0.76		248	262	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1557	AFM210ve7	0.74		177	201	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1558	AFM057xa5	0.68		152	166	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1559	AFM023xg1	0.77		137	159	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1560	AFM217xd6	0.59		235	253	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1561	AFM220xg3	0.69		226	250	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1562	AFM220ze1	0.64		168	192	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1563	AFM224xc9	0.67		216	226	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1564	AFM224ya3	0.72		176	190	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1565	AFM224zb12	0.80		239	245	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1566	AFM234tb8	0.68		225	247	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1567	AFM234ya11	0.56		241	253	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1569	AFM240xb2	0.26		277	297	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1570	AFM242xc3	0.52		259	271	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1571	AFM242yc9	0.62		160	184	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1572	AFM259va9	0.72		244	266	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1573	AFM126xg7	0.76		136	154	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1574	AFM155xd6	0.77		103	127	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1575	AFM263yh5	0.55		258	264	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1576	AFM267xd9	0.84		189	203	WEISSENBACH J: NATURE GENETIC, JUNE 1994

CH	LOCUS	ASSAY	HET	PIC	SIZE ANGE MIN	MAX	REFERENCE
03	D3S1577	AFM267zf9	0.69		221	235	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1578	AFM268wg9	0.60		140	166	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1579	AFM270yc5	0.76		151	159	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1580	AFM270zg9	0.69		139	155	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1581	AFM273ve9	0.80		78	102	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1582	AFM274yd5	0.58		154	178	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1583	AFM276vc9	0.64		149	173	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1584	AFM277wf9	0.70		148	162	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1585	AFM283vb5	0.63		126	144	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1586	AFM284xf9	0.70		291	309	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1587	AFM284ze5	0.73		215	227	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1588	AFM287yd9	0.80		212	236	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1589	AFM290zf1	0.56		159	169	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1591	AFM292xg5	0.66		241	251	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1592	AFM292xh1	0.63		281	287	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1593	AFM292ye5	0.74		137	153	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1594	AFM292ze1	0.46		266	334	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1595	AFM294zf9	0.72		295	317	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1597	AFM295yc9	0.67		162	180	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1598	AFM296vd5	0.83		290	316	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1599	AFM301ze9	0.77		134	140	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1600	AFM308xc9	0.61		182	198	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1601	AFM308yf1	0.63		184	214	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1602	AFM308zh9	0.81		275	297	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1603	AFM311vh1	0.58		159	177	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1604	AFM316vc1	0.73		247	255	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1605	AFM317xe1	0.68		141	163	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1606	AFM318wc5	0.66		236	252	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1607	AFM319yb1	0.71		230	244	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1608	AFM319zf1	0.70		184	206	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1609	AFM320wc9	0.76		253	269	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1610	AFM321xf5	0.81		171	189	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1611	AFM338xe5	0.79		252	268	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1612	AFM339xh1	0.83		206	226	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1613	AFM340xf1	0.49		225	253	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1614	AFM345th5	0.74		143	157	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1615	AFM347yg1	0.60		170	186	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1616	AFM348te9	0.83		101	107	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1617	AFM349xc5	0.80		250	254	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1618	AFM350te5	0.75		150	158	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1619	AFM350tf1	0.74		161	171	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1620	AFM351wc1	0.66		239	255	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1621	AFMa133wh1	0.74		97	139	WEISSENBACH J: NATURE GENETIC, JUNE 1994
03	D3S1744	GATA3C02	0.80		131	167	GDB
03	D3S1745	GATA8A06	0.58		213	213	GDB
03	D3S1746	GATA8F01	0.76		248	284	GDB
03	D3S1752	ATC3D09	0.81		181	208	GDB
03	D3S1753	GATA11F06	0.50		300	300	GDB
03	D3S1754	GATA14G12	1.00		187	187	GDB
03	D3S1759	GGAT2A01	0.68		280	280	GDB
03	D3S1763	GATA3H01	0.50		260	280	GDB
03	D3S1764	GATA4A10	0.69		225	253	GDB
03	D3S1765	GATA4G01	0.82		192	212	GDB
03	D3S1766	GATA6F06	0.86		208	232	GDB
03	D3S1767	GATA7A01	0.68		244	264	GDB
03	D3S1768	GATA8B05	0.75		186	206	GDB
03	D3S1769	GATA8D02	0.71		249	277	GDB
03	D3S1776	F127F91	0.75		205	217	TODD, S. ET AL. (1994) 3, 841.
03	D3S192	LIB 49-63	0.86		96	118	SCHMIDT, L. ET AL. (1993) HMG 2, 817-818.
03	D3S196	MFD 17	0.67	0.68	86	98	NAR 18(15):4635, 1990
03	D3S2384	ATA4D09	0.80		115	133	GDB
03	D3S2385	GAAT3E04	0.57		142	154	GDB
03	D3S2386	GATA13H08	0.88		274	310	GDB
03	D3S2387	GATA22G12	0.83		177	213	GDB
03	D3S2388	GATA24E11	0.63		101	125	GDB
03	D3S2389	GATA29C03	0.69		268	268	GDB
03	D3S2390	GATA31E08	0.83		247	247	GDB
03	D3S2391	GATA31G11	0.64		186	186	GDB
03	D3S2392	GATA41H09	0.75		132	132	GDB
03	D3S2393	GATA43D03	0.67		398	398	GDB

CH	LOCUS	ASSAY	HET	PIC	SIZE MIN	ANGE MAX	REFERENCE
03	D3S2394	GATA43D09	0.50		257	257	GDB
03	D3S2395	GATA49D12	1.00		196	196	GDB
03	D3S2396	GATA4F11	0.87		183	183	GDB
03	D3S2397	GATA51A05	0.62		200	200	GDB
03	D3S2398	GATA6G12	0.83		266	298	GDB
03	D3S2399	GCT3C11	0.18		183	183	GDB
03	D3S240	MFD 30	0.28	0.30	83	99	NAR 18(8):2203, 1990
03	D3S2400	GCT4B10	0.46		209	209	GDB
03	D3S2401	GCT5E11	1.00		254	254	GDB
03	D3S2402	GGAA13D09	0.73		250	250	GDB
03	D3S2403	GGAA4B09	0.75		248	292	GDB
03	D3S2404	GGAA6B07	0.91		106	158	GDB
03	D3S2405	GGAT2A11	0.85		109	109	GDB
03	D3S2406	GGAT2G03	0.88		306	350	GDB
03	D3S587	NA	0.77		125	143	LIM,L.& GILL,M.(1993) HMG 2,616.
03	D3S621	NA	0.79		208	212	GDB
03	D3S643	NA	0.77		113	113	JONES,M. ET AL.(1992) HMG 1,131-33.
03	D3S647	NA	0.73		102	102	JORDAN,S.A.ET AL.(1991)NAR 19,1171.
03	D3S656	CI3-326	0.57		96	96	JONES,M. ET AL.(1992) HMG 1,131-33.
03	D3S659	NA	0.73		113	113	JORDAN,S.A.ET AL.(1991)NAR 19,1171.
03	D3S663	NA	0.73		92	92	JORDAN,S.A.ET AL.(1991)NAR 19,1171.
03	D3S688	NA	0.73		110	110	JORDAN,S.A.ET AL.(1991)NAR 19,1171.
03	D3S769	LIB44-36ca	0.84		158	174	SCHMIDT,L. ET AL.(1993) HMG 2,89.
03	D3S966	NA	0.77		147	147	JORDAN,S.A.ET AL.(1991)NAR 19,1171.
03	GLUT2	NA	0.75		116	124	PATEL,P. ET AL.(1991) NAR 19,4017.
03	HRG	NA	0.82		233	267	HENNIS,B. ET AL.(1992) HMG 1,781.
03	IL5RA	NA	0.78		90	104	GDB
03	RHO	MFD 2	0.34	0.31	118	124	AM J HUM GEN 44:388-396, 1989
03	SST	MFD 4	0.51	0.46	163	175	AM J HUMAN GENET, 1993, IN PRESS
03	THRB	NA	0.50		197	209	SAKURAI,A.ET AL.(1991) NAR 19,6661.
03	THRB-5	THRB-5	0.66		189	201	GDB
04	ADRA2C	NA	0.73		179	193	193 RIESS,O. ET AL.(1992) HMG 1,452.
04	N/A	GATA29	0.71		N/A	N/A	GDB
04	D4S1089	MFD268	0.69	0.65	117	133	GENOMICS 8:400- , 1990
04	D4S1090	MFD258	0.85	0.84	192	216	GENOMICS 8:400- , 1990
04	D4S1091	MFD281	0.53	0.47	127	137	GENOMICS 8:400- , 1990
04	D4S126	C102	0.81		155	177	GDB
04	D4S127	NA	0.71	0.70	143	159	GUSELLA,J.F. ET AL.(1992)NAR 1,142.
04	D4S1534	AFM155xe11	0.51		146	158	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1535	AFM165zf8	0.65		177	191	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1536	AFM168xa5	0.79		262	272	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1537	AFM031yc7	0.70		121	133	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1538	AFM036xg3	0.67		149	161	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1539	AFM185xe1	0.45		221	229	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1540	AFM185xf8	0.56		185	193	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1541	AFM036yb2	0.48		151	159	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1542	AFM189yc3	0.80		215	219	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1543	AFM191xh2	0.65		144	170	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1544	AFM197ya5	0.81		243	251	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1545	AFM198xf10	0.83		199	209	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1546	AFM200wc11	0.78		146	160	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1547	AFM200yc7	0.62		224	238	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1548	AFM200zh12	0.77		245	271	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1549	AFM045xc1	0.58		203	217	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1550	AFM203yd4	0.72		206	212	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1551	AFM207we11	0.81		172	186	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1552	AFM210wd2	0.74		171	199	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1553	AFM210wh8	0.70		202	206	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1554	AFM218yb4	0.70		184	208	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1555	AFM220za3	0.61		275	283	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1556	AFM225zb2	0.70		157	171	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1557	AFM077yf11	0.59		140	148	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1558	AFM238xf6	0.69		274	292	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1559	AFM238zf10	0.86		259	269	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1560	AFM242xd8	0.55		250	270	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1561	AFM242ye9	0.63		294	306	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1562	AFM248td5	0.73		102	114	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1564	AFM248zg9	0.70		220	242	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1565	AFM261zg5	0.81		134	146	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1566	AFM025xd2	0.60		197	209	WEISSENBACH J: NATURE GENETIC, JUNE 1994

CH	LOCUS	ASSAY	HET	PIC	SIZE ANGE MIN	MAX	REFERENCE
04	D4S1567	AFM109xd12	0.80		100	114	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1569	AFM135xe1	0.88		279	291	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1570	AFM150yh6	0.76		181	193	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1571	AFM263yh9	0.64		163	166	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1572	AFM265va9	0.67		135	151	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1573	AFM266vf9	0.57		101	113	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1574	AFM267zc5	0.69		186	194	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1575	AFM268yb1	0.57		219	223	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1576	AFM273yg9	0.82		237	251	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1577	AFM274xb9	0.81		119	125	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1578	AFM274zc5	0.77		210	232	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1579	AFM275yc1	0.79		142	160	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1580	AFM277ye1	0.68		268	278	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1581	AFM280tf5	0.74		210	220	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1582	AFM283zf5	0.77		112	126	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1583	AFM284vc5	0.77		219	239	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1584	AFM284vh5	0.60		121	131	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1585	AFM284xh1	0.81		131	147	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1586	AFM288vb9	0.72		103	117	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1587	AFM288yb5	0.62		234	246	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1589	AFM290vc5	0.89		207	217	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1590	AFM290zg5	0.83		98	116	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1591	AFM291zd1	0.88		110	124	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1592	AFM292xe1	0.75		170	196	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1593	AFM294vb1	0.56		138	154	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1594	AFM294vb9	0.65		237	247	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1595	AFM294zh9	0.46		197	207	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1596	AFM295xd5	0.67		213	223	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1597	AFM295ye5	0.64		273	293	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1598	AFM296vd1	0.87		262	276	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1599	AFM296wb1	0.82		142	156	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1600	AFM297za5	0.55		129	141	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1601	AFM299zf5	0.85		124	146	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1602	AFM303xe5	0.80		222	233	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1603	AFM304vc9	0.69		190	208	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1604	AFM304wb9	0.78		254	260	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1605	AFM304xa5	0.44		127	133	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1607	AFM310wf5	0.77		160	184	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1608	AFM312wb1	0.81		231	247	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1609	AFM312wc9	0.78		163	177	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1610	AFM318xg9	0.55		169	183	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1611	AFM319yg9	0.75		277	285	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1612	AFM319zc1	0.61		160	186	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1613	AFM329vg1	0.34		257	263	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1614	AFM331xh5	0.79		143	149	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1615	AFM336xh5	0.37		115	125	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1616	AFM343wf9	0.80		245	257	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1617	AFM345wh9	0.65		181	191	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1618	AFM345xa5	0.76		241	273	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1619	AFM356tc5	0.83		145	153	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S1625	GATA107	0.76		182	210	GDB
04	D4S1626	GATA4C05	0.82		170	170	GDB
04	D4S1627	GATA7D01	0.57		177	201	GDB
04	D4S1628	GATA6D11	0.76		149	165	GDB
04	D4S1629	GATA8A05	0.68		151	151	GDB
04	D4S1630	GATA8D12	0.76		139	155	GDB
04	D4S1631	GATA2C02	0.92		131	145	GDB
04	D4S1632	GATA3A12	0.80		268	276	GDB
04	D4S1634	MFD206	0.64	0.59	168	184	GENOMICS 8:400- , 1990
04	D4S1635	MFD222	0.70	0.67	99	129	GENOMICS 8:400- , 1990
04	D4S1636	MFD223	0.80	0.78	118	140	GENOMICS 8:400- , 1990
04	D4S1637	MFD315	0.73	0.68	120	132	GENOMICS 8:400- , 1990
04	D4S1638	MFD324	0.77	0.74	138	150	GENOMICS 8:400- , 1990
04	D4S1643	GATA11D07	0.73		174	190	GDB
04	D4S1644	GATA11E09	0.81		186	206	GDB
04	D4S1645	GATA11F12	0.81		225	257	GDB
04	D4S1646	GATA2F08	0.44		163	183	GDB
04	D4S1647	GATA2F11	1.00		132	156	GDB
04	D4S1651	GATA4C04	1.00		194	206	GDB
04	D4S1652	GATA5B02	0.56		136	148	GDB

CH	LOCUS	ASSAY	HET	PIC	SIZE MIN	ANGE MAX	REFERENCE
04	D4S1653	GATA5B09	0.88		216	228	GDB
04	D4S1654	GATA6F07	0.59		147	167	GDB
04	D4S171	MFD 22	0.75	0.67	143	161	GENOMICS 14:209-219, 1992
04	D4S174	MFD 59	0.92	0.86	175	195	GENOMICS 15:251-258, 1993
04	D4S175	MFD 38	0.82	0.82	112	134	GENOMICS 15:251-258, 1993
04	D4S179	MFD 83	0.23	0.20	220	222	GENOMICS 15:251-258, 1993
04	D4S188	MFD 71	0.07	0.11	98	103	GENOMICS 15:251-258, 1993
04	D4S189	MFD 74	0.78	0.69	174	194	GENOMICS 15:251-258, 1993
04	D4S190	MFD106	0.55	0.55	148	164	GENOMICS 15:251-258, 1993
04	D4S191	MFD138	0.67	0.57	86	96	GENOMICS 15:251-258, 1993
04	D4S192	MFD140	0.80	0.73	85	97	GENOMICS 15:251-258, 1993
04	D4S193	MFD142	0.69	0.71	101	111	GENOMICS 15:251-258, 1993
04	D4S194	MFD146	0.66	0.78	97	115	GENOMICS 15:251-258, 1993
04	D4S230	MFD194	0.84	0.83	191	217	GENOMICS 15:251-258, 1993
04	D4S231	NA	0.71		157	169	MOORE,B. ET AL.(1992)NAR 20,929.
04	D4S2361	ATA2A03	0.62		149	164	GDB
04	D4S2362	GAAT1D1	0.31		146	154	GDB
04	D4S2363	GAAT1E01	0.57		121	129	GDB
04	D4S2364	GAAT1F09	0.44		120	128	GDB
04	D4S2365	GATA10A12	0.47		296	296	GDB
04	D4S2366	GATA22G05	0.83		120	144	GDB
04	D4S2367	GATA24H01	0.77		127	147	GDB
04	D4S2368	GATA27G03	0.70		304	328	GDB
04	D4S2369	GATA29G09	0.71		240	240	GDB
04	D4S2370	GATA30B10	0.75		253	253	GDB
04	D4S2371	GATA31H04	0.46		184	184	GDB
04	D4S2372	GATA32A10	0.86		165	165	GDB
04	D4S2373	GATA32A12	0.86		134	134	GDB
04	D4S2374	GATA42E01	0.89		220	220	GDB
04	D4S2375	GATA42E03	0.56		288	288	GDB
04	D4S2376	GATA44B07	0.50		201	201	GDB
04	D4S2377	GATA48D06	0.50		158	158	GDB
04	D4S2378	GATA50C05	0.50		240	240	GDB
04	D4S2379	GGAA16B09	0.79		319	319	GDB
04	D4S2380	GGAA18A06	1.00		194	194	GDB
04	D4S2381	GGAA6H08	0.62		256	288	GDB
04	D4S2382	GGAT2C10	0.81		230	246	GDB
04	D4S243	MIT-MH43	0.67		173	173	HUMAN GENET 87:401, 1991
04	D4S244	MIT-MS176	0.83		148	148	HUMAN GENET 87:401, 1991
04	D4S245	MIT-N133	0.52		107	107	HUMAN GENET 87:401, 1991
04	D4S246	MIT-MS205	0.67		163	163	HUMAN GENET 87:401, 1991
04	D4S247	MIT-MS240	0.82		173	173	HUMAN GENET 87:401, 1991
04	D4S250	MIT-N136	0.83		215	215	HUMAN GENET 87:401, 1991
04	D4S251	NA	0.83		100	124	PETRUKHIN,K. ET AL. (1992) HMG 1,349.
04	D4S349	MFD202	0.56		109	117	JONES,M. ET AL.(1992) HMG 1,131-33.
04	D4S350	MFD208	0.34		233	243	JONES,M. ET AL.(1992) HMG 1,131-33.
04	D4S391	AFM016xf3	0.86		165	185	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
04	D4S392	AFM022xc1	0.84		93	107	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
04	D4S393	AFM036xg11	0.73		104	118	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
04	D4S394	AFM037yg1	0.78		89	215	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
04	D4S395	AFM046xd4	0.85		113	137	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
04	D4S396	AFM057xe5	0.56		227	235	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
04	D4S397	AFM123xa9	0.66		192	214	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
04	D4S398	AFM135xc3	0.84		125	147	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
04	D4S399	AFM142xd4	0.87		128	150	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
04	D4S400	AFM143xh10	0.64		169	179	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
04	D4S401	AFM150xe5	0.56		187	193	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
04	D4S402	AFM151xc3	0.92		287	323	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
04	D4S403	AFM157xg3	0.78		217	231	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
04	D4S404	AFM158xc7	0.79		89	101	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
04	D4S405	AFM161yf6	0.87		279	299	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
04	D4S406	AFM164tf6	0.88		234	258	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
04	D4S407	AFM164wd10	0.87		111	135	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
04	D4S408	AFM165xc11	0.77		229	243	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
04	D4S409	AFM183xd6	0.76		275	305	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
04	D4S410	AFM183xh4	0.53		214	220	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
04	D4S411	AFM186xa7	0.67		135	143	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
04	D4S412	AFM196xb6	0.77		237	249	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
04	D4S413	AFM196xe1	0.86		128	166	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
04	D4S414	AFM196xf4	0.89		227	242	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801

CH	LOCUS	ASSAY	HET	PIC	SIZE RANGE MIN	SIZE RANGE MAX	REFERENCE
04	D4S415	AFM197xa11	0.81		172	202	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
04	D4S416	AFM198we3	0.47		211	215	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
04	D4S417	AFM200vb10	0.45		174	186	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
04	D4S418	AFM203xa9	0.78		210	226	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
04	D4S419	AFM207wa11	0.78		222	238	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
04	D4S420	AFM211zc7	0.60		196	216	WEISSENBACH J: NATURE GENETIC, JUNE 1994
04	D4S421	AFM218xh8	0.70		105	113	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
04	D4S422	AFM224xa9	0.80		75	97	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
04	D4S423	AFM224xd6	0.84		103	125	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
04	D4S424	AFM225xb10	0.84		178	192	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
04	D4S425	AFM234va1	0.76		178	192	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
04	D4S426	AFM238ve3	0.77		177	191	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
04	D4S427	AFM249ve1	0.84		142	166	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
04	D4S428	AFM255zf1	0.77		187	203	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
04	D4S429	AFM256za9	0.73		193	207	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
04	D4S430	AFM260za5	0.82		119	141	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
04	D4S431	AFM262vg9	0.84		246	270	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
04	D4S432	AFM266wb9	0.74		224	272	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
04	D4S620	AFM067xa5	0.65		109	121	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
04	D4S621	AFM070ya9	0.64		253	261	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
04	D4S622	AFM186xd2	0.65		84	92	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
04	DRD5	PCR1	0.77		134	156	GDB
04	EGF	NA	0.67		197	197	MURRAY, J. ET AL. (1992) NATURE GENETICS 2, 46-49.
04	F11	NA	0.45		372	376	BODFISH,P. ET AL.(1991)NAR 19,6979.
04	FABP2	NA	0.64		99	117	POLYMEROPOULOS,M. ET AL.(1991) NAR 19,7198.
04	FGA	NA	0.86		256	284	MILLS,K. ET AL.(1992) HMG 1,779.
04	GABRB1	NA	0.72	0.68	91	99	POLYMEROPOULOS,M. ET AL.(1991) NAR 19,6345.
04	HD	L191F1	0.80		247	247	GDB
04	HOX7	NA	0.62		169	175	MILLS,K. ET AL.(1992) GENOMICS 14,209-219.
04	IL2	NA	0.89		115	147	GDB
04	PDEB	E24	0.75		203	215	HMG,IN PRESS.
05	CFS1R	NA	0.86	0.85	95	127	POLYMEROPOULOS,M. ET AL.(1991) NAR 19,1160.
05	CRTL1	NA	0.85		222	240	HECHT,J.T. ET AL.(1991)NAR 19,6666.
05	CSF1R-T	PCR4	0.73		180	204	GDB
05	N/A	GATA43	0.85				GDB
05	N/A	GATA45	0.83				GDB
05	D5S107	MFD 27	0.82	0.78	133	155	NAR 18(13):4035, 1990
05	D5S108	MFD 34	0.51	0.45	83	93	NAR 18(13):4037, 1990
05	D5S111	MFD 40	0.53	0.51	167	171	NAR 18(13):4037, 1990
05	D5S117	MFD 48	0.55	0.62	147	163	NAR 18(13):4037, 1990
05	D5S118	MFD 63	0.56	0.48	78	92	NAR 18(13):4037, 1990
05	D5S119	MFD 6	0.49	0.50	190	202	AM J HUMAN GENET, 1993, IN PRESS
05	D5S1356	NA	0.73		107	129	VELASCO, E. ET AL. (1994) HMG 3, 1441.
05	D5S1357	NA	0.67		123	153	VELASCO, E. ET AL. (1994) HMG 3, 1441.
05	D5S1452	ATA3F07	0.50		102	117	GDB
05	D5S1453	ATA4D10	0.70		139	163	GDB
05	D5S1454	ATA4F06	0.70		184	208	GDB
05	D5S1455	GAAT1A12	0.44		141	149	GDB
05	D5S1456	GATA11A11	0.88		191	211	GDB
05	D5S1457	GATA21D04	0.77		97	127	GDB
05	D5S1458	GATA22C12	0.72		178	190	GDB
05	D5S1459	GATA23G12	0.73		90	110	GDB
05	D5S1460	GATA2A03	0.52		301	301	GDB
05	D5S1461	GATA30B07	0.70		201	201	GDB
05	D5S1462	GATA3H06	0.86		217	241	GDB
05	D5S1463	GATA41D11	0.86		194	194	GDB
05	D5S1464	GATA41G03	0.70		254	254	GDB
05	D5S1465	GATA47H08	0.78		262	262	GDB
05	D5S1466	GATA49C05	0.83		304	304	GDB
05	D5S1467	GATA49E12	0.92		157	157	GDB
05	D5S1468	GATA50C10	0.62		314	314	GDB
05	D5S1469	GATA51B01	0.62		179	179	GDB
05	D5S1470	GATA7C06	0.85		197	197	GDB
05	D5S1471	GATA7H10	0.71		152	172	GDB
05	D5S1472	GCT5E05	0.46		133	133	GDB
05	D5S1473	GGAA11C04	0.88		250	250	GDB
05	D5S1474	GGAA9A12	0.50		248	268	GDB
05	D5S1475	GGAT4A03	0.63		154	154	GDB
05	D5S178	NA	0.93		98	120	SZUBRYT,S. ET AL. (1993) HMG 2,90.
05	D5S207	MFD 43	0.69	0.63	135	143	GENOMICS 11:695-700, 1991

CH	LOCUS	ASSAY	HET	PIC	SIZE MIN	ANGE MAX	REFERENCE
05	D5S208	MFD 88	0.69	0.73	140	158	GENOMICS 12:607-609, 1992
05	D5S209	MFD116	0.71	0.67	194	214	GENOMICS 12:607-609, 1992
05	D5S210	MFD122	0.75	0.75	212	232	GENOMICS 12:607-609, 1992
05	D5S211	MFD154	0.72	0.69	186	204	GENOMICS 12:607-609, 1992
05	D5S253	NA	0.77	0.00	102	128	BERNARD,L.& WOOD,S.(1992)HMG 1:455.
05	D5S260	NA	0.74	0.69	146	154	BERNARD,L.E. ET AL.(1991)NAR 19,6970.
05	D5S268	NA	0.00	0.68	112	124	BERNARD,L ET AL.(1991) NAR 19,5794.
05	D5S299	NA	0.00	0.66	156	182	VAN LEEUWEN,C. ET AL.(1991) NAR 19,5805.
05	D5S318	NA	0.76	0.78	176	188	WIJNEN,J. ET AL.(1991) NAR 19,6965.
05	D5S346	NA	0.63		96	122	SPIRIO,L. ET AL.(1991) NAR 19,6348.
05	D5S349	MIT-A127	0.81		140	140	HUMAN GENET 87:401, 1991
05	D5S350	MIT-MS131	0.61		136	136	HUMAN GENET 87:401, 1991
05	D5S351	MIT-I105	0.75		197	197	HUMAN GENET 87:401, 1991
05	D5S352	MIT-MS158	0.96		149	149	HUMAN GENET 87:401, 1991
05	D5S353	MIT-MH98	0.83		133	133	HUMAN GENET 87:401, 1991
05	D5S354	MIT-MH96	0.76		135	135	HUMAN GENET 87:401, 1991
05	D5S355	MIT-MH91	0.67		194	194	HUMAN GENET 87:401, 1991
05	D5S356	NA	0.90	0.89	94	132	POLYMEROPOULOS,M. ET AL.(1992) HMG 1,290.
05	D5S357	MFD151	0.64	0.71	154	168	GENOMICS 8:400- , 1990
05	D5S365	2C7	0.76		75	105	GDB
05	D5S373	NA	0.70		88	100	GDB
05	D5S385	MFD234	0.63	0.57	145	157	GENOMICS 8:400- , 1990
05	D5S39	NA	0.78		212	220	MANKOO,BS. ET AL(1991)NAR 19,1963.
05	D5S392	AFM028xb12	0.92		83	117	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S393	AFM042xd12	0.84		162	182	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S394	AFM057xh8	0.72		141	153	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S395	AFM063yb6	0.82		189	213	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S396	AFM066xf11	0.66		122	136	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S397	AFM080xh11	0.65		267	281	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S398	AFM095zb7	0.82		109	121	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S399	AFM108xe3	0.80		116	132	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S400	AFM112yb6	0.83		218	236	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S401	AFM116xe1	0.63		193	199	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S402	AFM127xh4	0.63		165	169	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S403	AFM136xd2	0.65		155	163	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S404	AFM144yf4	0.73		180	198	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S405	AFM154xa3	0.66		119	137	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S406	AFM154xg3	0.79		160	186	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S407	AFM163xa11	0.87		107	127	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S408	AFM164xb8	0.74		247	265	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S409	AFM184yb6	0.69		138	154	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S410	AFM191xd8	0.78		159	177	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S411	AFM193xe11	0.47		142	162	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S412	AFM198we11	0.84		166	182	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S413	AFM198zh6	0.70		264	276	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S414	AFM200ya9	0.83		186	206	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S415	AFM203xa7	0.85		115	147	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S416	AFM205wh10	0.78		207	292	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S417	AFM205wh8	0.75		90	104	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S418	AFM205zh4	0.82		207	223	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S419	AFM207yc1	0.82		204	226	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S420	AFM207yg5	0.58		151	159	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S421	AFM210vd6	0.83		152	172	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S422	AFM211yc7	0.85		114	134	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S423	AFM214za9	0.77		179	191	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S424	AFM214zg9	0.77		116	146	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S425	AFM225xf2	0.77		224	248	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S426	AFM238we11	0.79		183	209	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S427	AFM238xa3	0.84		280	302	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S428	AFM238xf4	0.77		241	255	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S429	AFM242xb10	0.82		160	186	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S430	AFM248wd5	0.69		256	270	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S431	AFM254yb5	0.71		169	185	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S432	AFM255xb9	0.67		109	120	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S433	AFM273yf1	0.87		192	220	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S434	AFM276wd5	0.71		246	258	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S436	AFM263wg5	0.87		234	254	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
05	D5S455	AFM016yg5	0.64		170	190	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S456	AFM022te3	0.53		103	109	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S458	AFM042xf8	0.70		282	290	WEISSENBACH J: NATURE GENETIC, JUNE 1994

CH	LOCUS	ASSAY	HET	PIC	SIZE ANGE MIN	MAX	REFERENCE
05	D5S460	AFM072zf7	0.62		129	147	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S462	AFM102xc1	0.73		135	143	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S466	AFM122xe5	0.71		175	187	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S467	AFM127xd10	0.76		155	163	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S469	AFM137xf8	0.81		142	146	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S470	AFM144zh4	0.62		236	254	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S471	AFM151xe7	0.68		236	248	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S474	AFM164yg5	0.76		87	97	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S476	AFM176xa7	0.76		167	181	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S477	AFM177xb4	0.70		167	185	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S478	AFM179xd10	0.78		257	273	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S479	AFM196xc7	0.72		130	150	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S484	AFM203va3	0.78		261	281	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S486	AFM206zc1	0.61		163	191	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S487	AFM210vg3	0.80		253	267	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S488	AFM211yh6	0.73		221	243	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S489	AFM212yb8	0.55		184	192	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S490	AFM214yg1	0.58		92	114	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S491	AFM218xf12	0.74		161	169	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S492	AFM220xg9	0.74		248	260	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S494	AFM220yg5	0.56		112	134	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S495	AFM234vc1	0.66		219	241	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S496	AFM234vh4	0.50		188	194	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S497	AFM234yh8	0.67		193	223	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S498	AFM238xe11	0.64		171	189	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S500	AFM240xg3	0.71		188	214	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S501	AFM242xc5	0.77		208	214	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S502	AFM242yc1	0.46		282	296	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S504	AFM265vg5	0.74		167	183	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S505	AFM268vb1	0.75		117	137	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S519	IG90-3	0.82		99	113	DIXON,M. ET AL.(1993) AM.J.HUM.GENET.52,907-914.
05	D5S524	COS2	0.76		174	186	GDB
05	D5S529	MFD269	0.74	0.71	188	208	GENOMICS 8:400- , 1990
05	D5S530	MFD264	0.24	0.23	173	181	GENOMICS 8:400- , 1990
05	D5S556	NA	0.72		119	149	BURLET,P. ET AL.(1993) HMG 2, 1328.
05	D5S560	MS62	0.86		151	169	GDB
05	D5S616	AFM164zc5	0.68		197	229	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S617	AFM190xc11	0.85		171	203	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S618	AFM198wg9	0.53		165	185	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S619	AFM200vf6	0.58		250	262	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S620	AFM200wa5	0.69		191	197	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S621	AFM200za11	0.74		211	217	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S622	AFM205zd4	0.70		187	197	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S623	AFM207yg11	0.78		143	159	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S624	AFM207yh2	0.90		146	166	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S625	AFM210vf12	0.48		224	238	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S626	AFM214xe9	0.77		192	206	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S627	AFM217ye1	0.61		253	259	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S628	AFM254wf1	0.65		103	127	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S629	AFM265wf5	0.62		233	253	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S630	AFM268zd9	0.81		229	333	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S631	AFM270ve9	0.63		193	213	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S634	AFM276vb1	0.39		178	192	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S635	AFM276yb9	0.79		160	170	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S636	AFM277vf5	0.76		130	152	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S637	AFM281yh9	0.34		246	254	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S638	AFM282wd5	0.67		133	145	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S639	AFM283vb9	0.79		124	136	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S640	AFM283wb5	0.64		85	105	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S641	AFM284vd1	0.51		251	281	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S642	AFM286xg9	0.59		183	201	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S643	AFM287we9	0.45		134	168	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S644	AFM288va9	0.70		81	101	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S645	AFM289wf9	0.86		170	184	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S646	AFM290vf5	0.86		271	293	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S647	AFM292ve1	0.80		126	156	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S648	AFM292yg5	0.86		116	132	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S649	AFM292ze9	0.74		171	185	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S650	AFM294wd1	0.81		204	221	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S651	AFM302wd5	0.75		177	195	WEISSENBACH J: NATURE GENETIC, JUNE 1994

CH	LOCUS	ASSAY	HET	PIC	SIZE MIN	ANGE MAX	REFERENCE
05	D5S652	AFM302yg5	0.67		202	222	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S653	AFM304xd5	0.47		133	137	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S654	AFM304zb5	0.87		241	265	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S655	AFM308vf9	0.75		254	266	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S656	AFM308wa9	0.74		185	203	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S657	AFM308xf1	0.85		138	168	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S658	AFM308za9	0.66		264	282	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S659	AFM309vd5	0.59		207	231	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S660	AFM309vd9	0.60		240	260	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S661	AFM311vb9	0.62		117	131	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S662	AFM311vf9	0.74		91	111	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S663	AFM311wh9	0.73		205	238	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S664	AFM311yd1	0.54		119	143	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S666	AFM317xf5	0.69		231	253	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S667	AFM318zh5	0.66		285	315	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S668	AFM319yc1	0.66		255	275	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S669	AFM321yb5	0.57		190	210	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S670	AFM323we1	0.84		155	165	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S671	AFM324td5	0.64		197	211	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S672	AFM324wh1	0.85		172	178	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S673	AFM329tf5	0.67		260	288	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S674	AFM331ze9	0.86		266	276	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S675	AFM336tc1	0.82		155	173	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S676	AFM347yg9	0.67		229	239	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S677	AFM350xh1	0.46		214	222	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S678	AFMa139ya9	0.71		252	264	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	D5S683	JS1	0.91		119	157	GDB
05	D5S76	NA	0.77		97	109	GDB
05	D5S804	GATA5G04	0.81		241	241	GDB
05	D5S805	GATA5D11	0.73		156	156	GDB
05	D5S806	GATA5E10	0.67		196	196	GDB
05	D5S807	GATA3A04	0.76		161	161	GDB
05	D5S808	MFD213	0.74	0.70	116	132	GENOMICS 8:400- , 1990
05	D5S809	MFD247	0.52	0.49	89	107	GENOMICS 8:400- , 1990
05	D5S810	MFD317	0.64	0.57	192	200	GENOMICS 8:400- , 1990
05	D5S811	MFD343	0.79	0.76	216	248	GENOMICS 8:400- , 1990
05	D5S812	GAAT1D8	0.62		154	162	GDB
05	D5S813	GATA11G08	0.84		228	240	GDB
05	D5S814	GATA12A08	0.75		174	190	GDB
05	D5S815	GATA12G02	0.94		252	296	GDB
05	D5S816	GATA2H09	0.95		225	253	GDB
05	D5S817	GATA3E10	0.62		260	272	GDB
05	D5S818	GATA3F03	0.75		138	162	GDB
05	D5S819	GATA5C10	0.93		262	286	GDB
05	D5S82	NA	0.76	0.70	169	179	BREUKEL,C. ET AL.(1991) NAR 19,5804.
05	D5S820	GATA6E05	0.73		190	218	GDB
05	D5S821	GGAT3H04	0.50		162	170	GDB
05	D5S822	AFM224zh2	0.80		240	254	WEISSENBACH J: NATURE GENETIC, JUNE 1994
05	FBN2	NA	0.87		110	155	BIDDINGER,A. ET AL.(1993) HMG 2,1323.
05	FGFA	NA	0.88	0.61	245	259	GDB
05	GABRA1	NA	0.76	0.78	180	206	GDB
05	IG22	NA	0.79		142	159	GDB
05	IL9	NA	0.80	0.79	102	130	MOUT,R. ET AL. PERS. COMM. TO GDB.
05	IRF1	NA	0.74		174	184	HMG (1993) IN PRINT.
05	MCC	MBD	0.55		168	176	GDB
05	SPARC	SPARC	0.80		157	179	DRACOPOLI,N.& MEISLER,M.(1990) GENOMICS7,97-102.
05	TCOF1	IG52	0.89		196	236	DIXON,M. ET AL.(1992) HMG 1,249-253.
06	ACTBP2	NA	0.93	0.93	234	318	POLYMEROPOULOS,M. ET AL.(1992) NAR 20,1432.
06	ARG1	MFD 91	0.50		84	104	GENOMICS 1993, SUBMITTED
06	COL9A1	509-8B2	0.95		171	191	WARMAN, M. ET AL. (1993) GENOMICS 17, 694-698.
06	D1S1649	GATA30	0.71		N/A	N/A	GDB
06	D6S1003	ATA1F08	0.83		292	319	GDB
06	D6S1004	ATA1F12	0.56		217	217	GDB
06	D6S1005	ATA2C11	0.54		178	187	GDB
06	D6S1006	ATC4D09	0.56		194	203	GDB
06	D6S1007	GATA22G09	0.60		283	303	GDB
06	D6S1008	GATA31F06	0.77		246	246	GDB
06	D6S1009	GATA32B03	0.69		242	242	GDB
06	D6S1010	GATA41E03	0.75		228	228	GDB
06	D6S1011	GATA46H02	0.93		160	160	GDB

CH	LOCUS	ASSAY	HET	PIC	SIZE ANGE MIN	MAX	REFERENCE
06	D6S1012	GATA49A10	0.57		311	311	GDB
06	D6S1013	GATA52B08	0.60		167	167	GDB
06	D6S1014	GCT4B05	0.64		139	139	GDB
06	D6S1015	GCT5E07	0.48		179	179	GDB
06	D6S1016	GGAA10G12	0.67		232	256	GDB
06	D6S1017	GGAT3H10	0.69		151	171	GDB
06	D6S1018	GGAT4C01	0.45		148	156	GDB
06	D6S1019	GTAT1H06	0.70		212	234	GDB
06	D6S105	MFD 61	0.87	0.77	116	138	NAR 19:(4):968, 1991
06	D6S109	NA	0.78	0.78	169	193	RANUM L.P.W. ET AL.(1991)NAR 19,1171.
06	D6S202	NA	0.68		130	154	LE BORGNE-DEMARQUOY F.ET AL.(1991)NAR 19,6060.
06	D6S220	MIT-G119	0.68		175	175	HUMAN GENET 87:401, 1991
06	D6S223	NA	0.79		185	201	BOWCOCK,A. ET AL.(1992) HMG 1,66.
06	D6S224	MIT-MS135	0.59		245	245	HUMAN GENET 87:401, 1991
06	D6S225	MIT-E116	0.79		132	132	HUMAN GENET 87:401, 1991
06	D6S226	MIT-MS236	0.70		206	206	HUMAN GENET 87:401, 1991
06	D6S238	NA	0.64		388	404	GDB
06	D6S239	NA	0.78		162	176	GDB
06	D6S243	NA	0.53		170	250	GDB
06	D6S244	NA	0.74		362	368	GDB
06	D6S246	NA	0.68		220	230	GDB
06	D6S248	NA	0.80		269	287	GDB
06	D6S249	MFD 97	0.50		146	164	WILKIE,P. ET AL.(1993) GENOMICS 15,225-227.
06	D6S250	MFD118	0.80		150	174	GENOMICS 1993, SUBMITTED
06	D6S251	MFD131	0.86		144	162	GENOMICS 1993, SUBMITTED
06	D6S252	MFD171	0.70	0.64	142	168	GENOMICS 1993, SUBMITTED
06	D6S253	MFD181	0.70	0.55	267	291	GENOMICS 15:225-227, 1993
06	D6S254	MFD183	0.70		250	276	GENOMICS 1993, SUBMITTED
06	D6S255	MFD226	0.74	0.71	163	175	GENOMICS 1993, SUBMITTED
06	D6S257	AFM025te5	0.88		164	186	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S258	AFM031yh12	0.81		189	207	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S259	AFM035wc1	0.74		267	285	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S260	AFM056xe1	0.85		155	189	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S261	AFM059xh8	0.83		101	129	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S262	AFM059yd6	0.84		167	183	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S263	AFM066xf7	0.82		90	114	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S264	AFM079zb7	0.71		108	122	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S265	AFM101xa1	0.79		122	138	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S266	AFM102xf12	0.64		268	284	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S267	AFM114xd12	0.76		235	245	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S268	AFM115xh2	0.75		86	100	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S269	AFM123xe1	0.00		178	192	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S270	AFM127xb2	0.77		141	157	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S271	AFM136yf8	0.87		166	208	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S272	AFM142xe7	0.73		180	196	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S273	AFM142xh6	0.77		130	140	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S274	AFM144yf2	0.00		171	193	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S275	AFM158ya11	0.74		207	219	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S276	AFM158ye9	0.84		198	226	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S277	AFM158yh2	0.80		98	120	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S278	AFM162xc3	0.66		125	139	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S279	AFM163xa1	0.81		279	307	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S280	AFM168xh10	0.69		150	164	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S281	AFM176xh8	0.68		203	219	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S282	AFM184xa11	0.88		108	126	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S283	AFM190yf10	0.85		255	291	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S284	AFM191xa3	0.73		233	251	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S285	AFM192yf2	0.77		207	221	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S286	AFM198yc11	0.79		206	232	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S287	AFM198ze1	0.86		143	171	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S288	AFM199ye5	0.64		232	239	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S289	AFM200wc9	0.80		215	227	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S290	AFM200yb6	0.71		253	263	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S291	AFM203yg7	0.73		198	210	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S292	AFM203za9	0.83		141	161	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S294	AFM205yc7	0.83		86	108	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S295	AFM206xc11	0.74		93	107	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S296	AFM207xh2	0.79		260	300	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S297	AFM212yf6	0.67		210	224	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S299	AFM217xg7	0.00		206	226	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801

CH	LOCUS	ASSAY	HET	PIC	SIZE MIN	ANGE MAX	REFERENCE
06	D6S300	AFM218xg1	0.77		189	207	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S301	AFM220zf6	0.77		221	251	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S302	AFM224yb4	0.72		185	207	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S303	AFM225ya11	0.70		225	235	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S304	AFM238zd4	0.77		228	250	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S305	AFM242zg5	0.84		204	230	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S306	AFM248xh1	0.65		230	248	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S308	AFM262xe9	0.75		193	203	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S309	AFM265zh9	0.85		254	272	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S310	AFM267zg5	0.80		159	183	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S311	AFM276xf1	0.92		230	276	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S313	AFM191xd6	0.68		279	285	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S314	AFM254xh1	0.81		243	259	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S334	NA	0.89		178	190	GDB
06	D6S344	AFM092xb7	0.74		139	159	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
06	D6S348	MK6A	0.75		142	168	KANSARA, M. & ATHWAL, R. (1993) HMG 2, 1085.
06	D6S355	NA	0.65		348	362	GDB
06	D6S357	NA	0.80		264	272	GDB
06	D6S359	NA	0.78		181	193	GDB
06	D6S361	MFD284	0.00		125	135	GENOMICS 8:400- , 1990
06	D6S362	MFD312	0.72	0.70	174	194	GENOMICS 8:400- , 1990
06	D6S363	MFD314	0.63	0.57	110	130	GENOMICS 8:400- , 1990
06	D6S366	NA	0.82		138	162	GDB
06	D6S402	AFM190ye1	0.77		108	126	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S403	AFM190yg1	0.74		223	237	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S404	AFM190yg5	0.79		197	217	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S405	AFM036ye1	0.74		216	230	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S406	AFM038xc3	0.22		186	206	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S407	AFM198wg11	0.80		174	208	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S408	AFM199zh10	0.46		177	181	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S409	AFM200we11	0.78		164	180	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S410	AFM203xe11	0.72		243	251	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S411	AFM207wa1	0.78		151	159	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S412	AFM207xb6	0.74		192	208	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S413	AFM210vf8	0.71		184	196	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S414	AFM211xa11	0.64		104	108	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S415	AFM211yb10	0.66		261	277	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S416	AFM211ze5	0.77		255	263	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S417	AFM212zf10	0.33		237	253	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S418	AFM067xh3	0.67		162	176	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S419	AFM218xb6	0.72		129	141	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S420	AFM220yc9	0.85		180	192	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S421	AFM220zf2	0.65		103	117	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S422	AFM234xa3	0.67		191	211	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S423	AFM234xd8	0.58		197	211	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S424	AFM234ya7	0.66		247	263	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S425	AFM238yh10	0.55		142	152	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S426	AFM238zf4	0.58		203	221	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S427	AFM079xa5	0.70		182	200	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S428	AFM240yd6	0.48		229	247	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S429	AFM242za5	0.72		222	238	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S430	AFM254vb1	0.82		207	244	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S433	AFM260zb5	0.70		288	312	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S434	AFM123ya7	0.84		193	234	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S435	AFM135xh2	0.55		134	198	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S436	AFM029yd4	0.77		183	193	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S437	AFM266yb5	0.74		129	163	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S438	AFM268wa9	0.88		178	196	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S439	AFM268xe1	0.74		272	292	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S440	AFM268yf5	0.90		269	285	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S441	AFM269ze1	0.69		162	186	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S442	AFM277vf1	0.78		259	271	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S443	AFM277wb5	0.71		149	161	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S444	AFM278yd1	0.75		137	141	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S445	AFM286za5	0.62		205	233	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S446	AFM290xf5	0.79		201	211	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S447	AFM290zd9	0.68		196	212	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S448	AFM292yd5	0.77		169	183	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S449	AFM296ze5	0.80		193	199	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S450	AFM297zd5	0.42		227	239	WEISSENBACH J: NATURE GENETIC, JUNE 1994

CH	LOCUS	ASSAY	HET	PIC	SIZE MIN	ANGE MAX	REFERENCE
06	D6S451	AFM298we1	0.71		128	138	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S452	AFM301wb9	0.70		264	280	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S453	AFM301ya5	0.68		180	192	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S454	AFM301zb1	0.78		132	140	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S455	AFM303zd9	0.86		151	165	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S456	AFM308wf1	0.65		232	248	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S457	AFM310wg1	0.76		197	207	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S458	AFM311wa5	0.71		224	230	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S459	AFM312xc5	0.90		98	114	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S460	AFM312yb9	0.74		144	166	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S461	AFM316zg5	0.95		246	268	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S462	AFM317zb1	0.69		277	291	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S463	AFM319vh5	0.79		131	133	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S464	AFM323vb5	0.59		202	226	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S465	AFM323xd5	0.87		184	198	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S466	AFM331vh1	0.64		107	119	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S467	AFM331yh1	0.67		246	250	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S468	AFM345vd9	0.87		148	162	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S469	AFM347yf1	0.51		170	190	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S470	AFM349xh5	0.52		120	134	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S471	AFMa123xh1	0.59		108	116	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S472	AFMa128yd9	0.81		173	189	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S473	AFMa133xa5	0.57		166	196	WEISSENBACH J: NATURE GENETIC, JUNE 1994
06	D6S474	GATA31	0.64		165	170	GDB
06	D6S475	GATA44	0.89		139	167	GDB
06	D6S476	GATA2E07	0.50		187	187	GDB
06	D6S477	GATA3H05	0.90		239	239	GDB
06	D6S478	GATA4A03	1.00		287	287	GDB
06	D6S482	MFD348	0.72	0.67	113	125	GENOMICS 8:400- , 1990
06	D6S492	GATA108	0.55		200	212	GDB
06	D6S493	GATA11B08	0.71		236	240	GDB
06	D6S494	GATA2B06	0.43		169	193	GDB
06	D6S495	GATA2E12	0.65		177	183	GDB
06	D6S497	GATA14A08	0.89		260	260	GDB
06	D6S499	GATA7A07	0.00		183	183	GDB
06	D6S500	GATA7B06	0.00		286	286	GDB
06	D6S501	GATA5C03	0.75		166	182	GDB
06	D6S502	GATA7B06	0.85		N/A	N/A	GDB
06	D6S503	GGAA8D08	0.00		248	260	GDB
06	D6S510	NA	0.84		178	196	GDB
06	D6S87	MFD 47	0.60	0.53	137	155	NAR 18(15):4636, 1990
06	D6S89	NA	0.92	0.88	199	227	LITT,M.&LUTY,J.(1990) NAR 18,4301.
06	DHFRP2	NA	0.70	0.66	157	173	POLYMEROPOULOS,M. ET AL.(1991) NAR 19,7198.
06	EDN1	NA	0.78		197	217	PAGES,J. ET AL. (1993) HMG 2,90.
06	F13A1	NA	0.78	0.75	180	230	POLYMEROPOULOS,M. ET AL.(1991) NAR 19,4306.
06	FTHPI	NA	0.90		171	181	MAUVIEUX,V. ET AL(1991)NAR 19,6969.
06	IGF2R	MFD 56	0.70		99	101	GENOMICS 8:400- , 1990
06	IGF2R-3	NA	0.58		164	164	GDB
06	IGF2R-II	NA	0.00	0.41	158	166	GOTO,J. ET AL.(1992) NAR 20,923.
06	IID	NA	0.81	0.78	185	206	POLYMEROPOULOS,M. ET AL.(1991) NAR 19,4307.
06	RDS	NA	0.00	0.57	96	98	GDB
06	SCA1	CAGa/b	0.84		119	287	GDB
06	TCTE1	NA	83.30		171	171	KWIATKOWSKI,T. ET AL.(1991) GENOMICS10,921-26.
07	CD36	13-GT	0.84		68	80	GDB
07	CFTR	17B	0.89		414	414	ZIELENSKI, J (1991) AM. J. HUM. GENET. 49, 1256-1262.
07	COL1A2	NA	0.73		100	200	GDB
07	D7S1480	NA	0.74		98	120	VELASCO, E. ET AL. (1994) HMG 3, 1441.
07	D7S1789	ACT3E08	0.69		128	149	GDB
07	D7S1790	ATA1A10	0.31		242	257	GDB
07	D7S1791	ATA1B04	0.81		160	184	GDB
07	D7S1792	ATA3A10	0.31		206	206	GDB
07	D7S1793	ATA5H04	0.76		230	248	GDB
07	D7S1794	CTT8	0.29		221	221	GDB
07	D7S1795	GATA21B01	0.60		195	211	GDB
07	D7S1796	GATA21C11	0.86		288	304	GDB
07	D7S1797	GATA21D08	0.60		220	244	GDB
07	D7S1798	GATA23F01	0.60		245	257	GDB
07	D7S1799	GATA23F05	0.76		171	199	GDB
07	D7S1800	GATA2G07	0.50		193	201	GDB
07	D7S1801	GATA31D01	0.75		227	227	GDB

CH	LOCUS	ASSAY	HET	PIC	SIZE MIN	ANGE MAX	REFERENCE
07	D7S1802	GATA41G07	0.73		187	187	GDB
07	D7S1803	GATA42A01	0.88		381	381	GDB
07	D7S1804	GATA43C11	0.86		258	258	GDB
07	D7S1805	GATA4H10	0.92		198	223	GDB
07	D7S1806	GGAA11C11	0.40		195	195	GDB
07	D7S1807	GGAA2B12	0.78		286	286	GDB
07	D7S1808	GGAA3F06	0.81		252	276	GDB
07	D7S1809	GGAA9C07	0.76		200	228	GDB
07	D7S1810	GGAT2B11	0.56		226	226	GDB
07	D7S1830	GATA4E04	0.82		200	224	GDB
07	D7S1843	GTAT1A10	0.00				GDB
07	D7S23	NA	0.82		109	127	RICHARDS,B. ET AL.(1991) NAR 19,5798.
07	D7S435	MFD 20	0.59	0.53	122	134	NAR 18(13):4039, 1990
07	D7S440	MFD 50	0.75	0.70	169	191	NAR 18(15):4636, 1990
07	D7S460	MIT-MH26	0.95		180	196	HUMAN GENET 87:401, 1991
07	D7S461	MIT-MS97	0.87		177	177	HUMAN GENET 87:401, 1991
07	D7S462	MIT-MS262	0.52		150	150	HUMAN GENET 87:401, 1991
07	D7S463	MIT-G111	0.70		159	159	HUMAN GENET 87:401, 1991
07	D7S466	MIT-COS43	0.83		244	244	HUMAN GENET 87:401, 1991
07	D7S471	MFD123	0.80		181	199	HAUGE,X. ET AL.(1991) NAR 19,4308.
07	D7S472	MFD172	0.70		114	129	HAUGE,X. ET AL.(1991) NAR 19,4308.
07	D7S473	MFD148	0.85		126	148	H DONIS-KELLER ET AL, J. WEBER
07	D7S474	MFD107	0.80		120	144	GDB
07	D7S476	NA	0.79		186	210	XIAO,H. ET AL.(1992) HMG 1,549.
07	D7S477	AFM030xb4	0.71		175	185	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S478	AFM032xa1	0.70		118	130	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S479	AFM036xg5	0.84		105	135	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S480	AFM042xh10	0.87		189	206	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S481	AFM049xe3	0.85		186	204	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S482	AFM070yc1	0.74		166	198	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S483	AFM074xg5	0.83		166	188	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S484	AFM087yd11	0.75		99	113	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S485	AFM095xe9	0.79		244	256	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S486	AFM098xg9	0.81		133	146	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S487	AFM107yb6	0.75		174	188	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S488	AFM113xc11	0.85		136	156	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S489	AFM136xe3	0.38		140	144	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S490	AFM150yg7	0.79		92	106	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S491	AFM151xf10	0.75		115	131	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S492	AFM158xa1	0.78		145	155	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S493	AFM162xa7	0.89		194	224	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S494	AFM165zf4	0.79		173	191	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S495	AFM168xc3	0.82		150	168	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S496	AFM172xa1	0.76		129	141	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S497	AFM177xf10	0.53		101	111	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S498	AFM183ya3	0.63		137	153	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S499	AFM191xh6	0.84		236	252	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S500	AFM198zh8	0.88		188	210	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S501	AFM199vb2	0.82		163	179	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S502	AFM199vh8	0.85		273	291	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S503	AFM199xc3	0.88		148	180	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S504	AFM199xh12	0.80		145	159	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S505	AFM199zd4	0.70		262	278	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S506	AFM200wc7	0.88		117	143	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S507	AFM200we7	0.90		148	168	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S509	AFM203wg1	0.73		203	225	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S510	AFM207wb2	0.80		252	264	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S511	AFM210xc7	0.80		207	225	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S512	AFM214yb2	0.72		160	190	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S513	AFM217yc5	0.84		173	201	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S514	AFM218xf10	0.72		147	157	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S515	AFM220xc11	0.82		128	190	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S516	AFM224xg5	0.76		254	266	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S517	AFM225xa1	0.84		239	257	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S518	AFM225xg9	0.88		179	201	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S519	AFM238vb12	0.82		256	268	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S520	AFM240ve9	0.70		79	97	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S521	AFM240yh4	0.71		285	303	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S522	AFM242yc3	0.67		217	229	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S523	AFM242ye3	0.81		224	240	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801

CH	LOCUS	ASSAY	HET	PIC	SIZE ANGE MIN	MAX	REFERENCE
07	D7S524	AFM248ta5	0.75		234	246	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S525	AFM248tc5	0.66		219	235	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S526	AFM248vc9	0.72		125	135	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S527	AFM248vd9	0.76		273	297	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S528	AFM248ve5	0.73		108	116	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S529	AFM248zd1	0.70		218	226	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S530	AFM249xf9	0.79		106	255	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S531	AFM254yc9	0.77		225	255	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S547	NA	0.94		117	135	GREGG,R.& PARKER,M.(1992)HMG 1,659.
07	D7S550	AFM224xh4	0.83		177	200	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
07	D7S555	MFD253	0.80	0.77	110	134	GENOMICS 8:400- , 1990
07	D7S558	MFD267	0.58	0.50	95	119	GENOMICS 8:400- , 1990
07	D7S559	MFD265	0.81	0.78	196	216	GENOMICS 8:400- , 1990
07	D7S594	sAVH-6	0.84	0.82	217	235	HING,A. ET AL.(1993) AM.J.HUM.GENET. 53,509-517.
07	D7S629	AFM165xb10	0.76		249	263	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S630	AFM165yh12	0.77		198	222	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S631	AFM183xe11	0.89		108	124	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S632	AFM198ze5	0.65		209	221	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S634	AFM203vb6	0.73		136	148	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S635	AFM206xc1	0.56		216	234	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S636	AFM207za9	0.61		130	168	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S637	AFM211xc3	0.85		222	232	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S638	AFM217yb6	0.76		194	208	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S639	AFM220ya3	0.88		260	275	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S640	AFM220yg1	0.69		114	144	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S641	AFM224yb6	0.69		84	100	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S642	AFM074wc12	0.54		191	207	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S644	AFM234xc7	0.64		194	206	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S645	AFM238zc9	0.68		197	215	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S646	AFM240ve3	0.33		179	201	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S647	AFM240vh4	0.43		143	175	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S648	AFM078ze1	0.65		194	204	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S649	AFM240xe9	0.84		275	281	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S650	AFM240zh10	0.57		265	289	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S651	AFM249za5	0.88		173	191	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S652	AFM254xd5	0.92		269	281	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S653	AFM259zc1	0.51		201	229	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S654	AFM102xg7	0.63		207	227	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S655	AFM263wg9	0.76		251	269	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S656	AFM263xe9	0.58		243	275	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S657	AFM263yd9	0.61		246	264	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S658	AFM269zg1	0.88		264	272	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S659	AFM276yg1	0.69		191	215	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S660	AFM277vd5	0.79		189	197	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S661	AFM277zf5	0.77		252	282	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S662	AFM280vh9	0.76		204	234	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S663	AFM280zc5	0.71		153	173	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S664	AFM281vc9	0.76		203	215	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S665	AFM283xc5	0.66		204	224	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S666	AFM283za9	0.80		155	169	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S667	AFM284xg5	0.79		116	144	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S668	AFM284zf9	0.82		257	275	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S669	AFM286xf9	0.70		123	139	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S670	AFM288vb5	0.77		100	110	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S671	AFM288yg9	0.67		136	158	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S672	AFM289ve9	0.55		132	160	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S673	AFM290vg9	0.80		118	148	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S674	AFM294wf1	0.86		139	155	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S675	AFM295yg9	0.78		201	209	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S676	AFM302za5	0.63		148	166	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S677	AFM303vh9	0.69		275	295	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S678	AFM308wc5	0.78		166	180	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S679	AFM308zg1	0.77		140	164	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S680	AFM309yf1	0.80		119	131	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S681	AFM310yf9	0.71		249	261	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S682	AFM311xc5	0.73		271	283	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S683	AFM311zc5	0.60		258	284	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S684	AFM312wb5	0.67		169	187	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S685	AFM317yc5	0.73		178	192	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S686	AFM323wd5	0.45		254	266	WEISSENBACH J: NATURE GENETIC, JUNE 1994

CH	LOCUS	ASSAY	HET	PIC	SIZE ANGE MIN	MAX	REFERENCE
07	D7S687	AFM323yg5	0.69		238	244	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S688	AFM324zf9	0.61		147	161	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S689	AFM333wf5	0.86		125	135	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S690	AFM338wh1	0.71		264	274	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S691	AFM350va9	0.73		128	146	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S692	AFM357te1	0.64		161	171	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S793	GATA3B01	0.86		146	154	GDB
07	D7S794	GATA2C04	0.79		168	168	GDB
07	D7S795	GATA4B03	0.67		230	230	GDB
07	D7S796	GATA4E02	0.91		162	198	GDB
07	D7S798	AFM205va3	0.55		200	218	WEISSENBACH J: NATURE GENETIC, JUNE 1994
07	D7S799	MFD225	0.87	0.86	83	113	GENOMICS 8:400- , 1990
07	D7S800	MFD327	0.52	0.47	158	176	GENOMICS 8:400- , 1990
07	D7S801	MFD329	0.82	0.80	177	201	GENOMICS 8:400- , 1990
07	D7S802	MFD340	0.82	0.80	220	250	GENOMICS 8:400- , 1990
07	D7S803	MFD358	0.85	0.84	126	152	GENOMICS 8:400- , 1990
07	D7S804	PY5-18	0.78		123	141	GDB
07	D7S808	wg1a2	0.79		435	480	ARMOUR, J. ET AL. (1994) HMG 3, 599-605.
07	D7S809	wg1g9	0.86		241	289	ARMOUR, J. ET AL. (1994) HMG 3, 599-605.
07	D7S813	MS8-170	0.73		96	110	GDB
07	D7S814	ATC6	0.54		105	105	GDB
07	D7S815	GATA2G04	0.50		149	169	GDB
07	D7S817	GATA13G11	0.72		157	177	GDB
07	D7S818	GATA6G06	0.00		155	155	GDB
07	D7S820	GATA3F01	0.86		204	240	GDB
07	D7S821	GATA5D08	0.62		238	270	GDB
07	D7S822	wg1e12	0.82		266	298	ARMOUR, J. ET AL. (1994) HMG 3, 599-605.
07	EGFR	NA	0.72	0.67	114	128	CHI, D. ET AL.(1992) HMG 1,135.
07	ELN	NA	0.00		100	200	GDB
07	TCRB	VB6.7	0.83		127	153	GDB
08	CRH	NA	0.72		128	156	GU,J. ET AL.(1993) HMG 2,85.
08	D6S502	GATA7G07	0.82		162	194	GDB
08	D8S1098	ATA1G10	0.67		134	143	GDB
08	D8S1099	ATA3A02	0.59		249	262	GDB
08	D8S1100	ATC2D12	0.56		183	192	GDB
08	D8S1101	ATC2F06	0.40		165	168	GDB
08	D8S1102	GAAT1C11	0.60		184	200	GDB
08	D8S1103	GAAT2A02	0.31		313	313	GDB
08	D8S1104	GAAT2F03	0.80		129	145	GDB
08	D8S1105	GATA23C09	0.63		241	257	GDB
08	D8S1106	GATA23D06	0.80		127	151	GDB
08	D8S1107	GATA29D08	0.69		296	296	GDB
08	D8S1108	GATA50D10	0.71		256	256	GDB
08	D8S1109	GATA52F11	0.92		240	240	GDB
08	D8S1110	GATA8G10	0.85		262	286	GDB
08	D8S1111	GGAA2H06	0.00		146	146	GDB
08	D8S1112	GGAA8A04	0.62		206	234	GDB
08	D8S1113	GGAA8G07	0.77		217	245	GDB
08	D8S133	c24E10	0.77	0.74	94	112	WOOD,S.&SCHERTZER,M.(1992)GENOMICS13,232.
08	D8S136	NA	0.88		71	85	GDB
08	D8S137	NA	0.67	0.61	150	162	WOOD,S.& SCHERTZER,M.(1991)NAR19,6664.
08	D8S161	NA	0.75	0.77	151	165	COUCH,F. ET AL.(1991) NAR 19,5093.
08	D8S164	MFD104	0.86	0.79	165	199	GENOMICS 14:144-152, 1992
08	D8S165	MFD117	0.54	0.50	138	152	GENOMICS 15:225-227, 1993
08	D8S166	MFD159	0.88	0.83	110	132	GENOMICS 15:225-227, 1993
08	D8S167	MFD185	0.84	0.86	105	135	GENOMICS 15:225-227, 1993
08	D8S198	MFD169	0.83	0.81	155	173	GENOMICS 15:225-227, 1993
08	D8S199	MFD177	0.83	0.81	204	230	GENOMICS 15:225-227, 1993
08	D8S200	MFD196	0.76	0.67	184	196	GENOMICS 15:225-227, 1993
08	D8S201	MFD199	0.92	0.89	178	178	GENOMICS 15:225-227, 1993
08	D8S205	MIT-MS45	0.78		108	108	HUMAN GENET 87:401, 1991
08	D8S206	MIT-MS61	0.67		127	127	HUMAN GENET 87:401, 1991
08	D8S207	MIT-MS142	0.74		142	142	HUMAN GENET 87:401, 1991
08	D8S208	MIT-MS91	0.75		155	155	HUMAN GENET 87:401, 1991
08	D8S251	MFD229	0.83	0.80	153	181	GENOMICS 8:400- , 1990
08	D8S254	MFD210	0.58	0.49	65	75	GENOMICS 8:400- , 1990
08	D8S255	AFM023xc1	0.74		107	129	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
08	D8S256	AFM073yb7	0.84		210	232	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
08	D8S257	AFM077ya5	0.73		106	122	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
08	D8S258	AFM107xb6	0.71		144	154	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801

CH	LOCUS	ASSAY	HET	PIC	SIZE RANGE MIN	SIZE RANGE MAX	REFERENCE
08	D8S259	AFM107yb2	0.60		200	216	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
08	D8S260	AFM114xe7	0.83		187	213	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
08	D8S261	AFM123xg5	0.78		128	144	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
08	D8S262	AFM127xh2	0.72		114	128	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
08	D8S263	AFM141xa5	0.76		275	289	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
08	D8S264	AFM143xd8	0.85		121	145	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
08	D8S265	AFM144zb2	0.79		208	231	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
08	D8S266	AFM151ye3	0.53		153	165	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
08	D8S268	AFM156xa3	0.61		255	265	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
08	D8S269	AFM156xc3	0.49		215	221	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
08	D8S270	AFM165xh4	0.80		181	195	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
08	D8S271	AFM165yb10	0.78		192	271	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
08	D8S272	AFM175xb4	0.82		192	239	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
08	D8S273	AFM179yf6	0.81		127	139	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
08	D8S274	AFM182xa3	0.78		108	118	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
08	D8S275	AFM185xe9	0.76		139	157	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
08	D8S276	AFM192xc5	0.66		239	245	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
08	D8S277	AFM198wd2	0.74		148	180	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
08	D8S278	AFM200ye1	0.65		232	240	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
08	D8S279	AFM203wc1	0.88		229	273	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
08	D8S280	AFM205wc5	0.54		224	232	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
08	D8S281	AFM205yh4	0.65		112	124	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
08	D8S282	AFM234vf4	0.73		260	272	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
08	D8S283	AFM238yh12	0.80		103	129	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
08	D8S284	AFM248td9	0.84		243	273	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
08	D8S285	AFM255yb9	0.79		108	124	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
08	D8S286	AFM268ve9	0.82		220	238	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
08	D8S296E	FB12B7	0.81		102	114	KHAN, A. ET AL. (1992) NATURE GENETICS 2, 180-185.
08	D8S298	AFM234yh10	0.70		155	167	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
08	D8S306	KW97	0.78		244	297	NELSON, L. ET AL. (1994) HMG 3, 1209.
08	D8S307	NA	0.80		337	418	WARD,K. ET AL.(1993) HMG 2,615.
08	D8S315	KW328	0.73		336	384	GDB
08	D8S320	KW205	0.84		386	420	GDB
08	D8S322	KW218	0.76		210	238	GDB
08	D8S339	WT251	0.30		162	176	GDB
08	D8S344	KW371	0.75		292	328	WARD, K. ET AL. (1993) HMG 2, 1087.
08	D8S346	KW400	0.63		232	252	GDB
08	D8S347	KW426	0.79		322	382	GDB
08	D8S350	MFD287	0.45	0.36	158	164	GENOMICS 8:400- , 1990
08	D8S351	MFD295	0.84	0.82	109	127	GENOMICS 8:400- , 1990
08	D8S373	UT721	0.80		194	218	GDB
08	D8S439	MFD280	0.84	0.82	115	133	GENOMICS 8:400- , 1990
08	D8S474	KW401	0.71		355	373	GDB
08	D8S501	AFM172xf10	0.72		197	217	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S502	AFM182xh12	0.52		211	221	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S503	AFM193xh4	0.84		212	226	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S504	AFM197xg5	0.57		193	203	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S505	AFM198tb2	0.45		203	213	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S506	AFM200wb6	0.78		121	127	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S507	AFM203wh10	0.63		155	169	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S508	AFM203xd10	0.62		214	232	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S509	AFM203zc1	0.73		269	277	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S510	AFM203zg1	0.68		217	225	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S511	AFM205tb10	0.76		133	143	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S513	AFM211wa9	0.52		194	206	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S514	AFM212we1	0.78		209	227	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S515	AFM214ya7	0.74		268	276	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S516	AFM214zb4	0.74		158	168	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S517	AFM224xh6	0.42		212	226	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S518	AFM234ve1	0.69		229	253	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S519	AFM234vh8	0.61		116	130	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S520	AFM234ye5	0.68		179	199	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S521	AFM078za9	0.60		118	128	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S522	AFM240wc9	0.35		203	217	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S523	AFM240zb4	0.67		243	257	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S524	AFM242zf6	0.73		248	272	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S525	AFM244td1	0.81		149	267	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S526	AFM255yb1	0.81		233	239	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S527	AFM259vb1	0.65		272	284	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S528	AFM080xh7	0.65		166	180	WEISSENBACH J: NATURE GENETIC, JUNE 1994

CH	LOCUS	ASSAY	HET	PIC	SIZE MIN	ANGE MAX	REFERENCE
08	D8S529	AFM259xc5	0.74		244	262	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S530	AFM259yg5	0.74		201	227	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S531	AFM081xe9	0.84		114	126	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S532	AFM081yd11	0.58		239	255	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S533	AFM088xb3	0.61		159	169	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S534	AFM144xh4	0.74		176	210	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S535	AFM265zc9	0.81		168	180	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S537	AFM269zh1	0.23		146	176	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S538	AFM274ve5	0.53		169	181	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S539	AFM277yc1	0.58		190	200	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S540	AFM281yb9	0.57		112	118	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S541	AFM286xf1	0.68		194	228	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S542	AFM287we5	0.49		229	249	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S543	AFM289ye5	0.65		116	140	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S544	AFM294vd5	0.85		132	140	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S545	AFM294we9	0.67		183	195	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S546	AFM295xh5	0.67		162	178	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S547	AFM295xh9	0.80		185	195	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S548	AFM301zc9	0.38		228	248	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S549	AFM303zc1	0.75		166	172	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S550	AFM304ze9	0.77		249	277	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S551	AFM308ze5	0.81		253	285	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S552	AFM320yf1	0.80		168	182	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S553	AFM326te5	0.67		221	253	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S554	AFM337wg5	0.71		161	177	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S555	AFM337xe1	0.43		165	177	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S556	AFM343xa5	0.64		161	175	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S557	AFM347zg1	0.64		233	251	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S558	AFM351zh1	0.71		160	180	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S559	AFM352td9	0.51		218	230	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S560	AFMa127ye5	0.79		131	156	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S561	AFMa132xe5	0.83		172	182	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S565	GATA8G03	0.67		252	252	GDB
08	D8S566	GATA8H08	0.87		152	152	GDB
08	D8S568	GATA2E01	0.83		255	259	GDB
08	D8S569	GATA3H11	0.76		179	179	GDB
08	D8S570	MFD311	0.63	0.60	153	167	GENOMICS 8:400- , 1990
08	D8S571	MFD333	0.10	0.10	136	158	GENOMICS 8:400- , 1990
08	D8S572	MFD352	0.57	0.48	143	151	GENOMICS 8:400- , 1990
08	D8S580	wg1c4	0.90		187	272	ARMOUR, J. ET AL. (1994) HMG 3, 599-605.
08	D8S586	GATA11E08	0.86		244	244	GDB
08	D8S587	GATA11G11	0.81		154	186	GDB
08	D8S588	GATA12F02	0.75		177	197	GDB
08	D8S589	GATA12H01	0.50		137	149	GDB
08	D8S591	GATA4E09	0.84		150	162	GDB
08	D8S592	GATA6B02	0.62		150	162	GDB
08	D8S593	GATA6F11	0.94		194	194	GDB
08	D8S594	GATA8A03	0.75		201	204	GDB
08	D8S601	AFM276xe9	0.68		223	236	WEISSENBACH J: NATURE GENETIC, JUNE 1994
08	D8S602	NA	0.80		289	315	GDB
08	D8S639	KW413	0.73		331	331	NELSON, L. ET AL. (1994) HMG 3, 1209.
08	D8S640	KW306	0.75		448	480	LU, J. ET AL. (1994) HMG 3, 839.
08	D8S84	MFD 8	0.58	0.58	181	195	GENOMICS 15:225-227, 1993
08	D8S85	MFD 18	0.79	0.69	73	83	GENOMICS 15:225-227, 1993
08	D8S87	MFD 39	0.71	0.65	145	157	GENOMICS 15:225-227, 1993
08	D8S88	MFD 45	0.88	0.82	76	100	GENOMICS 15:225-227, 1993
08	FGFR1	FGFR1CA4	0.69		159	173	YU, C. ET AL. (1994) HMG 3, 212.
08	LPL	c114C11	0.83	0.81	118	142	TOMFOHRDE,J. ET AL.(1992) GENOMICS14,144-152.
08	LPL-TET	GZ14/15	0.71	0.66	119	135	ZULIANI,G.& HOBBS,H.(1990)NAR 18,4958.
08	MYC	PCR3	0.86	0.85	87	125	POLYMEROPOULOS,M. ET AL.(1992) HMG 1,65.
08	NEFL	NA	0.81	0.00	137	147	ROGAEV,E. ET AL.(1992) HMG 1,781.
08	PENK	MFD 31	0.60	0.43	75	83	NAR 14:144-152, 1992
08	PLAT	NA	0.77	0.75	105	149	DAIGER,SP.ET AL.(1991)NAR 19,6058.
08	PLAT2	NA	0.69		108	138	W. THOMAS & D. DRAYNA (1992) HMG 1, 138.
09	ABL1	NA	0.67	0.64	89	117	KWIATKOWSKI,D.J.(1991) NAR 19,967
09	ABO	NA	0.67	0.00	220	220	GDB
09	ASS	NA	0.64	0.00	121	147	YUILLE,M.A.R. ET AL.(1990)NAR 18,7472.
09	D9S103	MFD 77	0.58	0.46	89	97	GENOMICS 14:144-152, 1992
09	D9S104	MFD121	0.81	0.79	181	199	GENOMICS 14:144-152, 1992
09	D9S105	MFD178	0.77	0.78	187	203	GENOMICS 14:144-152, 1992

CH	LOCUS	ASSAY	HET	PIC	SIZE ANGE MIN	MAX	REFERENCE
09	D9S106	MFD189	0.74	0.75	99	111	GENOMICS 14:144-152, 1992
09	D9S109	NA	0.70		219	229	FURLONG,R. ET AL.(1992) NAR 20,925.
09	D9S112	NA	0.85		115	135	KWIATKOWSKI,D.&GUSELLA,J.(1992) NAR20,932.
09	D9S113	5B1	0.82		118	132	GDB
09	D9S114	5B11	0.79		93	111	GDB
09	D9S115	NA	0.78		115	141	KWIATKOWSKI,D.&GUSELLA,J.(1992) NAR20,930.
09	D9S116	NA	0.85		88	112	KWIATKOWSKI,D.&GUSELLA,J.(1992) NAR20,931.
09	D9S117	NA	0.78		106	120	KWIATKOWSKI,D.&GUSELLA,J.(1992) NAR20,933.
09	D9S118	NA	0.83		69	93	KWIATKOWSKI,D.&GUSELLA,J.(1992) NAR20,932.
09	D9S119	NA	0.52		130	138	KWIATKOWSKI,D.&GUSELLA,J.(1992) NAR20,934.
09	D9S12	PCR2	0.92		126	129	YUILLE,M. ET AL.(1992) HMG 1,351.
09	D9S120	NA	0.76		141	155	KWIATKOWSKI,D.&GUSELLA,J.(1992) NAR20,933.
09	D9S121	NA	0.79		126	142	KWIATKOWSKI,D.ET AL.(1992) GENOMICS 12,229-240.
09	D9S122	10G11	0.78		146	160	GDB
09	D9S123	NA	0.55		74	92	KWIATKOWSKI,D.&GUSELLA,J.(1992) NAR20,934.
09	D9S125	3AB12	0.85		113	155	GDB
09	D9S126	NA	0.68		238	248	GDB
09	D9S127	NA	0.72		149	159	LYALL,J. ET AL.(1992)NAR 20,925.
09	D9S129	MIT-MS47	0.67		135	135	HUMAN GENET 87:401, 1991
09	D9S12II	2635	0.83		67	91	KWIATKOWSKI,T. ET AL.(1991) GENOMICS10,921-26.
09	D9S130	MIT-E117	0.58		184	184	HUMAN GENET 87:401, 1991
09	D9S131	MIT-MS202	0.83		100	100	HUMAN GENET 87:401, 1991
09	D9S132	MIT-G115	0.75		156	156	HUMAN GENET 87:401, 1991
09	D9S133	MIT-MS67	0.63		150	150	HUMAN GENET 87:401, 1991
09	D9S135	MIT-MS93	0.63		99	99	HUMAN GENET 87:401, 1991
09	D9S143	9CMP3	0.54		111	123	GDB
09	D9S144	NA	0.72		137	155	FURLONG,R. ET AL.(1992) HMG 1,447.
09	D9S146	NA	0.74		80	104	FUTREAL,P.A.,ET AL(1992) HMG 1,66.
09	D9S147E	NA	0.78		189	201	POLYMEROPOULOS,M. ET AL.(1992) HMG 1,549.
09	D9S148	C10	0.61		99	123	GDB
09	D9S149	D3	0.88		146	176	GDB
09	D9S15	NA	0.74		197	205	KWIATKOWSKI,D. ET AL.(1992)GENOMICS12,229-240.
09	D9S150	B1	0.72		87	99	GDB
09	D9S151	NA	0.75		293	442	GDB
09	D9S152	AFM015ya5	0.84		120	136	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
09	D9S153	AFM025yb2	0.77		143	155	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
09	D9S154	AFM028ye5	0.85		139	171	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
09	D9S155	AFM042xh4	0.74		120	129	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
09	D9S156	AFM051xd6	0.80		133	155	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
09	D9S157	AFM067xd3	0.85		133	149	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
09	D9S158	AFM073yb11	0.70		213	231	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
09	D9S159	AFM077xa9	0.78		293	309	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
09	D9S160	AFM079ze1	0.64		136	146	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
09	D9S161	AFM087yd3	0.78		119	135	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
09	D9S162	AFM115yb4	0.75		172	196	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
09	D9S163	AFM120xf6	0.71		271	279	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
09	D9S164	AFM122xf4	0.80		187	199	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
09	D9S165	AFM136xc5	0.76		202	226	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
09	D9S166	AFM144zg7	0.82		233	261	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
09	D9S167	AFM157xb12	0.00		260	286	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
09	D9S168	AFM158xf12	0.76		227	275	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
09	D9S169	AFM164xg7	0.84		259	275	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
09	D9S170	AFM164ya11	0.75		108	126	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
09	D9S171	AFM186xc3	0.80		159	177	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
09	D9S172	AFM199xf10	0.53		291	305	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
09	D9S173	AFM200vd6	0.51		242	250	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
09	D9S174	AFM207xa1	0.67		147	159	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
09	D9S175	AFM224zh10	0.86		200	230	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
09	D9S176	AFM225xf10	0.82		129	147	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
09	D9S177	AFM234yc5	0.87		211	237	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
09	D9S178	AFM242xh6	0.68		93	99	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
09	D9S179	AFM248wf1	0.77		231	251	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
09	D9S180	AFM168xb6	0.64		220	230	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
09	D9S195	AFM193yg5	0.75		204	240	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
09	D9S196	AFM212yb4	0.65		254	260	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
09	D9S197	AFM238va7	0.68		199	215	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
09	D9S199	NA	0.75		144	164	GRIMSBY,J. ET AL.(1992)NAR 20,924.
09	D9S200	NA	0.83		107	127	GRAW,S.& KWIATKOWSKI,D.(1993) HMG2,614.
09	D9S205	MFD271	0.47	0.43	199	207	GENOMICS 8:400- , 1990
09	D9S256	AFM161xd6	0.67		166	178	WEISSENBACH J: NATURE GENETIC, JUNE 1994

CH	LOCUS	ASSAY	HET	PIC	SIZE MIN	ANGE MAX	REFERENCE
09	D9S257	AFM183xh10	0.33		259	285	WEISSENBACH J: NATURE GENETIC, JUNE 1994
09	D9S258	AFM185xe3	0.64		155	167	WEISSENBACH J: NATURE GENETIC, JUNE 1994
09	D9S259	AFM186xc7	0.54		130	144	WEISSENBACH J: NATURE GENETIC, JUNE 1994
09	D9S260	AFM206za9	0.75		182	186	WEISSENBACH J: NATURE GENETIC, JUNE 1994
09	D9S261	AFM210ze7	0.73		90	104	WEISSENBACH J: NATURE GENETIC, JUNE 1994
09	D9S262	AFM211wc9	0.63		153	165	WEISSENBACH J: NATURE GENETIC, JUNE 1994
09	D9S263	AFM212yg1	0.55		177	185	WEISSENBACH J: NATURE GENETIC, JUNE 1994
09	D9S264	AFM218xh10	0.79		153	171	WEISSENBACH J: NATURE GENETIC, JUNE 1994
09	D9S265	AFM220zd4	0.68		84	94	WEISSENBACH J: NATURE GENETIC, JUNE 1994
09	D9S266	AFM225yh2	0.53		240	268	WEISSENBACH J: NATURE GENETIC, JUNE 1994
09	D9S267	AFM248wc9	0.72		163	175	WEISSENBACH J: NATURE GENETIC, JUNE 1994
09	D9S268	AFM259ya5	0.82		240	250	WEISSENBACH J: NATURE GENETIC, JUNE 1994
09	D9S269	AFM261zh9	0.61		171	183	WEISSENBACH J: NATURE GENETIC, JUNE 1994
09	D9S270	AFM123xd10	0.86		87	101	WEISSENBACH J: NATURE GENETIC, JUNE 1994
09	D9S271	AFM263zf1	0.72		199	213	WEISSENBACH J: NATURE GENETIC, JUNE 1994
09	D9S272	AFM276ye1	0.79		195	211	WEISSENBACH J: NATURE GENETIC, JUNE 1994
09	D9S273	AFM280th5	0.73		199	217	WEISSENBACH J: NATURE GENETIC, JUNE 1994
09	D9S274	AFM282zh9	0.71		157	171	WEISSENBACH J: NATURE GENETIC, JUNE 1994
09	D9S275	AFM286yc5	0.71		190	200	WEISSENBACH J: NATURE GENETIC, JUNE 1994
09	D9S276	AFM287xd1	0.74		178	186	WEISSENBACH J: NATURE GENETIC, JUNE 1994
09	D9S277	AFM291xc5	0.85		167	199	WEISSENBACH J: NATURE GENETIC, JUNE 1994
09	D9S278	AFM295ye1	0.76		247	275	WEISSENBACH J: NATURE GENETIC, JUNE 1994
09	D9S279	AFM297wb1	0.64		244	254	WEISSENBACH J: NATURE GENETIC, JUNE 1994
09	D9S280	AFM304td9	0.66		140	160	WEISSENBACH J: NATURE GENETIC, JUNE 1994
09	D9S281	AFM304th1	0.68		188	212	WEISSENBACH J: NATURE GENETIC, JUNE 1994
09	D9S282	AFM308vb1	0.68		231	242	WEISSENBACH J: NATURE GENETIC, JUNE 1994
09	D9S283	AFM318xc9	0.63		179	203	WEISSENBACH J: NATURE GENETIC, JUNE 1994
09	D9S284	AFM331yh5	0.78		136	150	WEISSENBACH J: NATURE GENETIC, JUNE 1994
09	D9S285	AFM339xd9	0.69		107	129	WEISSENBACH J: NATURE GENETIC, JUNE 1994
09	D9S286	AFM344yc9	0.86		139	165	WEISSENBACH J: NATURE GENETIC, JUNE 1994
09	D9S287	AFM347zf5	0.74		168	180	WEISSENBACH J: NATURE GENETIC, JUNE 1994
09	D9S288	AFMa123xg1	0.66		124	140	WEISSENBACH J: NATURE GENETIC, JUNE 1994
09	D9S289	AFMa131yc1	0.77		75	87	WEISSENBACH J: NATURE GENETIC, JUNE 1994
09	D9S290	AFMa131yg9	0.29		148	160	WEISSENBACH J: NATURE GENETIC, JUNE 1994
09	D9S299	GATA27	0.70		216	216	GDB
09	D9S301	GATA7D12	0.73		234	234	GDB
09	D9S301	GATA7D12	0.75		N/A	N/A	GDB
09	D9S302	GATA4D10	0.89		294	294	GDB
09	D9S303	GATA3D04	0.86		163	163	GDB
09	D9S304	GATA5E06	0.85		164	164	GDB
09	D9S306	MFD220	0.87	0.85	102	136	GENOMICS 8:400- , 1990
09	D9S307	MFD308	0.65	0.63	166	178	GENOMICS 8:400- , 1990
09	D9S318	GATA11A07	0.50		272	292	GDB
09	D9S319	GATA12C06	0.88		166	179	GDB
09	D9S322	GGAT2B03	0.38		207	207	GDB
09	D9S324	GATA8E06	0.75		236	260	GDB
09	D9S328E	220	0.84		175	203	GDB
09	D9S43	MFD 14	0.83	0.74	80	102	NAR 18:( ):6465, 1990
09	D9S50	MFD 85	0.47	0.56	76	100	GENOMICS 12:607-609, 1992
09	D9S51	MFD 94	0.84	0.78	135	159	GENOMICS 14:144-152, 1992
09	D9S52	MFD110	0.59	0.55	213	229	GENOMICS 14:144-152, 1992
09	D9S53	MFD135	0.87	0.82	116	150	GENOMICS 14:144-152, 1992
09	D9S54	MFD141	0.54	0.52	110	126	GENOMICS 14:144-152, 1992
09	D9S55	NA	0.00	0.75	165	197	SHARMA,V. ET AL.(1992) NAR 19,1722.
09	D9S58	NA	0.83		105	137	KWIATKOWSKI,D.&GUSELLA,J.(1992) NAR20,930.
09	D9S59	NA	0.86		94	116	KWIATKOWSKI,D.&GUSELLA,J.(1992) NAR20,930.
09	D9S60	NA	0.80		136	154	KWIATKOWSKI,D.&GUSELLA,J.(1992) NAR20,930.
09	D9S61	NA	0.80		112	150	KWIATKOWSKI,D.&GUSELLA,J.(1992) NAR20,930.
09	D9S62	2BH10	0.85		156	188	GDB
09	D9S63	NA	0.80		116	142	KWIATKOWSKI,D.&GUSELLA,J.(1992) NAR20,930.
09	D9S64	NA	0.93		99	117	KWIATKOWSKI,D.&GUSELLA,J.(1992) NAR20,930.
09	D9S65	NA	0.71		151	161	KWIATKOWSKI,T. ET AL.(1991) GENOMICS10,921-26.
09	D9S66	NA	0.87		109	135	KWIATKOWSKI,D.ET AL.(1992) GENOMICS 12,229-240.
09	D9S67	NA	0.72		71	113	KWIATKOWSKI,D.&GUSELLA,J.(1992) NAR20,930.
09	D9S741	MFD361	0.79	0.77	188	206	WEBER, J. PERSONAL COMMUNICATION
09	D9S749	9CMP9	0.86		246	280	GDB
09	D9S903	ATA2G03	0.07		264	264	GDB
09	D9S904	ATA3H11	0.40		196	202	GDB
09	D9S905	GAAT1C06	0.36		289	301	GDB
09	D9S906	GATA46C09	0.64		132	132	GDB

CH	LOCUS	ASSAY	HET	PIC	SIZE ANGE MIN	MAX	REFERENCE
09	D9S907	GGAA16D02	0.87		270	270	GDB
09	DBH	NA	0.69	0.69	253	276	NAHMIAS,J. ET AL.(1992) HMG 1,286.
09	DBHII	DH59AC	0.54		170	180	PORTER, C. (1992) NAR 20, 1429.
09	GSN	NA	0.70	0.76	111	147	KWIATKOWSKI D.J.&PERMAN,S.(1991)NAR 19,967.
09	HXB	NA	0.79		107	117	OZELIUS,L. ET AL.(1992) NAR 1,141.
09	IFNA	NA	0.72		138	150	KWIATKOWSKI,D.& DIAZ,M. (1992) HMG1,658.
09	RPS6	SACA	0.91		119	135	GDB
10	D10S107	MFD 78	0.53	0.46	155	167	GENOMICS 13:532-536, 1992
10	D10S108	MFD100	0.64	0.55	117	147	GENOMICS 13:532-536, 1992
10	D10S109	MFD150	0.71	0.73	82	98	GENOMICS 13:532-536, 1992
10	D10S1099	sSSD-2	0.78		194	228	CARLSON, K. ET AL. (1994) HMG 3, 1207.
10	D10S110	MFD157	0.58	0.55	184	194	GENOMICS 13:532-536, 1992
10	D10S111	MFD164	0.67	0.62	144	154	GENOMICS 12:604-606, 1992
10	D10S1207	ATA1F04	0.58		107	107	GDB
10	D10S1208	ATA5A04	0.67		179	200	GDB
10	D10S1209	GAAT2G10	0.26		262	262	GDB
10	D10S1210	GAAT4C01	0.35		266	266	GDB
10	D10S1211	GATA47G05	0.73		140	140	GDB
10	D10S1212	GGAA3D10	0.56		175	175	GDB
10	D10S1213	GGAA5D10	0.73		93	133	GDB
10	D10S1214	GGAA6E04	0.97		320	432	GDB
10	D10S1215	GGAA7H02	0.81		243	243	GDB
10	D10S1216	GGAA8G02	0.69		211	231	GDB
10	D10S1217	GGAA9B08	0.53		143	179	GDB
10	D10S141	PCR1	0.85		118	139	GDB
10	D10S168	MFD175	0.44	0.57	159	175	GENOMICS 13:532-536, 1992
10	D10S169	MFD187	0.73	0.68	99	117	GENOMICS 13:532-536, 1992
10	D10S172	MIT-MX5	0.64		152	152	HUMAN GENET 87:401, 1991
10	D10S173	MIT-MX6	0.81		155	155	HUMAN GENET 87:401, 1991
10	D10S174	MIT-MS88	0.75		150	150	HUMAN GENET 87:401, 1991
10	D10S175	MIT-MS77	0.70		233	233	HUMAN GENET 87:401, 1991
10	D10S179	NA	0.82	0.80	117	141	COUCH,FJ.ET AL.(1992) NAR 20,1431.
10	D10S183	MFD200	0.83	0.80	124	158	GENOMICS 8:400- , 1990
10	D10S184	MFD214	0.63	0.60	174	186	GENOMICS 8:400- , 1990
10	D10S185	AFM019th6	0.77		143	159	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
10	D10S186	AFM023xc3	0.81		147	163	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
10	D10S187	AFM042xa9	0.84		92	114	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
10	D10S188	AFM063xc5	0.57		121	127	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
10	D10S189	AFM063xf4	0.73		180	188	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
10	D10S190	AFM065yh11	0.84		203	219	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
10	D10S191	AFM066xa1	0.82		124	152	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
10	D10S192	AFM094tc9	0.78		180	198	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
10	D10S193	AFM095zh7	0.82		213	231	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
10	D10S195	AFM098xg1	0.69		170	184	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
10	D10S196	AFM115xf2	0.79		101	990	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
10	D10S197	AFM119xh12	0.75		161	173	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
10	D10S198	AFM126yd6	0.70		185	203	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
10	D10S199	AFM137xh4	0.86		171	183	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
10	D10S200	AFM151xd2	0.00		291	303	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
10	D10S201	AFM155zc3	0.83		281	305	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
10	D10S202	AFM161xc1	0.68		83	93	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
10	D10S203	AFM162xf6	0.56		231	239	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
10	D10S204	AFM164tg9	0.75		290	330	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
10	D10S205	AFM164yd8	0.90		224	244	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
10	D10S206	AFM168xe1	0.58		254	268	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
10	D10S207	AFM175xh10	0.51		108	114	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
10	D10S208	AFM183xg1	0.80		170	186	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
10	D10S209	AFM185xe11	0.75		181	216	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
10	D10S210	AFM186xe5	0.80		130	140	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
10	D10S211	AFM198wf8	0.84		195	211	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
10	D10S212	AFM198zb4	0.71		189	201	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
10	D10S213	AFM199zb6	0.83		173	191	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
10	D10S214	AFM200yh6	0.50		110	118	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
10	D10S215	AFM205wd12	0.81		152	208	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
10	D10S216	AFM205zd8	0.62		202	223	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
10	D10S217	AFM212xd6	0.82		196	212	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
10	D10S218	AFM220xe5	0.67		239	248	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
10	D10S219	AFM240wg7	0.78		89	103	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
10	D10S220	AFM240xa9	0.84		267	291	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
10	D10S221	AFM249vf1	0.81		95	115	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801

CH	LOCUS	ASSAY	HET	PIC	SIZE MIN	ANGE MAX	REFERENCE
10	D10S222	AFM249wc5	0.72		264	276	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
10	D10S223	AFM254wd9	0.67		221	231	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
10	D10S224	AFM254xb1	0.67		125	131	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
10	D10S225	AFM256yf9	0.72		262	278	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
10	D10S226	AFM260zc5	0.54		156	168	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
10	D10S245	MFD248	0.76	0.72	123	141	GENOMICS 8:400- ,1990
10	D10S246	MFD228	0.44	0.42	106	122	GENOMICS 8:400- ,1990
10	D10S247E	FB7F11	0.85		179	197	KHAN, A. ET AL. (1992) NATURE GENETICS 2, 180-185.
10	D10S249	AFM207wd12	0.75		118	134	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
10	D10S254	MFD249	0.73	0.75	98	114	GENOMICS 8:400- ,1990
10	D10S412I	8472/3	0.95		117	141	GDB
10	D10S412II	8986/7	0.86		109	129	GDB
10	D10S463	SE3	0.65		90	102	GDB
10	D10S464	MFD274	0.78	0.75	130	146	GENOMICS 8:400- ,1990
10	D10S465	MFD277	0.61	0.54	110	118	GENOMICS 8:400- ,1990
10	D10S466	MFD289	0.78	0.75	120	140	GENOMICS 8:400- ,1990
10	D10S467	MFD296	0.56	0.51	147	155	GENOMICS 8:400- ,1990
10	D10S468	MFD298	0.69	0.65	82	96	GENOMICS 8:400- ,1990
10	D10S469	CRAD0	0.87		128	146	GDB
10	D10S505	UT517	0.00		247	247	GDB
10	D10S507	UT539	0.00		148	148	GDB
10	D10S521	UT5027	0.00		155	215	GDB
10	D10S522	UT1270	0.00		341	341	GDB
10	D10S525	UT1357	0.00		347	347	GDB
10	D10S526	UT1699	0.00		270	270	GDB
10	D10S527	UT1729	0.00		268	268	GDB
10	D10S529	AFM162xa1	0.69		144	160	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S530	AFM177xe3	0.78		182	208	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S531	AFM185yc9	0.64		190	198	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S532	AFM192xa5	0.59		265	275	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S533	AFM038xe9	0.85		114	116	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S534	AFM198yh8	0.63		220	250	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S535	AFM200wf4	0.66		126	136	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S536	AFM022xe3	0.57		126	140	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S537	AFM203xc5	0.55		136	160	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S538	AFM203zf8	0.67		190	196	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S539	AFM205tg11	0.79		96	114	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S540	AFM205xe11	0.78		255	259	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S541	AFM205xe3	0.85		247	273	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S542	AFM052xg5	0.79		225	231	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S544	AFM210we3	0.53		274	290	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S545	AFM211we7	0.51		232	252	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S546	AFM063xb4	0.28		148	160	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S547	AFM214yc9	0.81		236	250	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S548	AFM220xd4	0.66		140	148	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S549	AFM234wh6	0.77		177	191	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S550	AFM238we5	0.73		246	260	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S551	AFM240vf10	0.79		119	127	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S552	AFM240wg5	0.62		85	103	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S553	AFM240xd6	0.46		205	235	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S554	AFM240yc7	0.77		150	162	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S555	AFM242yc7	0.71		229	233	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S556	AFM242zg7	0.61		191	201	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S557	AFM079ya1	0.69		139	147	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S558	AFM248vb9	0.32		192	212	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S559	AFM248vg1	0.55		180	192	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S560	AFM248ze1	0.79		184	208	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S561	AFM254xd9	0.52		265	315	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S562	AFM259xg5	0.65		185	205	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S563	AFM084xf11	0.72		135	141	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S564	AFM029xh12	0.72		252	262	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S565	AFM137xb8	0.48		109	120	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S566	AFM154xh2	0.78		200	212	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S567	AFM263yd5	0.81		192	208	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S568	AFM265vd9	0.72		147	181	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S569	AFM265zg5	0.82		138	154	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S570	AFM268zh1	0.68		287	305	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S571	AFM269xg9	0.46		187	213	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S572	AFM269yb9	0.72		275	283	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S573	AFM269zd1	0.75		167	175	WEISSENBACH J: NATURE GENETIC, JUNE 1994

CH	LOCUS	ASSAY	HET	PIC	SIZE ANGE MIN	MAX	REFERENCE
10	D10S574	AFM269zd9	0.86		124	135	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S575	AFM270xb1	0.77		251	269	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S576	AFM275yg9	0.88		172	174	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S577	AFM276xb5	0.83		199	213	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S578	AFM282ya9	0.59		160	184	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S579	AFM282yc1	0.86		260	276	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S580	AFM284vf5	0.75		91	105	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S581	AFM287yf9	0.55		129	155	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S582	AFM289zd1	0.84		114	138	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S583	AFM289zh5	0.74		201	219	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S584	AFM294wh9	0.62		182	190	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S585	AFM294zd9	0.74		233	249	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S586	AFM295th1	0.59		124	132	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S587	AFM296zg9	0.78		172	186	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S588	AFM298zf5	0.75		136	142	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S589	AFM302wb5	0.42		173	193	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S590	AFM304wh1	0.74		241	255	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S591	AFM309yd9	0.72		212	232	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S593	AFM311yb1	0.84		126	144	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S594	AFM317zd9	0.77		100	108	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S595	AFM324xc1	0.81		185	207	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S596	AFM329xa9	0.82		262	274	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S597	AFM331xa9	0.64		206	222	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S599	AFM337ya5	0.78		209	215	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S600	AFM338ta5	0.85		175	193	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S601	AFM342xe9	0.73		216	232	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S602	AFM343vd9	0.63		233	255	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S603	AFM350wa5	0.58		232	256	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S604	AFM362tb1	0.62		177	187	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S605	AFMa120xc5	0.63		90	110	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S606	AFMa131yc5	0.82		216	240	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S607	AFMa133zg5	0.64		166	178	WEISSENBACH J: NATURE GENETIC, JUNE 1994
10	D10S608	C10-GT02	0.80		166	190	GDB
10	D10S609	GATA2G08	0.68		240	242	GDB
10	D10S610	GATA5A02	0.46		268	272	GDB
10	D10S611	GATA3G07	0.79		N/A	N/A	GDB
10	D10S611	GATA3G07	0.80		151	151	GDB
10	D10S674	GATA6E06	0.73		218	254	GDB
10	D10S675	GATA6H05	0.62		102	122	GDB
10	D10S676	GATA7B01	0.77		175	199	GDB
10	D10S677	GGAA2F11	0.86		197	225	GDB
10	D10S681	JY4069-9	0.85		157	183	GDB
10	D10S88	MFD 7	0.54	0.51	205	217	GENOMICS 13:532-536, 1992
10	D10S89	MFD 28	0.80	0.71	142	156	GENOMICS 13:532-536, 1992
10	D10S91	MFD 29	0.67	0.60	115	125	GENOMICS 13:532-536, 1992
10	D1S537	ATC3	0.62		158	158	GDB
10	GLUDP2	PCR1	0.78		191	203	GDB
10	GLUDP5	C10-GT01	0.71		286	340	GOULIELMOS,G. ET AL.(1993) HMG 2,1328.
10	RBP3	NA	0.79		355	387	PAPI,L. ET AL.(1992) HMG 1,450.
10	RBP3-2	sJRH-1	0.90		274	308	HOWE, J. AM. J. HUM. GENET. 51, 1430-1442.
10	RET	STCL	0.71		N/A	N/A	GDB
10	TCF8	NA	0.73		135	149	GDB
10	ZNF22	NA	0.84		151	175	GDB
11	CD3D	MFD 69	0.74	0.69	85	99	NAR 18(13):4036, 1990
11	D11S1240	ms73	0.82		160	160	IIZUKA, M. ET AL. (1994) GENOMICS 19, 581-584.
11	D11S1242	ms65	0.71		170	170	IIZUKA, M. ET AL. (1994) GENOMICS 19, 581-584.
11	D11S1244	ms61	0.68		270	270	IIZUKA, M. ET AL. (1994) GENOMICS 19, 581-584.
11	D11S1245	ms60	0.73		180	180	IIZUKA, M. ET AL. (1994) GENOMICS 19, 581-584.
11	D11S1246	ms58	0.77		190	190	IIZUKA, M. ET AL. (1994) GENOMICS 19, 581-584.
11	D11S1247	ms52	0.69		200	200	IIZUKA, M. ET AL. (1994) GENOMICS 19, 581-584.
11	D11S1249	ms44	0.68		190	190	IIZUKA, M. ET AL. (1994) GENOMICS 19, 581-584.
11	D11S1250	ms42	0.70		200	200	IIZUKA, M. ET AL. (1994) GENOMICS 19, 581-584.
11	D11S1251	ms39	0.80		150	150	IIZUKA, M. ET AL. (1994) GENOMICS 19, 581-584.
11	D11S1253	ms31	0.78		160	160	IIZUKA, M. ET AL. (1994) GENOMICS 19, 581-584.
11	D11S1256	nrms1	0.91		200	200	IIZUKA, M. ET AL. (1994) GENOMICS 19, 581-584.
11	D11S1257	nrms2	0.72		150	150	IIZUKA, M. ET AL. (1994) GENOMICS 19, 581-584.
11	D11S1258	nrms4	0.70		220	220	IIZUKA, M. ET AL. (1994) GENOMICS 19, 581-584.
11	D11S1259	nrms7	0.70		270	270	IIZUKA, M. ET AL. (1994) GENOMICS 19, 581-584.
11	D11S1263	c4	0.81		200	200	IIZUKA, M. ET AL. (1994) GENOMICS 19, 581-584.
11	D11S1264	rms7	0.68		500	500	IIZUKA, M. ET AL. (1994) GENOMICS 19, 581-584.

CH	LOCUS	ASSAY	HET	PIC	SIZE MIN	ANGE MAX	REFERENCE
11	D11S1294	UT928	0.83				VANGAITE, L. ET AL. (1994) GENOMICS 22, 231-233.
11	D11S1300	UT1004	0.75				VANGAITE, L. ET AL. (1994) GENOMICS 22, 231-233.
11	D11S1307	AFM166ze1	0.63		120	138	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1308	AFM189xc1	0.50		233	239	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1309	AFM200vg5	0.59		237	249	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1310	AFM200zf10	0.79		222	228	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1311	AFM203ve1	0.70		127	147	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1312	AFM045xf2	0.35		197	233	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1313	AFM211xe1	0.81		184	204	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1314	AFM212xe3	0.79		209	227	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1315	AFM212yf12	0.57		158	162	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1316	AFM214xg7	0.81		200	208	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1317	AFM214xh6	0.63		214	228	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1318	AFM218xe1	0.67		123	145	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1319	AFM234wd2	0.51		182	198	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1320	AFM234yf10	0.61		225	233	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1321	AFM238xe7	0.62		197	215	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1322	AFM248tg9	0.81		224	230	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1323	AFM248xf9	0.83		201	207	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1324	AFM079zd3	0.58		110	128	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1325	AFM254zd5	0.50		80	84	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1326	AFM255zg1	0.84		247	255	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1327	AFM256vc1	0.51		248	254	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1328	AFM265wa9	0.83		151	165	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1329	AFM269za5	0.73		257	269	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1330	AFM270vb1	0.83		156	160	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1331	AFM277wg1	0.72		191	205	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1332	AFM281wf9	0.47		182	204	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1333	AFM282we5	0.84		254	274	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1334	AFM283wg1	0.79		134	150	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1335	AFM284xd9	0.65		172	183	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1336	AFM286xf5	0.80		232	252	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1337	AFM289ya9	0.72		279	295	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1338	AFM289yc5	0.75		255	265	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1339	AFM291yh1	0.71		120	144	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1340	AFM295xg5	0.78		188	200	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1341	AFM295yd5	0.83		167	181	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1342	AFM296xg9	0.56		257	267	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1343	AFM296yd9	0.87		226	252	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1344	AFM298vc9	0.76		273	293	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1345	AFM302xb9	0.71		232	240	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1346	AFM319wa9	0.64		263	281	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1347	AFM320xh1	0.81		177	203	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1348	AFM323vc1	0.46		172	175	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1349	AFM323wf5	0.75		260	280	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1350	AFM323ye1	0.61		201	219	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1351	AFM324zh9	0.72		252	270	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1352	AFM329wb5	0.61		231	249	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1353	AFM331yc5	0.54		196	209	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1354	AFM338xe1	0.66		167	179	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1355	AFM344tb9	0.88		141	147	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1356	AFM344zc1	0.87		193	213	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1357	AFM344zg1	0.83		134	140	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1358	AFM345zd1	0.66		138	146	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1359	AFM347tc5	0.68		210	234	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1360	AFM362tb9	0.75		103	117	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1361	AFMa131xd5	0.75		206	220	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1362	AFMa132xh9	0.56		187	207	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1363	AFMa134wh5	0.86		242	252	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1364	AFMa139yg1	0.63		134	144	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1365	AFMa141xd1	0.79		98	124	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S1366	GATA3B05	0.71		241	241	GDB
11	D11S1367	GATA7A03	1.00		220	244	GDB
11	D11S1368	GATA8A08	0.67		178	198	GDB
11	D11S1369	GATA5C04	0.86		179	179	GDB
11	D11S1377	MFD316	0.78	0.76	128	146	GENOMICS 8:400- , 1990
11	D11S1378	MFD322	0.68	0.65	149	160	GENOMICS 8:400- , 1990
11	D11S1383	26BH1	0.70		88	98	GDB
11	D11S1384	GATA11A02	0.76		288	300	GDB
11	D11S1385	GATA2A01	0.77		197	217	GDB

CH	LOCUS	ASSAY	HET	PIC	SIZE ANGE MIN	MAX	REFERENCE
11	D11S1386	ATC3C12	0.00		166	166	GDB
11	D11S1390	GATA4B05	0.64		145	165	GDB
11	D11S1391	GATA4E01	0.86		158	178	GDB
11	D11S1392	GATA6B09	0.80		200	220	GDB
11	D11S1393	GATA6C04	0.64		198	210	GDB
11	D11S1394	GATA6C11	0.81		222	238	GDB
11	D11S1395	GATA6G03	0.70		220	232	GDB
11	D11S1396	GGAA2C10	0.73		136	176	GDB
11	D11S1397	GGAT1C7	0.68		142	150	GDB
11	D11S1974	ACT2E05	0.25		116	125	GDB
11	D11S1975	GAAT1B01	0.47		104	112	GDB
11	D11S1976	GAAT2C05	0.57		172	172	GDB
11	D11S1977	GAAT2D08	0.90		110	110	GDB
11	D11S1978	GATA22D05	0.77		250	298	GDB
11	D11S1979	GATA45H10	0.85		246	246	GDB
11	D11S1980	GATA47F03	0.79		197	197	GDB
11	D11S1981	GATA48E02	0.83		160	160	GDB
11	D11S1982	GATA49B02	0.60		133	133	GDB
11	D11S1983	GATA5G01	0.90		208	255	GDB
11	D11S1984	GGAA17G05	0.77		166	206	GDB
11	D11S1985	GGAA5C04	0.73		234	286	GDB
11	D11S1986	GGAA7G08	0.79		176	252	GDB
11	D11S1987	GTAT1D06	0.57		185	209	GDB
11	D11S29	NA	0.83	0.77	143	163	GDB
11	D11S35	NA	0.88	0.79	152	162	LITT,M. ET AL. (1990) NAR 18,5921.
11	D11S387	1H2	0.85		168	196	GDB
11	D11S419	MFD 58	0.49	0.43	112	118	NAR 18(13):4039, 1990
11	D11S420	NA		0.66	188	208	LUO,X.Y. ET AL.(1990) NAR 18,5920.
11	D11S436	NA	0.68		176	188	GUO,Z. ET AL.(1991) NAR 19,6981.
11	D11S439	NA	0.80		160	188	HAUGE,X. ET AL.(1992) HMG 1,548.
11	D11S480	MF	0.60		189	201	GDB
11	D11S488	NA	0.87		243	295	BROWNE,D. ET AL.(1993) HMG 2,89.
11	D11S490	NA		0.72	147	167	LUO,X.Y. ET AL.(1990) NAR 18,7470.
11	D11S527	NA		0.88	142	166	BROWNE,D. ET AL.(1991) NAR 19,4790.
11	D11S528	NA		0.60	73	91	HAUGE,X.Y. ET AL.(1991) NAR 19,1964.
11	D11S534	NA		0.74	228	244	HAUGE,X. ET AL.(1991) NAR 19,4308.
11	D11S554	388II	0.84		174	254	GDB
11	D11S569	434	0.84		139	158	GDB
11	D11S614	8D11	0.85		160	178	GDB
11	D11S787	NA	0.79		164	182	OVERBECK,L., ET AL.(1993)HMG 2,611.
11	D11S809	C43	0.91		200	232	GDB
11	D11S836	MFD108	0.70	0.62	66	80	J. WEBER, PERS. COMM.
11	D11S860	BS48	0.80		154	196	MCNODE,L. ET AL.(1992) NAR 20,1161.
11	D11S861	MIT-A136	0.70		154	154	HUMAN GENET 87:401, 1991
11	D11S862	MIT-MS7	0.83		152	152	HUMAN GENET 87:401, 1991
11	D11S863	MIT-MS20	0.65		133	133	HUMAN GENET 87:401, 1991
11	D11S865	MIT-E137	0.81		170	170	HUMAN GENET 87:401, 1991
11	D11S870	MFD 90	0.60		154	160	GENOMICS 8:400- , 1990
11	D11S871	MFD132	0.77	0.65	188	194	GENOMICS 8:400- , 1990
11	D11S872	MFD105	0.50		158	166	GENOMICS 8:400- , 1990
11	D11S873	MFD127	0.90		176	204	JONES,M. ET AL.(1992) HMG 1,131-33.
11	D11S874	MFD161	0.75		158	170	JONES,M. ET AL.(1992) HMG 1,131-33.
11	D11S875	MFD166	0.90		103	125	GENOMICS 8:400- , 1990
11	D11S876	MFD212	0.89	0.88	216	242	JONES,M. ET AL.(1992) HMG 1,131-33.
11	D11S896	MFD216	0.74	0.70	169	183	GENOMICS 8:400- , 1990
11	D11S897	MFD231	0.84	0.83	98	120	GENOMICS 8:400- , 1990
11	D11S898	AFM022te1	0.38		140	156	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S899	AFM022th2	0.69		87	111	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
11	D11S900	AFM059yc5	0.79		91	109	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
11	D11S901	AFM063yg1	0.83		160	176	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
11	D11S902	AFM072yd3	0.81		145	163	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
11	D11S903	AFM077xe1	0.75		99	109	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
11	D11S904	AFM081za5	0.83		185	201	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
11	D11S905	AFM105xb10	0.75		208	228	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
11	D11S906	AFM107xc7	0.75		291	303	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
11	D11S907	AFM109ya1	0.74		163	173	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
11	D11S908	AFM120xe9	0.77		141	151	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
11	D11S909	AFM154xb6	0.62		113	125	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
11	D11S910	AFM154yh2	0.73		249	261	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
11	D11S911	AFM155xh10	0.86		159	203	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801

CH	LOCUS	ASSAY	HET	PIC	SIZE MIN	ANGE MAX	REFERENCE
11	D11S912	AFM157xh6	0.82		101	123	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
11	D11S913	AFM164zf12	0.00		221	227	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
11	D11S914	AFM165yf10	0.73		275	285	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
11	D11S915	AFM178xf12	0.82		254	274	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
11	D11S916	AFM185ya1	0.74		135	153	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
11	D11S917	AFM198yb4	0.81		143	157	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
11	D11S918	AFM203vg1	0.63		181	199	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
11	D11S919	AFM203vg7	0.81		245	261	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
11	D11S920	AFM207ze3	0.64		243	253	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
11	D11S921	AFM212xa11	0.71		243	255	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
11	D11S922	AFM217yb10	0.94		88	138	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
11	D11S923	AFM218ya7	0.72		201	225	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
11	D11S924	AFM220xh6	0.73		245	253	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
11	D11S925	AFM220yb6	0.85		173	199	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
11	D11S926	AFM224zc7	0.74		135	145	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
11	D11S927	AFM225yb4	0.85		129	149	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
11	D11S928	AFM234wh12	0.71		277	289	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
11	D11S929	AFM234xc3	0.88		218	240	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
11	D11S930	AFM238xe5	0.68		225	237	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
11	D11S931	AFM238xh10	0.74		251	267	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
11	D11S932	AFM240vh2	0.64		150	164	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
11	D11S933	AFM240ye1	0.80		247	263	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
11	D11S934	AFM248wf5	0.85		180	206	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
11	D11S935	AFM254zb9	0.75		196	208	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
11	D11S936	AFM256za5	0.42		250	256	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
11	D11S937	AFM256zb5	0.88		230	264	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
11	D11S938	AFM259yc1	0.50		207	219	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
11	D11S939	AFM267yh5	0.69		240	248	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
11	D11S940	AFM268vd5	0.74		163	185	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
11	D11S956	SMSH3	0.88		247	303	GDB
11	D11S969	AFM205vf10	0.76		141	149	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
11	D11S975	MFD251	0.75	0.72	216	242	GENOMICS 8:400- , 1990
11	D11S976	MFD254	0.83	0.81	117	139	GENOMICS 8:400- , 1990
11	D11S982E	NA	0.77		112	128	XAIO, H. ET AL. (1993) HMG 2, 1081.
11	D11S986	AFM255ye1	0.68		137	169	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S987	AFMa131ye5	0.83		82	118	WEISSENBACH J: NATURE GENETIC, JUNE 1994
11	D11S988	MFD257	0.83	0.81	112	138	GENOMICS 8:400- , 1990
11	D11S989	MFD282	0.00		144	156	GENOMICS 8:400- , 1990
11	D11S990	MFD290	0.79	0.76	73	89	GENOMICS 8:400- , 1990
11	D11S991	MFD309	0.61	0.58	192	206	GENOMICS 8:400- , 1990
11	D11S992	MFD263	0.60	0.57	159	175	GENOMICS 8:400- , 1990
11	D11S995	591/1	0.79		123	139	BROWNE,D. ET AL.(1993) HMG 2,1332.
11	DRD2	NA	0.00	0.76	80	86	HAUGE,X.Y. ET AL(1991) GENOMICS 27,527-30.
11	FCERIB	NA	0.69		112	128	GDB
11	FGR3	PCR1.	0.82		161	177	POLYMEROPOULOS (1990) NAR 18, 7468.
11	GSTP1	PCR1	0.79		190	235	HARADA, S. (1994) HUM. GENET. 93, 223-224.
11	HBB	NA	0.72		141	149	HAUGE,X.Y. ET AL.(1991) NAR 19,1964.
11	HBE1	NA	0.75		136	144	ROGAEV,E. ET AL.(1992) HMG 1,285.
11	HRAS1	NA	0.52		106	118	TANCI,P. ET AL.(1992)NAR 20,1157.
11	INT-2	NA	0.85		161	177	POLYMEROPOULOS,M. ET AL.(1990) NAR 18,7468.
11	NCAM	16-F	0.89		94	138	TELATAR, M. ET AL. (1994) HMG 3, 842.
11	PYGM(AT)	NA	0.71	0.71	367	615	IN PREPARATION
11	PYGM(CA)	NA	0.89	0.89	162	188	IWASAKI,H. ET AL.(1992) GENOMICS13,7-15.
11	TH	NA	0.78	0.75	244	260	POLYMEROPOULOS,M. ET AL.(1991) NAR 19,3753.
11	TH01	NA	0.79		183	207	FEENER,ET AL.(1991) AM.J.HUM.GENET. 48,621-627.
11	TYR	PCR6	0.58	0.52	286	298	GDB
12	CACNL1A1	NA	0.75	0.73	210	214	POWERS,P. ET AL.(1992) GENOMICS 14,206-207.
12	CD4	PCR1	0.79		85	115	GDB
12	D12S100	AFM220zc7	0.73		137	153	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
12	D12S101	AFM234tg11	0.81		194	231	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
12	D12S102	AFM238yb10	0.78		241	259	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
12	D12S1022	GAAT1D02	0.40		169	193	GDB
12	D12S1023	GATA27G11	0.67		153	169	GDB
12	D12S1024	GATA30D01	0.54		147	147	GDB
12	D12S1025	GATA31D06	0.83		186	186	GDB
12	D12S1026	GATA31D11	0.79		217	217	GDB
12	D12S1027	GATA31F05	0.79		244	244	GDB
12	D12S1028	GATA32A08	1.00		140	140	GDB
12	D12S1029	GATA47G02	0.60		134	134	GDB
12	D12S103	AFM249vf9	0.37		267	273	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801

CH	LOCUS	ASSAY	HET	PIC	SIZE ANGE MIN	MAX	REFERENCE
12	D12S1030	GATA6H09	0.57		243	271	GDB
12	D12S104	AFM259zc9	0.52		180	184	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
12	D12S105	AFM262xb9	0.72		137	155	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
12	D12S106	AFM262zd9	0.43		229	239	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
12	D12S1074	GGAA2G02	0.32		164	164	GDB
12	D12S1075	GGAT1A12	0.45		189	205	GDB
12	D12S129	NA	1.00		194	202	GDB
12	D12S137	NA	0.76		416	424	GDB
12	D12S161	NA	0.90		169	197	GDB
12	D12S172	NA	0.85		298	316	GDB
12	D12S183	NA	0.80		164	176	GDB
12	D12S188	NA	0.70		239	245	GDB
12	D12S191	NA	0.80		303	311	GDB
12	D12S218	NA	0.80		176	184	GDB
12	D12S221	NA	0.82		270	290	GDB
12	D12S234	NA	0.85		311	329	GDB
12	D12S262	NA	0.76		223	235	GDB
12	D12S269	MFD259	0.80	0.78	107	131	GENOMICS 8:400- , 1990
12	D12S270	MFD305	0.71	0.68	143	169	GENOMICS 8:400- , 1990
12	D12S271	GT22	0.65		96	102	THISELTON,D. ET AL.(1993)HMG 2,613.
12	D12S305	AFM184yf2	0.58		119	131	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S306	AFM198wc3	0.87		200	230	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S308	AFM198yf6	0.78		222	228	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S309	AFM199wb10	0.75		136	146	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S310	AFM205xg3	0.61		243	253	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S311	AFM206yc1	0.72		199	207	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S312	AFM207va9	0.64		238	248	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S313	AFM207xf2	0.80		137	153	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S314	AFM207xf8	0.78		246	260	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S316	AFM210yc9	0.84		252	264	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S318	AFM214xc9	0.82		247	263	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S319	AFM217xd10	0.82		210	222	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S320	AFM073wh7	0.72		196	216	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S321	AFM220zf4	0.57		155	177	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S322	AFM224xf12	0.39		170	182	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S323	AFM224yf10	0.70		180	186	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S324	AFM234tb10	0.84		207	221	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S325	AFM234tf12	0.37		209	229	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S326	AFM238wa1	0.47		237	261	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S327	AFM248tg1	0.73		182	201	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S328	AFM248xc5	0.46		205	209	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S329	AFM249xh9	0.72		151	165	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S330	AFM086xd7	0.69		156	198	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S331	AFM092wd11	0.68		175	181	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S332	AFM263zd1	0.65		317	341	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S333	AFM265zb1	0.62		240	250	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S334	AFM269ye1	0.66		208	218	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S335	AFM273vg9	0.64		249	265	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S336	AFM273zc9	0.45		195	209	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S337	AFM283wf9	0.80		133	139	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S338	AFM291wd9	0.89		259	275	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S339	AFM294wc5	0.69		257	277	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S340	AFM294xg1	0.80		237	241	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S341	AFM294yd9	0.52		114	130	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S342	AFM294ze9	0.83		217	237	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S343	AFM295ye9	0.73		111	123	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S344	AFM296xd9	0.73		132	138	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S345	AFM296yg5	0.42		210	238	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S346	AFM298xe5	0.62		154	174	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S347	AFM298zb1	0.51		96	114	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S348	AFM299zc5	0.77		286	296	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S349	AFM299zd5	0.76		165	181	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S350	AFM302wb9	0.76		179	189	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S351	AFM302wd9	0.67		151	169	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S352	AFM303xd9	0.67		148	164	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S353	AFM304wg5	0.65		89	105	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S354	AFM304wh5	0.50		187	205	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S355	AFM309xh1	0.65		172	197	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S356	AFM309zf9	0.77		205	225	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S357	AFM310vd5	0.67		193	225	WEISSENBACH J: NATURE GENETIC, JUNE 1994

CH	LOCUS	ASSAY	HET	PIC	SIZE MIN	ANGE MAX	REFERENCE
12	D12S358	AFM320xb5	0.79		238	270	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S359	AFM329zh9	0.82		210	226	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S360	AFM330yd5	0.79		187	223	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S361	AFM331th9	0.80		244	258	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S362	AFM336yf9	0.33		155	163	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S363	AFM337zg5	0.84		205	215	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S364	AFM345we1	0.57		269	299	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S365	AFM345ze5	0.72		161	170	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S366	AFM351tb9	0.61		185	201	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S367	AFMa123xe1	0.65		135	149	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S368	AFMa128yd5	0.67		200	216	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S369	AFMa142zc5	0.65		201	225	WEISSENBACH J: NATURE GENETIC, JUNE 1994
12	D12S371	AL4	0.84		153	171	GDB
12	D12S372	GATA4H03	0.69		174	190	GDB
12	D12S373	GATA6C01	0.82		209	217	GDB
12	D12S374	GATA7F09	0.91		274	294	GDB
12	D12S375	GATA3F02	0.91		175	176	GDB
12	D12S376	GATA5F06	0.75		147	147	GDB
12	D12S377	GATA4A06	0.93		224	224	GDB
12	D12S378	GATA5H03	0.68		160	160	GDB
12	D12S379	GATA4B09	0.70		196	212	GDB
12	D12S385	MFD331	0.82	0.79	125	151	GENOMICS 8:400- , 1990
12	D12S386	MFD353	0.73	0.68	209	249	GENOMICS 8:400- , 1990
12	D12S388	M758B6-1	0.70		89	131	GDB
12	D12S389	M758B6-21	0.80		139	159	GDB
12	D12S390	GATA11B02	0.82		141	153	GDB
12	D12S391	GATA11H08	0.88		225	233	GDB
12	D12S392	GATA13D05	0.75		136	153	GDB
12	D12S393	GATA15A03	0.69		243	247	GDB
12	D12S395	GATA4H01	0.59		223	243	GDB
12	D12S396	GATA6G11	0.00		200	212	GDB
12	D12S397	GATA8A09	0.93		194	206	GDB
12	D12S398	GGAT2G06	0.62		120	144	GDB
12	D12S43	MFD 84	0.72	0.71	99	113	NAR 18(15):4637, 1990
12	D12S58	MFD 73	0.61	0.70	75	91	JANSEN,G. ET AL.(1993) 2,333.
12	D12S59	MFD75	0.81	0.71	164	182	WEBER, J. PERSONAL COMMUNICATION
12	D12S60	MFD109	0.77	0.75	67	85	WEBER, J. PERSONAL COMMUNICATION
12	D12S61	MFD114	0.76	0.69	99	117	JANSEN,G. ET AL.(1993) 2,333.
12	D12S62	MFD129	0.92	0.82	189	209	JANSEN,G. ET AL.(1993) 2,333.
12	D12S63	MFD133	0.72	0.66	161	175	JANSEN,G. ET AL.(1993) 2,333.
12	D12S64	MFD155	0.64	0.60	107	121	JANSEN,G. ET AL.(1993) 2,333.
12	D12S68	MIT-G117	0.50		201	202	HUMAN GENET 87:401, 1991
12	D12S69	MIT-MS6	0.56		214	214	HUMAN GENET 87:401, 1991
12	D12S70	MIT-MS54	0.72		177	177	HUMAN GENET 87:401, 1991
12	D12S71	MIT-MS159	0.78		179	179	HUMAN GENET 87:401, 1991
12	D12S72	MIT-MS263	0.83		159	159	HUMAN GENET 87:401, 1991
12	D12S75	MFD244	0.63	0.59	129	143	GENOMICS 8:400- , 1990
12	D12S755E	CTG-B37	0.85		141	141	GDB
12	D12S76	AFM010th7	0.72		112	124	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
12	D12S77	AFM026tb5	0.88		163	193	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
12	D12S78	AFM026tf3	0.92		171	207	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
12	D12S79	AFM067yc5	0.00		155	179	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
12	D12S80	AFM102xd6	0.79		207	223	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
12	D12S81	AFM102xg9	0.82		146	166	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
12	D12S82	AFM107xc11	0.91		124	145	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
12	D12S83	AFM112yf4	0.82		85	99	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
12	D12S84	AFM116xb8	0.00		199	219	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
12	D12S85	AFM122xf6	0.69		105	125	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
12	D12S86	AFM123xh2	0.90		124	160	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
12	D12S87	AFM135xe3	0.80		142	168	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
12	D12S88	AFM158yb4	0.85		217	255	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
12	D12S89	AFM164ta5	0.79		254	288	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
12	D12S90	AFM172xd8	0.75		166	182	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
12	D12S91	AFM182xf10	0.70		176	181	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
12	D12S92	AFM203va7	0.77		188	204	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
12	D12S93	AFM205ve5	0.81		271	291	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
12	D12S94	AFM206ze5	0.79		183	211	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
12	D12S95	AFM207ve1	0.77		146	160	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
12	D12S96	AFM207yh10	0.62		201	227	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
12	D12S97	AFM210zd6	0.00		265	279	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801

CH	LOCUS	ASSAY	HET	PIC	SIZE MIN	ANGE MAX	REFERENCE
12	D12S98	AFM211wb6	0.61		228	238	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
12	D12S99	AFM217xa7	0.84		208	232	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
12	F8VWF	MFD 92	0.75	0.71	115	129	J. WEBER, ET AL, GENOMICS, IN PRESS
12	IFNG	NA	0.67		122	134	LINARES-RUIZ,A.(1993) HMG 2,1508.
12	IGF1	MFD 1	0.54	0.53	176	196	AM J HUMAN GENET, 44:388-396, 1989
12	NT3	NA	0.67		139	151	GDB
12	PAH	PCR9	0.80		229	257	GDB
12	PLA2	NA	0.73		122	143	POLYMEROPOULOS,M. ET AL.(1991) NAR 19,1718.
12	VWF-DNR	NA	0.77		176	190	GDB
12	VWF-TNR	NA	0.73	0.78	138	162	KIMPTON,C. ET AL.(1992) HMG 1,287.
12	Y21	pY21/1	0.83		193	217	BROWNE, D. & LITT, M. (1994) HMG 3, 842
13	D13S115	MIT-MS34	0.82		169	169	HUMAN GENET 87:401, 1991
13	D13S118	Utsw 1312	0.73	0.67	187	201	BOWCOCK,A. ET AL.(1993) GENOMICS 15,376-386.
13	D13S119	Utsw 1310	0.72	0.74	124	140	BOWCOCK,A. ET AL.(1993) GENOMICS 15,376-386.
13	D13S120	Utsw 1353	0.81	0.76	112	136	BOWCOCK,A. ET AL.(1993) GENOMICS 15,376-386.
13	D13S121	Utsw 1305	0.73	0.74	160	178	BOWCOCK,A. ET AL.(1993) GENOMICS 15,376-386.
13	D13S122	Utsw 1334	0.87	0.80	85	111	BOWCOCK,A. ET AL.(1993) GENOMICS 15,376-386.
13	D13S124	MFD179	0.50		184	192	GENOMICS 8:400- , 1990
13	D13S125	Utsw 1334	0.72	0.67	129	155	BOWCOCK,A. ET AL.(1993) GENOMICS 15,376-386.
13	D13S126	Utsw1303	0.63		100	112	BOWCOCK,A. ET AL.(1993) GENOMICS 15,376-386.
13	D13S127	1341	0.61		130	142	GDB
13	D13S128	NA	0.90		144	178	GENOMICS 8:400- , 1990
13	D13S129	NA	0.88		414	434	GENOMICS 8:400- , 1990
13	D13S130	NA	0.85		165	189	PHILLIPS,H. ET AL(1991)NAR 19,6664.
13	D13S131	NA	0.71		156	180	PETRUKHIN,K. ET AL. (1993) GENOMICS15,76-85.
13	D13S132	NA	0.84		141	163	GENOMICS 8:400- , 1990
13	D13S133	ca006	0.83		130	187	PHILLIPS,H. ET AL(1991)NAR 19,6664.
13	D13S134	NA	0.82		168	188	GENOMICS 8:400- , 1990
13	D13S135	NA	0.82		168	184	GENOMICS 8:400- , 1990
13	D13S136	NA	0.80		124	152	GENOMICS 8:400- , 1990
13	D13S137	ca010	0.72		113	135	PHILLIPS,H. ET AL(1991)NAR 19,6664.
13	D13S138	ca011	0.69		102	110	PHILLIPS,H. ET AL(1991)NAR 19,6664.
13	D13S139	ca012	0.68		127	143	PHILLIPS,H. ET AL(1991)NAR 19,6664.
13	D13S140	NA	0.59		120	138	GENOMICS 8:400- , 1990
13	D13S141	CU13	0.60		115	131	GDB
13	D13S142	NA	0.52		163	171	PHILLIPS,H. ET AL(1991)NAR 19,6664.
13	D13S144	Utsw1348	0.72	0.76	183	199	BOWCOCK,A. ET AL.(1993) GENOMICS 15,376-386.
13	D13S147	NA	0.90		93	107	GENOMICS 8:400- , 1990
13	D13S148	NA	0.82		123	139	GENOMICS 8:400- , 1990
13	D13S152	AFM037xa1	0.78		133	143	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
13	D13S153	AFM058xd6	0.82		212	236	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
13	D13S154	AFM065xe9	0.00		243	277	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
13	D13S155	AFM070xb7	0.83		204	218	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
13	D13S156	AFM093ye1	0.81		272	286	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
13	D13S157	AFM095wh3	0.73		250	264	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
13	D13S158	AFM135xb8	0.82		99	113	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
13	D13S159	AFM155yb6	0.91		169	203	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
13	D13S160	AFM157xa11	0.81		229	241	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
13	D13S161	AFM184xf12	0.58		88	100	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
13	D13S162	AFM190xa1	0.86		182	202	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
13	D13S163	AFM190ya3	0.70		198	204	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
13	D13S164	AFM198yd4	0.73		208	219	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
13	D13S165	AFM203xa1	0.75		183	195	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
13	D13S166	AFM205wg3	0.76		115	125	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
13	D13S167	AFM207wb8	0.59		184	192	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
13	D13S168	AFM210wa5	0.77		173	197	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
13	D13S169	AFM210xc11	0.48		183	189	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
13	D13S170	AFM240wh2	0.90		113	137	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
13	D13S171	AFM255ze9	0.73		227	241	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
13	D13S172	AFM259zg1	0.63		131	141	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
13	D13S173	AFM261yg5	0.83		166	178	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
13	D13S174	AFM248tf1	0.00		175	199	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
13	D13S175	AFM249xb1	0.77		101	113	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
13	D13S176	AFM263yd1	0.80		211	227	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
13	D13S192	HKCA3	0.90		88	124	HONG,H. ET AL.(1993) HMG 2,86.
13	D13S193	HKCA5	0.75		129	149	HUDSON,T. ET AL.(1992) GENOMICS 13,622-29.
13	D13S197	HKCA1	0.86		97	142	HONG,H. ET AL.(1993) HMG 2,337.
13	D13S217	AFM205xh12	0.68		160	174	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
13	D13S218	AFM210zb2	0.68		187	195	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
13	D13S219	AFM225xe5	0.64		117	127	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801

CH	LOCUS	ASSAY	HET	PIC	SIZE MIN	ANGE MAX	REFERENCE
13	D13S220	AFM234yb8	0.67		191	203	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
13	D13S221	AFM248wc1	0.83		223	243	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
13	D13S225	MFD250	0.80	0.77	150	168	GENOMICS 8:400- , 1990
13	D13S231	NA	0.75		208	224	SAKSOVA, L. ET AL. (1993) HMG 2, 1082.
13	D13S232	MFD299	0.76	0.73	108	126	GENOMICS 8:400- , 1990
13	D13S233	MFD262	0.75	0.72	89	109	GENOMICS 8:400- , 1990
13	D13S251	UT 329	0.75		212	212	GDB
13	D13S259	AFM164zf4	0.75		139	155	WEISSENBACH J: NATURE GENETIC, JUNE 1994
13	D13S260	AFM177xf4	0.41		158	173	WEISSENBACH J: NATURE GENETIC, JUNE 1994
13	D13S261	AFM200yh8	0.64		166	172	WEISSENBACH J: NATURE GENETIC, JUNE 1994
13	D13S262	AFM205vh2	0.84		285	303	WEISSENBACH J: NATURE GENETIC, JUNE 1994
13	D13S263	AFM210yg11	0.77		145	165	WEISSENBACH J: NATURE GENETIC, JUNE 1994
13	D13S264	AFM218xd12	0.78		103	117	WEISSENBACH J: NATURE GENETIC, JUNE 1994
13	D13S265	AFM218yf10	0.65		104	132	WEISSENBACH J: NATURE GENETIC, JUNE 1994
13	D13S266	AFM224za11	0.64		125	135	WEISSENBACH J: NATURE GENETIC, JUNE 1994
13	D13S267	AFM234ve3	0.75		148	162	WEISSENBACH J: NATURE GENETIC, JUNE 1994
13	D13S268	AFM238vc3	0.72		296	312	WEISSENBACH J: NATURE GENETIC, JUNE 1994
13	D13S269	AFM240xg5	0.34		116	140	WEISSENBACH J: NATURE GENETIC, JUNE 1994
13	D13S270	AFM084xc5	0.20		79	99	WEISSENBACH J: NATURE GENETIC, JUNE 1994
13	D13S271	AFM115xa11	0.80		120	142	WEISSENBACH J: NATURE GENETIC, JUNE 1994
13	D13S272	AFM120xa3	0.86		131	143	WEISSENBACH J: NATURE GENETIC, JUNE 1994
13	D13S273	AFM126za5	0.49		235	259	WEISSENBACH J: NATURE GENETIC, JUNE 1994
13	D13S274	AFM126zg5	0.55		208	216	WEISSENBACH J: NATURE GENETIC, JUNE 1994
13	D13S275	AFM263zb5	0.65		236	242	WEISSENBACH J: NATURE GENETIC, JUNE 1994
13	D13S276	AFM269yb1	0.75		203	219	WEISSENBACH J: NATURE GENETIC, JUNE 1994
13	D13S277	AFM282wa5	0.63		171	173	WEISSENBACH J: NATURE GENETIC, JUNE 1994
13	D13S278	AFM283wf5	0.43		195	215	WEISSENBACH J: NATURE GENETIC, JUNE 1994
13	D13S279	AFM284za9	0.57		241	257	WEISSENBACH J: NATURE GENETIC, JUNE 1994
13	D13S280	AFM286zh5	0.60		226	256	WEISSENBACH J: NATURE GENETIC, JUNE 1994
13	D13S281	AFM288yc1	0.62		170	176	WEISSENBACH J: NATURE GENETIC, JUNE 1994
13	D13S282	AFM289za9	0.76		235	241	WEISSENBACH J: NATURE GENETIC, JUNE 1994
13	D13S283	AFM296vb9	0.68		128	155	WEISSENBACH J: NATURE GENETIC, JUNE 1994
13	D13S284	AFM303zg5	0.81		197	227	WEISSENBACH J: NATURE GENETIC, JUNE 1994
13	D13S285	AFM309va9	0.81		92	106	WEISSENBACH J: NATURE GENETIC, JUNE 1994
13	D13S286	AFM310yd5	0.62		175	195	WEISSENBACH J: NATURE GENETIC, JUNE 1994
13	D13S287	AFM317wa9	0.76		139	153	WEISSENBACH J: NATURE GENETIC, JUNE 1994
13	D13S288	AFM318xa9	0.80		224	228	WEISSENBACH J: NATURE GENETIC, JUNE 1994
13	D13S289	AFM321xb1	0.81		260	276	WEISSENBACH J: NATURE GENETIC, JUNE 1994
13	D13S290	AFM323vh1	0.76		176	194	WEISSENBACH J: NATURE GENETIC, JUNE 1994
13	D13S291	AFM330ya5	0.60		241	253	WEISSENBACH J: NATURE GENETIC, JUNE 1994
13	D13S292	AFM351xd9	0.85		201	207	WEISSENBACH J: NATURE GENETIC, JUNE 1994
13	D13S293	AFMa127xh5	0.76		92	96	WEISSENBACH J: NATURE GENETIC, JUNE 1994
13	D13S308	EST 00444	0.71		147	147	GDB
13	D13S309	AD2C	0.81	0.79	165	185	GDB
13	D13S310	AD13J	0.70	0.65	142	162	GDB
13	D13S312	AD11C	0.84	0.82	148	170	GDB
13	D13S313	AD23J	0.79	0.75	126	138	GDB
13	D13S317	GATA7G10	0.87		175	199	GDB
13	D13S318	GATA5H09	0.83		283	283	GDB
13	D13S321	GATA10D05	0.79		186	200	GDB
13	D13S328	AFM098xd3	0.57		265	283	WEISSENBACH J: NATURE GENETIC, JUNE 1994
13	D13S64	NA	0.70		174	190	PHILLIPS,H. ET AL(1991)NAR 19,6664.
13	D13S71	MFD 44	0.76	0.67	67	79	NAR 18(15):4638, 1990
13	D13S762	GATA30G10	1.00		315	315	GDB
13	D13S763	GATA44G10	0.50		221	221	GDB
13	D13S764	GATA46C12	1.00		323	323	GDB
13	D13S765	GATA47C08	0.75		192	192	GDB
13	D13S766	GATA48F08	0.45		128	128	GDB
13	D13S767	GATA50D02	0.83		158	170	GDB
13	D13S768	GGAT2A09	0.57		144	156	GDB
13	ERCC5	NA	0.84		245	277	GDB
14	AACT	NA	0.72		999	999	BYTH,B & COX,D (1993) HMG 2, 1085.
14	AACT	NA	0.72		141	173	GDB
14	D14S118	UT562	0.00		248	274	GDB
14	D14S241	ms54	0.69		183	203	IIZUKA, M. ET AL. (1993) HMG 2, 1979.
14	D14S244	ms46	0.75		215	237	IIZUKA, M. ET AL. (1993) HMG 2, 1979.
14	D14S245	ms68	0.81		164	176	IIZUKA, M. ET AL. (1993) HMG 2, 1979.
14	D14S250	AFM203ve5	0.56		157	179	WEISSENBACH J: NATURE GENETIC, JUNE 1994
14	D14S251	AFM203zb2	0.81		298	318	WEISSENBACH J: NATURE GENETIC, JUNE 1994
14	D14S252	AFM207xb12	0.66		247	259	WEISSENBACH J: NATURE GENETIC, JUNE 1994

CH	LOCUS	ASSAY	HET	PIC	SIZE MIN	ANGE MAX	REFERENCE
14	D14S253	AFM212ze3	0.67		141	151	WEISSENBACH J: NATURE GENETIC, JUNE 1994
14	D14S254	AFM214xf8	0.51		123	159	WEISSENBACH J: NATURE GENETIC, JUNE 1994
14	D14S255	AFM220zh4	0.61		197	207	WEISSENBACH J: NATURE GENETIC, JUNE 1994
14	D14S256	AFM224xb4	0.27		134	156	WEISSENBACH J: NATURE GENETIC, JUNE 1994
14	D14S257	AFM224yb8	0.68		174	194	WEISSENBACH J: NATURE GENETIC, JUNE 1994
14	D14S258	AFM224zf12	0.21		170	182	WEISSENBACH J: NATURE GENETIC, JUNE 1994
14	D14S259	AFM225yb10	0.56		252	270	WEISSENBACH J: NATURE GENETIC, JUNE 1994
14	D14S260	AFM238we7	0.59		188	202	WEISSENBACH J: NATURE GENETIC, JUNE 1994
14	D14S261	AFM238yd6	0.65		169	199	WEISSENBACH J: NATURE GENETIC, JUNE 1994
14	D14S262	AFM240vc5	0.83		196	206	WEISSENBACH J: NATURE GENETIC, JUNE 1994
14	D14S263	AFM254xa9	0.66		148	158	WEISSENBACH J: NATURE GENETIC, JUNE 1994
14	D14S264	AFM084ya1	0.73		216	234	WEISSENBACH J: NATURE GENETIC, JUNE 1994
14	D14S265	AFM136yb4	0.26		148	168	WEISSENBACH J: NATURE GENETIC, JUNE 1994
14	D14S266	AFM150xa7	0.81		132	138	WEISSENBACH J: NATURE GENETIC, JUNE 1994
14	D14S267	AFM263wh9	0.69		193	225	WEISSENBACH J: NATURE GENETIC, JUNE 1994
14	D14S268	AFM265vb9	0.73		91	128	WEISSENBACH J: NATURE GENETIC, JUNE 1994
14	D14S269	AFM267zd5	0.57		213	229	WEISSENBACH J: NATURE GENETIC, JUNE 1994
14	D14S270	AFM270zh1	0.49		214	224	WEISSENBACH J: NATURE GENETIC, JUNE 1994
14	D14S271	AFM277xe9	0.69		226	246	WEISSENBACH J: NATURE GENETIC, JUNE 1994
14	D14S272	AFM281wg1	0.76		224	232	WEISSENBACH J: NATURE GENETIC, JUNE 1994
14	D14S273	AFM283va9	0.59		179	199	WEISSENBACH J: NATURE GENETIC, JUNE 1994
14	D14S274	AFM288vg1	0.63		114	134	WEISSENBACH J: NATURE GENETIC, JUNE 1994
14	D14S275	AFM291za9	0.67		195	205	WEISSENBACH J: NATURE GENETIC, JUNE 1994
14	D14S276	AFM292we1	0.70		86	98	WEISSENBACH J: NATURE GENETIC, JUNE 1994
14	D14S277	AFM295zd5	0.76		140	158	WEISSENBACH J: NATURE GENETIC, JUNE 1994
14	D14S278	AFM296zd5	0.51		150	158	WEISSENBACH J: NATURE GENETIC, JUNE 1994
14	D14S279	AFM301wh9	0.62		196	212	WEISSENBACH J: NATURE GENETIC, JUNE 1994
14	D14S280	AFM304ya5	0.75		229	241	WEISSENBACH J: NATURE GENETIC, JUNE 1994
14	D14S281	AFM309xh5	0.75		173	181	WEISSENBACH J: NATURE GENETIC, JUNE 1994
14	D14S282	AFM311wd1	0.64		164	174	WEISSENBACH J: NATURE GENETIC, JUNE 1994
14	D14S283	AFM312xh1	0.82		125	153	WEISSENBACH J: NATURE GENETIC, JUNE 1994
14	D14S284	AFM318th1	0.79		156	180	WEISSENBACH J: NATURE GENETIC, JUNE 1994
14	D14S285	AFM319vf5	0.31		171	191	WEISSENBACH J: NATURE GENETIC, JUNE 1994
14	D14S286	AFM324tb1	0.81		157	195	WEISSENBACH J: NATURE GENETIC, JUNE 1994
14	D14S287	AFM324va9	0.79		240	254	WEISSENBACH J: NATURE GENETIC, JUNE 1994
14	D14S288	AFM328yc5	0.82		189	209	WEISSENBACH J: NATURE GENETIC, JUNE 1994
14	D14S289	AFM330za9	0.78		192	206	WEISSENBACH J: NATURE GENETIC, JUNE 1994
14	D14S290	AFM343ze5	0.81		233	253	WEISSENBACH J: NATURE GENETIC, JUNE 1994
14	D14S291	AFMa120xc1	0.68		210	216	WEISSENBACH J: NATURE GENETIC, JUNE 1994
14	D14S292	AFMa120xg5	0.40		110	118	WEISSENBACH J: NATURE GENETIC, JUNE 1994
14	D14S293	AFMa143xb5	0.64		153	165	WEISSENBACH J: NATURE GENETIC, JUNE 1994
14	D14S297	GATA5H04	0.68		99	99	GDB
14	D14S298	GATA2B03	0.62		201	201	GDB
14	D14S299	wg1c5	0.86		294	318	ARMOUR, J. ET AL. (1994) HMG 3, 599-605.
14	D14S301	GATA10H04	0.75		265	277	GDB
14	D14S302	GATA13B06	0.62		178	182	GDB
14	D14S304	GGAT2D06			249	249	GDB
14	D14S306	GATA4B04	0.85		190	210	GDB
14	D14S34	MFD 42	0.59	0.59	107	117	NAR 18(15):4640, 1990
14	D14S42	NA		0.65	115	133	JORDAN,S.A.ET AL.(1991)NAR 19,1171.
14	D14S43	NA		0.72	152	190	SHARMA,V. ET AL.(1992) NAR 19,1722.
14	D14S45	NA	0.79	0.74	79	95	LUTY,J.& LITT,M. (1991) NAR 19,4308.
14	D14S47	MFD 86	0.70	0.58	75	89	GENOMICS 13:532-536, 1992
14	D14S48	MFD101	0.82	0.73	259	277	GENOMICS 14:209-219, 1992
14	D14S49	MFD119	0.81	0.84	168	179	GENOMICS 14:209-219, 1992
14	D14S50	MFD130	0.77	0.73	166	180	GENOMICS 14:209-219, 1992
14	D14S51	MFD165	0.77	0.80	131	149	GENOMICS 14:209-219, 1992
14	D14S52	MFD167	0.68	0.58	79	99	GENOMICS 14:209-219, 1992
14	D14S53	MFD190	0.71	0.68	135	161	GENOMICS 14:209-219, 1992
14	D14S54	MFD192	0.77	0.72	254	258	GENOMICS 14:209-219, 1992
14	D14S55	MFD198	0.48	0.51	123	129	GENOMICS 14:209-219, 1992
14	D14S57	MIT-MS16	0.63		135	151	HUMAN GENET 87:401, 1991
14	D14S579	ATA3B06	0.73		292	316	GDB
14	D14S58	MIT-MS162	0.63		205	205	HUMAN GENET 87:401, 1991
14	D14S580	ATA4B10	0.45		208	232	GDB
14	D14S581	GAAT1B03	0.50		191	199	GDB
14	D14S582	GATA29G12	0.62		204	204	GDB
14	D14S583	GATA46A06	0.92		265	265	GDB
14	D14S584	GATA48E09	0.69		284	284	GDB
14	D14S585	GATA49B10	0.57		247	247	GDB

CH	LOCUS	ASSAY	HET	PIC	SIZE MIN	ANGE MAX	REFERENCE
14	D14S586	GATA51F02	0.58		214	214	GDB
14	D14S587	GGAA10C09	0.94		263	263	GDB
14	D14S588	GGAA4A12	0.69		117	141	GDB
14	D14S59	MIT-MH90	0.88		99	111	HUMAN GENET 87:401, 1991
14	D14S61	AFM025tc9	0.81		197	227	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
14	D14S62	AFM027xb3	67.00		118	127	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
14	D14S63	AFM058yh2	0.77		199	217	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
14	D14S64	AFM079ze5	0.77		126	136	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
14	D14S65	AFM093yg5	0.80		125	149	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
14	D14S66	AFM109ya3	0.67		186	194	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
14	D14S67	AFM137xh12	0.00		133	167	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
14	D14S68	AFM164tb12	0.89		148	203	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
14	D14S69	AFM164xf10	0.70		205	213	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
14	D14S70	AFM191ve1	0.77		212	220	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
14	D14S71	AFM197xf12	0.75		191	211	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
14	D14S72	AFM199zf4	0.83		257	273	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
14	D14S73	AFM203za11	0.63		99	119	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
14	D14S74	AFM210zh4	0.80		291	313	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
14	D14S75	AFM214yg5	0.77		184	202	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
14	D14S76	AFM214zg3	0.70		167	207	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
14	D14S77	AFM218zh4	0.94		207	247	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
14	D14S78	AFM234we5	0.68		211	233	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
14	D14S79	AFM240zd4	0.67		81	89	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
14	D14S80	AFM242xa9	0.84		132	160	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
14	D14S81	AFM260xb1	0.84		175	209	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
14	D14S99E	NA	0.75		142	168	GDB
14	MYH6	NA	0.81	0.79	108	132	FOUGEROUSSE,F. ET AL.(1992) HMG 1,64.
14	MYH7	NA	0.82	0.80	130	158	WARLICK,C. ET AL.(1992) HMG 1, 136.
14	P1	NA	0.65		101	113	ROGEAV,E.&KERYANOV,S.(1992) HMG 1,657.
14	P1-5	NA	0.90		155	189	GDB
14	PCI	NA	0.80		115	139	GDB
14	PCI-II	NA	0.80		128	156	GDB
14	PI-1	NA	0.83		225	239	GDB
14	SSTR1	NA	0.84		185	211	GDB
14	TCRA	NA	0.77		186	200	CORNELIS,F.ET AL.(1992) GENOMICS13,820-825.
14	TCRD	NA	0.00	0.74	118	128	JORDAN,S.A.ET AL.(1991)NAR 19,1959.
15	ACTC	NA	0.41		219	239	WATKINS,C. ET AL(1991) NAR 19,6980.
15	AS/PWS	NA	0.74	0.70	243	263	MUTIRANGURA,A. ET AL(1992)HMG 1,139.
15	CYP19	NA	0.91		168	192	POLYMEROPOULOS,M. ET AL.(1991) NAR 19,195.
15	D15S100	MIT-MS164	0.79		183	183	HUMAN GENET 87:401, 1991
15	D15S101	MIT-MS178	0.83		104	104	HUMAN GENET 87:401, 1991
15	D15S102	MIT-N130	0.85		217	217	HUMAN GENET 87:401, 1991
15	D15S103	MIT-G113	0.58		233	233	HUMAN GENET 87:401, 1991
15	D15S104	MIT-M131	0.83		170	170	HUMAN GENET 87:401, 1991
15	D15S106	MFD 81	0.24	0.21	101	109	GENOMICS 8:400- , 1990
15	D15S107	MFD 87	0.77	0.63	132	146	JONES,M. ET AL.(1992) HMG 1,131-33.
15	D15S108	MFD102	0.52	0.52	185	205	JONES,M. ET AL.(1992) HMG 1,131-33.
15	D15S11	NA	0.74		243	263	GDB
15	D15S111	Utws 1513	0.75		143	157	BECKMAN, J. ET AL. (1993) HMG 2, 2019-2030.
15	D15S112	Utsw 1547	0.70		133	151	BECKMAN, J. ET AL. (1993) HMG 2, 2019-2030.
15	D15S113	NA	0.73		130	140	SUBMITTED
15	D15S114	AFM019tc9	0.71		177	187	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
15	D15S115	AFM029yg1	0.55		172	200	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
15	D15S116	AFM078zf7	0.84		164	184	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
15	D15S117	AFM098yg1	0.79		132	150	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
15	D15S118	AFM112xa1	0.76		218	230	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
15	D15S119	AFM150xf4	0.73		185	197	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
15	D15S120	AFM164zc9	0.75		150	174	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
15	D15S121	AFM189yc1	0.67		258	264	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
15	D15S122	AFM200wb4	0.79		143	159	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
15	D15S123	AFM205ye1	0.81		191	207	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
15	D15S124	AFM207xa3	0.35		99	107	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
15	D15S125	AFM214xd10	0.80		157	169	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
15	D15S126	AFM218yf12	0.83		188	218	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
15	D15S127	AFM224xe11	0.87		114	147	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
15	D15S128	AFM273yf9	0.79		193	209	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
15	D15S129	AFM-Z1280	0.84		208	224	GDB
15	D15S130	AFM072yb11	0.71		218	232	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
15	D15S131	AFM262xb1	0.84		238	274	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
15	D15S132	AFM265xf9	0.76		69	83	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801

CH	LOCUS	ASSAY	HET	PIC	SIZE ANGE MIN	MAX	REFERENCE
15	D15S143	AFM016yg1	0.80		189	199	WEISSENBACH J: NATURE GENETIC, JUNE 1994
15	D15S144	AFM019td6	0.85		156	170	WEISSENBACH J: NATURE GENETIC, JUNE 1994
15	D15S145	AFM046xc9	0.78		84	104	WEISSENBACH J: NATURE GENETIC, JUNE 1994
15	D15S148	AFM135yc1	0.72		134	142	WEISSENBACH J: NATURE GENETIC, JUNE 1994
15	D15S151	AFM182yb4	0.73		253	261	WEISSENBACH J: NATURE GENETIC, JUNE 1994
15	D15S152	AFM200zg11	0.78		200	234	WEISSENBACH J: NATURE GENETIC, JUNE 1994
15	D15S153	AFM205ye3	0.68		194	226	WEISSENBACH J: NATURE GENETIC, JUNE 1994
15	D15S155	AFM211zc1	0.74		237	269	WEISSENBACH J: NATURE GENETIC, JUNE 1994
15	D15S156	AFM214xg11	0.54		217	229	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
15	D15S157	AFM217zg1	0.23		115	125	WEISSENBACH J: NATURE GENETIC, JUNE 1994
15	D15S158	AFM234vf12	0.75		85	99	WEISSENBACH J: NATURE GENETIC, JUNE 1994
15	D15S159	AFM234zd6	0.81		163	169	WEISSENBACH J: NATURE GENETIC, JUNE 1994
15	D15S160	AFM248yh1	0.66		143	159	WEISSENBACH J: NATURE GENETIC, JUNE 1994
15	D15S161	AFM261xb9	0.82		277	291	WEISSENBACH J: NATURE GENETIC, JUNE 1994
15	D15S165	AFM248vc5	0.80		184	208	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
15	D15S169	UTSW1591	0.77		142	158	GDB
15	D15S170	Utsw 1560	0.69		216	226	BECKMAN, J. ET AL. (1993) HMG 2, 2019-2030.
15	D15S172	MH22	0.68		123	159	BECKMAN, J. ET AL. (1993) HMG 2, 2019-2030.
15	D15S175	MFD288	0.78	0.74	165	177	GENOMICS 8:400- , 1990
15	D15S197	AFM200ve9	0.71		144	148	WEISSENBACH J: NATURE GENETIC, JUNE 1994
15	D15S198	AFM234ye11	0.78		204	216	WEISSENBACH J: NATURE GENETIC, JUNE 1994
15	D15S199	AFM107xg7	0.73		100	108	WEISSENBACH J: NATURE GENETIC, JUNE 1994
15	D15S200	AFM265vb1	0.75		190	208	WEISSENBACH J: NATURE GENETIC, JUNE 1994
15	D15S201	AFM281yh1	0.71		212	224	WEISSENBACH J: NATURE GENETIC, JUNE 1994
15	D15S202	AFM282wg5	0.66		226	247	WEISSENBACH J: NATURE GENETIC, JUNE 1994
15	D15S203	AFM286zb5	0.87		116	140	WEISSENBACH J: NATURE GENETIC, JUNE 1994
15	D15S204	AFM290ya5	0.47		116	134	WEISSENBACH J: NATURE GENETIC, JUNE 1994
15	D15S205	AFM291zh5	0.81		128	170	WEISSENBACH J: NATURE GENETIC, JUNE 1994
15	D15S206	AFM299yf9	0.68		267	279	WEISSENBACH J: NATURE GENETIC, JUNE 1994
15	D15S207	AFM309vg9	0.78		160	174	WEISSENBACH J: NATURE GENETIC, JUNE 1994
15	D15S208	AFM310we1	0.82		224	230	WEISSENBACH J: NATURE GENETIC, JUNE 1994
15	D15S209	AFM312wd1	0.56		190	212	WEISSENBACH J: NATURE GENETIC, JUNE 1994
15	D15S210	AFM320vd9	0.64		124	146	WEISSENBACH J: NATURE GENETIC, JUNE 1994
15	D15S211	AFM323yd9	0.67		207	259	WEISSENBACH J: NATURE GENETIC, JUNE 1994
15	D15S212	AFM331vb5	0.71		203	217	WEISSENBACH J: NATURE GENETIC, JUNE 1994
15	D15S213	AFM350vh1	0.75		186	194	WEISSENBACH J: NATURE GENETIC, JUNE 1994
15	D15S214	AFMa123xc5	0.76		266	276	WEISSENBACH J: NATURE GENETIC, JUNE 1994
15	D15S215	AFMa132yb1	0.73		243	249	WEISSENBACH J: NATURE GENETIC, JUNE 1994
15	D15S216	AFMa143xc5	0.77		225	233	WEISSENBACH J: NATURE GENETIC, JUNE 1994
15	D15S217	GATA8B06	0.91		191	191	GDB
15	D15S219	MFD209	0.48	0.46	145	185	GENOMICS 8:400- , 1990
15	D15S220	MFD351	0.60	0.52	139	145	GENOMICS 8:400- , 1990
15	D15S221	M770F4-12	0.77		149	173	GDB
15	D15S222	T38	0.78		184	198	GDB
15	D15S230	GATA8C05	0.89		196	196	GDB
15	D15S231	GTAT1B2	0.44		100	112	GDB
15	D15S233	wg1d1	0.78		186	218	ARMOUR, J. ET AL. (1994) HMG 3, 599-605.
15	D15S234	M770F4-12	0.85		149	173	GDB
15	D15S519	CYP11A	0.63		170	180	BECKMAN, J. ET AL. (1993) HMG 2, 2019-2030.
15	D15S641	ATA3E11	0.53		96	108	GDB
15	D15S642	GATA27A03	0.87		200	218	GDB
15	D15S643	GATA50G06	0.80		211	211	GDB
15	D15S644	GGAA5F05	0.73		204	204	GDB
15	D15S87	MFD 49	0.87	0.85	78	98	NAR 18(15):4640, 1990
15	D15S95	MX8	0.79		140	150	BECKMAN, J. ET AL. (1993) HMG 2, 2019-2030.
15	D15S96	G127	0.83		148	174	BECKMAN, J. ET AL. (1993) HMG 2, 2019-2030.
15	D15S97	MIT-MS14	0.75		159	159	HUMAN GENET 87:401, 1991
15	D15S98	MIT-MS112	0.81		152	152	HUMAN GENET 87:401, 1991
15	D15S99	MIT-MS149	0.50		135	135	HUMAN GENET 87:401, 1991
15	FES	NA	0.75	0.70	142	158	POLYMEROPOULOS,M. ET AL.(1991) NAR 19,4018.
15	GABRA5	NA	0.78		137	149	GLATT,K. ET AL.(1992) HMG 1,348.
15	GABRB3	NA	0.82	0.83	181	201	MUTIRANGURA,A. ET AL(1992)HMG 1,67.
15	IGF1R	NA	0.44	0.35	90	96	MELONI,R. ET AL(1992)NAR 20,1427.
16	D16S186	NA	0.57		130	178	PHILLIPS,H. ET AL(1991)NAR 19,6964.
16	D16S260	MFD 12	0.57	0.43	234	240	NAR 18(13):4036, 1990
16	D16S261	MFD 24	0.67	0.66	88	100	NAR 18(13):4034, 1990
16	D16S265	MFD 23	0.73	0.75	89	117	NAR 18(13):4036, 1990
16	D16S266	MFD 62	0.59	0.54	94	104	NAR 18(13):4036, 1990
16	D16S267	MFD 65	0.47	0.45	154	170	NAR 18(13):4036, 1990
16	D16S283	NA	0.63		81	107	HARRIS,P. ET AL.,(1991) LANCET 338,1484-86.

CH	LOCUS	ASSAY	HET	PIC	SIZE MIN	ANGE MAX	REFERENCE
16	D16S285	NA	0.82		105	129	KONARDI,C. ET AL.(1991) NAR 19,5449.
16	D16S287	NA	0.78		201	225	PHILLIPS,H. ET AL(1991)NAR 19,6664.
16	D16S288	NA	0.73	0.69	154	166	SHEN,Y. ET AL.(1991) NAR 19,5545.
16	D16S289	NA	0.77		156	172	GENOMICS 8:400- ,1990
16	D16S291	NA	0.79	0.76	154	170	THOMPSON,A. ET AL.(1992) GENOMICS13,402-08.
16	D16S292	NA	0.74	0.71	180	198	TROFATTER,J.A. ET AL. (1991) NAR19,2802.
16	D16S295	16AC62F3	0.66		110	124	GDB
16	D16S298	NA	0.79	0.77	172	192	TROFATTER,J.A. ET AL. (1991) NAR19,2802.
16	D16S299	NA	0.72	0.69	110	124	TROFATTER,J.A. ET AL. (1991) NAR19,2802.
16	D16S305	NA	0.82	0.80	172	200	TROFATTER,J.A. ET AL. (1991) NAR19,2802.
16	D16S308	NA	0.77	0.73	137	167	TROFATTER,J.A. ET AL. (1991) NAR19,2802.
16	D16S310	MIT-MH20	0.67		162	162	HUMAN GENET 87:401, 1991
16	D16S312	MIT-I103	0.75		176	176	HUMAN GENET 87:401, 1991
16	D16S313	MIT-MS79	0.57		141	141	HUMAN GENET 87:401, 1991
16	D16S320	NA	0.86		151	175	SHEN,Y. ET AL.(1992) HMG 1,773.
16	D16S347	16AC12F8	0.76		186	206	GDB
16	D16S363	A6AC51G1	0.78		242	258	GDB
16	D16S389	16AC10B3	0.77		90	114	GDB
16	D16S390	16AC10F5	0.80		177	197	GDB
16	D16S392	16AC305E9	0.78		113	151	GDB
16	D16S393	16AC323H4	0.87		130	158	GDB
16	D16S395	16AC33G11	0.69		98	124	GDB
16	D16S397	MFD 98	0.70		104	112	GENOMICS 8:400- ,1990
16	D16S398	MFD168	0.90		180	196	GENOMICS 8:400- ,1990
16	D16S400	AFM024xg1	0.62		192	202	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
16	D16S401	AFM025tg9	0.00		166	180	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
16	D16S402	AFM031xa5	0.87		161	187	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
16	D16S403	AFM049xd2	0.86		134	152	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
16	D16S404	AFM056yf6	0.82		117	137	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
16	D16S405	AFM070ya1	0.78		116	150	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
16	D16S406	AFM079yh3	0.82		180	198	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
16	D16S407	AFM113xe3	0.86		150	170	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
16	D16S408	AFM137xf8	0.69		241	251	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
16	D16S409	AFM161xa1	0.71		135	147	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
16	D16S410	AFM165yb6	0.57		134	140	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
16	D16S411	AFM186xa3	0.79		209	229	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
16	D16S412	AFM191wb10	0.76		101	125	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
16	D16S413	AFM196xg1	0.85		131	149	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
16	D16S414	AFM205za11	0.61		152	161	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
16	D16S415	AFM205ze5	0.74		208	234	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
16	D16S416	AFM210yg3	0.43		217	223	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
16	D16S417	AFM220xb10	0.73		124	142	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
16	D16S418	AFM225xd2	0.83		166	188	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
16	D16S419	AFM225zf2	0.76		146	164	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
16	D16S420	AFM238xb2	0.82		179	201	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
16	D16S421	AFM240yh6	0.57		206	216	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
16	D16S422	AFM249xc5	0.80		188	212	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
16	D16S423	AFM249yc5	0.75		121	139	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
16	D16S446	MFD272	0.31	0.28	225	229	GENOMICS 8:400- ,1990
16	D16S449	16AC51A4	0.85		183	201	GDB
16	D16S451	16AC69F12	0.84		83	113	GDB
16	D16S452	16AC33A4	0.68		132	144	GDB
16	D16S454	16AC45G5	0.75		175	193	GDB
16	D16S468	C28	0.90		198	208	GDB
16	D16S494	AFM193xh10	0.74		171	195	WEISSENBACH J: NATURE GENETIC, JUNE 1994
16	D16S495	AFM199zb10	0.54		136	142	WEISSENBACH J: NATURE GENETIC, JUNE 1994
16	D16S496	AFM214zg5	0.37		209	226	WEISSENBACH J: NATURE GENETIC, JUNE 1994
16	D16S497	AFM218ya11	0.68		150	174	WEISSENBACH J: NATURE GENETIC, JUNE 1994
16	D16S498	AFM218yb10	0.86		208	246	WEISSENBACH J: NATURE GENETIC, JUNE 1994
16	D16S499	AFM259xb9	0.80		209	217	WEISSENBACH J: NATURE GENETIC, JUNE 1994
16	D16S500	AFM112xg5	0.56		187	203	WEISSENBACH J: NATURE GENETIC, JUNE 1994
16	D16S501	AFM113xa9	0.70		122	130	WEISSENBACH J: NATURE GENETIC, JUNE 1994
16	D16S502	AFM266xg9	0.79		253	265	WEISSENBACH J: NATURE GENETIC, JUNE 1994
16	D16S503	AFM274ya5	0.80		221	235	WEISSENBACH J: NATURE GENETIC, JUNE 1994
16	D16S504	AFM292xh5	0.73		268	284	WEISSENBACH J: NATURE GENETIC, JUNE 1994
16	D16S505	AFM296tb1	0.74		239	261	WEISSENBACH J: NATURE GENETIC, JUNE 1994
16	D16S506	AFM297yg5	0.75		146	162	WEISSENBACH J: NATURE GENETIC, JUNE 1994
16	D16S507	AFM301zb9	0.74		175	195	WEISSENBACH J: NATURE GENETIC, JUNE 1994
16	D16S508	AFM304xf1	0.80		89	93	WEISSENBACH J: NATURE GENETIC, JUNE 1994
16	D16S509	AFM308yf9	0.84		207	223	WEISSENBACH J: NATURE GENETIC, JUNE 1994

CH	LOCUS	ASSAY	HET	PIC	SIZE ANGE MIN	MAX	REFERENCE
16	D16S510	AFM312vd5	0.24		271	287	WEISSENBACH J: NATURE GENETIC, JUNE 1994
16	D16S511	AFM312xd1	0.82		182	222	WEISSENBACH J: NATURE GENETIC, JUNE 1994
16	D16S512	AFM320wf1	0.76		201	211	WEISSENBACH J: NATURE GENETIC, JUNE 1994
16	D16S513	AFM321th5	0.74		244	274	WEISSENBACH J: NATURE GENETIC, JUNE 1994
16	D16S514	AFM330vd9	0.71		117	129	WEISSENBACH J: NATURE GENETIC, JUNE 1994
16	D16S515	AFM340ye5	0.82		222	244	WEISSENBACH J: NATURE GENETIC, JUNE 1994
16	D16S516	AFM350vd1	0.60		164	176	WEISSENBACH J: NATURE GENETIC, JUNE 1994
16	D16S517	AFMa132we9	0.50		257	277	WEISSENBACH J: NATURE GENETIC, JUNE 1994
16	D16S518	AFMa132xg9	0.66		272	290	WEISSENBACH J: NATURE GENETIC, JUNE 1994
16	D16S519	AFMa133xd5	0.77		135	157	WEISSENBACH J: NATURE GENETIC, JUNE 1994
16	D16S520	AFMa135xg5	0.87		181	197	WEISSENBACH J: NATURE GENETIC, JUNE 1994
16	D16S521	AFMa139wg1	0.58		156	172	WEISSENBACH J: NATURE GENETIC, JUNE 1994
16	D16S522	16AC8.21	0.69		103	119	GDB
16	D16S523	16AC13H1	0.68		77	87	GDB
16	D16S524	16AC40A7	0.76		143	169	GDB
16	D16S525	16AC308G7	0.91		143	175	GDB
16	D16S531	16AC8.15	0.86		116	154	GDB
16	D16S533	NA	0.78		199	215	GDB
16	D16S539	GATA11C06	0.60		148	172	GDB
16	D16S540	GATA7B02	0.68		232	248	GDB
16	D16S541	GATA7E02	0.77		144	164	GDB
16	D16S543	wg1f2	0.83		280	502	ARMOUR, J. ET AL. (1994) HMG 3, 599-605.
16	D16S663	CW2	0.83		113	129	PERAL, B. ET AL (1994) AM. J. HUM. GENET. 54, 899-908.
16	D16S665	SM6	0.69		86	140	PERAL, B. ET AL (1994) AM. J. HUM. GENET. 54, 899-908.
16	D16S668	MFD180	0.70		117	127	AM J HUM GEN 44:388-396, 1989
16	D16S747	ATA2D09	0.42		270	270	GDB
16	D16S748	ATA3A07	0.80		187	214	GDB
16	D16S749	GAAT1E9	0.69		138	150	GDB
16	D16S750	GAAT2B10	0.47		109	117	GDB
16	D16S751	GATA49B09	1.00		223	223	GDB
16	D16S752	GATA51G03	0.92		101	276	GDB
16	D16S753	GGAA3G05	0.88		252	276	GDB
16	HBAP1	NA	0.76	0.73	95	107	FOUGEROUSSE,F. ET AL.(1992) NAR 20,1165.
16	SPN	NA	0.96		145	185	ROGEAV,E.&KERYANOV,S.(1992) HMG 1,657.
17	CACNLB1	PCR2	0.82	0.80	81	121	GDB
17	CHRNB1	c15F4	0.88		165	165	GUZZETTA,V. ET AL.(1992) GENOMICS13,551-559.
17	D17S107	VAW134	0.72		234	246	BARKER, D. ET AL. (1993) HMG 2, 1086.
17	D17S113	NA	0.64		146	160	BARKER,D. ET AL.(1992) NAR 20,923.
17	D17S122	pRM11-GT	0.74		157	167	HARRIS,P. ET AL.,(1991) LANCET 338,1484-86.
17	D17S1288	ATA1H07	0.93		160	216	GDB
17	D17S1289	GATA41E09	0.58		148	148	GDB
17	D17S1290	GATA49C09	0.92		199	199	GDB
17	D17S1291	GCT1E1	0.31		147	147	GDB
17	D17S1292	GCT8D06	0.62		96	96	GDB
17	D17S1293	GGAA7D11	0.88		262	290	GDB
17	D17S1294	GGAA9D03	0.70		248	272	GDB
17	D17S250	MFD 15	0.91	0.81	151	169	NAR 18(8):2200, 1990
17	D17S250II	MFD 46	0.94	0.82	144	160	GENOMICS 8:400- , 1990
17	D17S261	MFD 41	0.43	0.44	157	171	NAR 18(8):2200, 1990
17	D17S379	NA	0.74		342	362	CARROZZO,R.& LEDBETTER,D.(1993) HMG2,615.
17	D17S513	NA	0.89		183	203	OLIPHANT,A. ET AL.(1991) NAR 19,4794.
17	D17S518	NA	0.76	0.67	88	100	COUCH,F. ET AL.(1991) NAR 19,5093.
17	D17S520	MFD144	0.77	0.83	130	144	IN PREPARATION
17	D17S525	cCI17-453	0.72		101	107	STACK, M. ET AL. (1994) HMG 3, 1443.
17	D17S559	cCI17-713	0.70		110	135	STACK, M. ET AL. (1994) HMG 3, 1443.
17	D17S578	MFD152	0.63	0.65	148	174	GENOMICS 8:400- , 1990
17	D17S579	MFD188	0.87	0.79	111	133	J. WEBER, ET AL, GENOMICS, IN PRESS
17	D17S581	MIT-MS52	0.75		155	155	HUMAN GENET 87:401, 1991
17	D17S582	MIT-MS105	0.63		123	123	HUMAN GENET 87:401, 1991
17	D17S583	MIT-N127	0.54		178	178	HUMAN GENET 87:401, 1991
17	D17S584	MIT-MS246	0.52		139	139	HUMAN GENET 87:401, 1991
17	D17S588	42D6	0.85		154	174	GDB
17	D17S740	2KZ14-B4	0.77		103	151	O'CONNELL,P. ET AL.(1993) GENOMICS 15,38-47.
17	D17S776	MFD191	0.40		111	129	GENOMICS 8:400- , 1990
17	D17S783	AFM026vh7	0.71		241	255	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
17	D17S784	AFM044xg3	0.79		226	238	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
17	D17S785	AFM049xc1	0.84		181	207	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
17	D17S786	AFM051xd10	0.77		135	157	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
17	D17S787	AFM095tc5	0.82		138	166	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
17	D17S788	AFM095zd11	0.70		186	198	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801

CH	LOCUS	ASSAY	HET	PIC	SIZE MIN	ANGE MAX	REFERENCE
17	D17S789	AFM107yb8	0.83		154	170	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
17	D17S790	AFM151xa11	0.79		187	201	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
17	D17S791	AFM155xd12	0.88		165	199	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
17	D17S792	AFM158xc3	0.60		190	200	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
17	D17S793	AFM165zd4	0.70		95	109	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
17	D17S794	AFM168xd12	0.00		226	236	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
17	D17S795	AFM175xg3	0.82		105	121	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
17	D17S796	AFM177xh6	0.82		144	174	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
17	D17S797	AFM179xa1	0.59		198	204	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
17	D17S798	AFM179xg11	0.00		209	229	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
17	D17S799	AFM192yh2	0.69		186	200	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
17	D17S800	AFM200zf4	0.74		168	178	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
17	D17S801	AFM203xg5	0.86		258	336	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
17	D17S802	AFM210xa5	0.83		166	188	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
17	D17S804	AFM225zc1	0.62		156	170	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
17	D17S805	AFM234ta1	0.59		216	228	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
17	D17S806	AFM234td2	0.91		153	185	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
17	D17S807	AFM234xc9	0.86		114	138	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
17	D17S808	AFM238yf8	0.68		147	161	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
17	D17S809	AFM248tb9	0.72		229	247	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
17	D17S810	AFM248yg1	0.51		236	242	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
17	D17S836	AFM163yg1	0.48		202	210	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S839	AFM200yb12	0.56		155	175	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S840	AFM207vf4	0.59		238	252	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S841	AFM238vb10	0.42		253	273	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S842	AFM240xe5	0.64		112	122	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S846	227	0.83		215	255	GDB
17	D17S849	AFM234wg3	0.68		251	261	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
17	D17S855	248yg9	0.82		145	145	ANDERSON, L. ET AL. (1993) GENOMICS 17, 618-623
17	D17S856	OF2	0.39		260	260	ANDERSON, L. ET AL. (1993) GENOMICS 17, 618-623
17	D17S857	OF1	0.81		106	122	GDB
17	D17S858	OF3	0.63		113	127	GDB
17	D17S859	OF4	0.78		133	133	ANDERSON, L. ET AL. (1993) GENOMICS 17, 618-623
17	D17S920	AFM186xa1	0.68		97	109	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S921	AFM191xh12	0.62		169	185	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S922	AFM197xh6	0.61		178	192	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S923	AFM200va9	0.76		285	291	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S924	AFM203wh4	0.66		124	132	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S925	AFM206yh8	0.60		151	165	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S926	AFM207xa11	0.62		243	260	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S928	AFM217yd10	0.30		135	165	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S929	AFM074zf7	0.61		217	229	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S930	AFM240yg7	0.74		104	110	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S931	AFM248tg5	0.53		218	238	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S932	AFM248yg9	0.66		185	201	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S933	AFM254vg5	0.52		188	206	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S934	AFM256vb9	0.48		170	190	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S935	AFM260yd5	0.80		150	156	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S936	AFM260yg5	0.78		93	103	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S937	AFM107ye3	0.73		125	149	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S938	AFM263wh5	0.90		164	182	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S939	AFM267xh1	0.54		191	215	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S940	AFM268yd5	0.67		207	215	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S941	AFM269xd1	0.74		269	277	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S942	AFM269yf1	0.77		168	176	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S943	AFM269zb1	0.65		181	199	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S944	AFM277vg9	0.69		212	224	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S945	AFM282yd1	0.62		186	208	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S946	AFM283zb9	0.79		128	142	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S947	AFM290vc9	0.87		250	282	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S948	AFM291ve9	0.76		125	149	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S949	AFM292vh9	0.67		207	221	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S950	AFM298wa5	0.67		174	198	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S951	AFM298wg5	0.58		170	188	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S952	AFM302wh9	0.71		129	141	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S953	AFM304xh5	0.69		119	131	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S954	AFM316vg1	0.85		218	281	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S955	AFM317yg1	0.68		181	189	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S956	AFM319wc1	0.47		162	174	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S957	AFM323wd9	0.67		128	134	WEISSENBACH J: NATURE GENETIC, JUNE 1994

CH	LOCUS	ASSAY	HET	PIC	SIZE ANGE MIN	MAX	REFERENCE
17	D17S958	AFM323yg9	0.75		101	115	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S959	AFM350zb1	0.81		127	137	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S960	AFMa120xd5	0.74		127	135	WEISSENBACH J: NATURE GENETIC, JUNE 1994
17	D17S968	GATA10F10	0.75		163	163	GDB
17	D17S969	GATA10H07	0.72		127	127	GDB
17	D17S974	GATA8C04	1.00		201	217	GDB
17	D17S975	GGAT2C07	0.70		256	264	GDB
17	HGH	NA	0.83		201	253	POLYMEROPOULOS,M. ET AL.(1993) HMG 2,689.
17	HOX2B	NA	0.74	0.70	130	140	DEINARD,A. ET AL.(1992)NAR 20,1171.
17	KRT9	K9CAI4F	0.78		182	198	GDB
17	MPO	NA	0.55	0.45	104	110	POLYMEROPOULOS,M. ET AL.(1991) NAR 19,1961.
17	NF1	NA	0.82		171	187	GDB
17	NF1-TET	NA	0.72		235	274	ANDERSON, L. ET AL. (1993) HMG 2, 1083.
17	SCN4A	PCR2	0.79		357	377	GDB
17	SSTR2	NA	0.88		132	148	GDB
17	THRA1	NA	0.81		158	176	FUTREAL,P.A.,ET AL(1992) HMG 1,66.
17	TP53	NA	0.90		103	135	JONES,M.& NAKAMURA,Y.(1992) GENES,CHROM.& CANCER 5,89-90.
18	D18S19	NA	0.73		104	128	BARE,J. ET AL. (1992) HMG 1,553.
18	D18S34	MFD 26	0.81	0.78	103	119	NAR 18(8):2201, 1990
18	D18S35	MFD 32	0.72	0.65	104	124	NAR 18:( ):6465, 1990
18	D18S36	MIT-MS156	0.75		146	146	HUMAN GENET 87:401, 1991
18	D18S363	MFD273	0.87	0.86	177	247	GENOMICS 8:400- , 1990
18	D18S364	MFD261	0.76	0.74	112	136	GENOMICS 8:400- , 1990
18	D18S37	NA	0.63	0.56	98	108	SZUBRYT,S. ET AL. (1993) HMG 2,90.
18	D18S379	UT574			263	323	GDB
18	D18S38	MFD153	0.53	0.52	162	172	GENOMICS 8:400- , 1990
18	D18S39	MFD193	0.65	0.70	204	222	GENOMICS 8:400- , 1990
18	D18S391	UT1302	0.75		182	182	GDB
18	D18S40	CU18-001	0.80	0.78	79	109	STRAUB,R. ET AL.(1993) GENOMICS 15,48-56.
18	D18S41	CU18-002	0.72	0.69	185	209	SZUBRYT,S. ET AL. (1993) HMG 2,90.
18	D18S42	NA	0.78		180	196	ROJAS,KATHERINE ET AL.(1992)GENOMICS 14,1095-97.
18	D18S43	NA	76.00		68	88	ROSEN,D.& BROWN,JR.,R.(1993) HMG 2,617.
18	D18S44	NA	0.79		95	111	ROSEN,D.& BROWN,JR.,R.(1993) HMG 2,617.
18	D18S45	CU18-006	0.62	0.57	181	203	SZUBRYT,S. ET AL. (1993) HMG 2,90.
18	D18S450	AFM191vc7	0.79		266	278	WEISSENBACH J: NATURE GENETIC, JUNE 1994
18	D18S451	AFM199zg7	0.53		89	99	WEISSENBACH J: NATURE GENETIC, JUNE 1994
18	D18S452	AFM206xf4	0.00		123	141	WEISSENBACH J: NATURE GENETIC, JUNE 1994
18	D18S452	AFM206xh4	0.00		123	141	WEISSENBACH J: NATURE GENETIC, JUNE 1994
18	D18S453	AFM214xd2	0.65		135	165	WEISSENBACH J: NATURE GENETIC, JUNE 1994
18	D18S454	AFM217ya9	0.69		142	148	WEISSENBACH J: NATURE GENETIC, JUNE 1994
18	D18S455	AFM070yd11	0.64		136	150	WEISSENBACH J: NATURE GENETIC, JUNE 1994
18	D18S456	AFM234zb6	0.72		233	243	WEISSENBACH J: NATURE GENETIC, JUNE 1994
18	D18S457	AFM238vb6	0.77		114	126	WEISSENBACH J: NATURE GENETIC, JUNE 1994
18	D18S458	AFM238xc11	0.78		208	218	WEISSENBACH J: NATURE GENETIC, JUNE 1994
18	D18S459	AFM238yd8	0.66		137	151	WEISSENBACH J: NATURE GENETIC, JUNE 1994
18	D18S46	CU18-007	0.80	0.78	129	153	SZUBRYT,S. ET AL. (1993) HMG 2,90.
18	D18S460	AFM240xa7	0.80		183	197	WEISSENBACH J: NATURE GENETIC, JUNE 1994
18	D18S461	AFM240yd2	0.56		160	169	WEISSENBACH J: NATURE GENETIC, JUNE 1994
18	D18S462	AFM079xb3	0.62		179	193	WEISSENBACH J: NATURE GENETIC, JUNE 1994
18	D18S463	AFM242zh6	0.63		172	184	WEISSENBACH J: NATURE GENETIC, JUNE 1994
18	D18S464	AFM259vh9	0.69		283	291	WEISSENBACH J: NATURE GENETIC, JUNE 1994
18	D18S465	AFM260yh1	0.84		233	251	WEISSENBACH J: NATURE GENETIC, JUNE 1994
18	D18S466	AFM094ye5	0.83		212	216	WEISSENBACH J: NATURE GENETIC, JUNE 1994
18	D18S467	AFM107xh6	0.72		168	180	WEISSENBACH J: NATURE GENETIC, JUNE 1994
18	D18S468	AFM116ye9	0.65		273	285	WEISSENBACH J: NATURE GENETIC, JUNE 1994
18	D18S469	AFM116yg11	0.61		234	244	WEISSENBACH J: NATURE GENETIC, JUNE 1994
18	D18S47	CU18-008	0.60	0.56	197	211	SZUBRYT,S. ET AL. (1993) HMG 2,90.
18	D18S470	AFM276za9	0.81		270	292	WEISSENBACH J: NATURE GENETIC, JUNE 1994
18	D18S471	AFM283xg9	0.80		251	264	WEISSENBACH J: NATURE GENETIC, JUNE 1994
18	D18S472	AFM284ve1	0.73		149	163	WEISSENBACH J: NATURE GENETIC, JUNE 1994
18	D18S473	AFM292wg1	0.75		231	243	WEISSENBACH J: NATURE GENETIC, JUNE 1994
18	D18S474	AFM295xh1	0.66		119	139	WEISSENBACH J: NATURE GENETIC, JUNE 1994
18	D18S475	AFM298wf9	0.77		158	170	WEISSENBACH J: NATURE GENETIC, JUNE 1994
18	D18S476	AFM299yf1	0.74		263	275	WEISSENBACH J: NATURE GENETIC, JUNE 1994
18	D18S477	AFM301xf5	0.60		244	254	WEISSENBACH J: NATURE GENETIC, JUNE 1994
18	D18S478	AFM311we5	0.83		240	252	WEISSENBACH J: NATURE GENETIC, JUNE 1994
18	D18S479	AFM312vc5	0.72		204	304	WEISSENBACH J: NATURE GENETIC, JUNE 1994
18	D18S480	AFM320yc9	0.62		120	142	WEISSENBACH J: NATURE GENETIC, JUNE 1994
18	D18S481	AFM321xc9	0.81		183	203	WEISSENBACH J: NATURE GENETIC, JUNE 1994
18	D18S482	AFM323ze9	0.70		295	311	WEISSENBACH J: NATURE GENETIC, JUNE 1994

CH	LOCUS	ASSAY	HET	PIC	SIZE MIN	ANGE MAX	REFERENCE
18	D18S483	AFM324wc9	0.63		197	225	WEISSENBACH J: NATURE GENETIC, JUNE 1994
18	D18S484	AFM326tc9	0.84		260	266	WEISSENBACH J: NATURE GENETIC, JUNE 1994
18	D18S485	AFM330yd9	0.78		176	190	WEISSENBACH J: NATURE GENETIC, JUNE 1994
18	D18S486	AFM333wd5	0.65		105	109	WEISSENBACH J: NATURE GENETIC, JUNE 1994
18	D18S487	AFM344ta9	0.51		115	127	WEISSENBACH J: NATURE GENETIC, JUNE 1994
18	D18S488	AFM344zf1	0.67		239	264	WEISSENBACH J: NATURE GENETIC, JUNE 1994
18	D18S49	MFD245	0.73	0.69	102	118	GENOMICS 8:400- , 1990
18	D18S498	CU18-014	0.75		171	191	GDB
18	D18S499	MIT-T38	0.71		150	178	GDB
18	D18S50	CU18-009	0.73	0.69	176	190	SZUBRYT,S. ET AL. (1993) HMG 2,90.
18	D18S51	UT574	0.86	0.85	267	319	SZUBRYT,S. ET AL. (1993) HMG 2,90.
18	D18S52	AFM020tf12	0.77		116	130	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
18	D18S53	AFM036ya1	0.80		159	179	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
18	D18S535	GATA13	0.92		150	150	GDB
18	D18S536	GATA8E05	0.72		146	170	GDB
18	D18S537	GATA2E06	0.74		190	190	GDB
18	D18S539	GATA3G05	0.63		252	252	GDB
18	D18S54	AFM080xd7	0.82		205	221	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
18	D18S541	GATA10A09	0.79		272	283	GDB
18	D18S542	GATA11A06	0.88		182	194	GDB
18	D18S548	GATA4H06	0.78		212	224	GDB
18	D18S55	AFM122xc1	0.78		134	152	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
18	D18S554	AFM296wd5	0.64		212	228	WEISSENBACH J: NATURE GENETIC, JUNE 1994
18	D18S556	PCR1	0.88		172	202	GDB
18	D18S56	AFM123ya1	0.75		197	209	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
18	D18S57	AFM147yg7	0.88		88	110	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
18	D18S58	AFM164xe3	0.74		144	160	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
18	D18S59	AFM178xc3	0.82		148	164	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
18	D18S60	AFM178xe3	0.38		156	172	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
18	D18S61	AFM193yf8	0.88		157	183	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
18	D18S62	AFM197xh12	0.67		187	195	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
18	D18S63	AFM205td6	0.80		255	279	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
18	D18S64	AFM212xg5	0.75		188	209	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
18	D18S65	AFM240vh6	0.73		168	178	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
18	D18S66	AFM240xc7	0.86		244	262	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
18	D18S67	AFM248te1	0.82		113	129	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
18	D18S68	AFM248yb9	0.80		270	290	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
18	D18S69	AFM248yf1	0.79		194	210	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
18	D18S70	AFM254vd5	0.84		111	126	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
18	D18S71	AFM254yd5	0.82		252	282	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
18	D18S72	AFM256vd5	0.54		193	203	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
18	D18S73	AFM266wa5	0.71		140	144	WEISSENBACH J: NATURE GENETIC, JUNE 1994
18	D18S74E	NA	0.82		119	137	POLYMEROPOULOS,M. ET AL.(1992) HMG 1,779.
18	D18S78	MFD 80	0.47	0.47	151	163	GENOMICS 8:400- , 1990
18	D18S843	ACT1A01	0.80		179	191	GDB
18	D18S844	ATA1H06	0.77		182	200	GDB
18	D18S845	ATA5B08	0.60		218	236	GDB
18	D18S846	GAAT1E07	0.38		162	166	GDB
18	D18S847	GATA25H01	0.87		212	236	GDB
18	D18S848	GATA27H10	0.63		87	115	GDB
18	D18S849	GATA30B03	0.80		269	269	GDB
18	D18S850	GATA53B01	0.67		202	202	GDB
18	D18S851	GATA6D09	0.55		256	276	GDB
18	D18S852	GCT5D07	0.58		117	117	GDB
18	N/A	GATA24	0.72		N/A	N/A	GDB
18	DCC	NA	0.82		106	160	RISINGER,J.& BOYD,J. (1992) HMG 1,657.
18	FECH	NA	0.68		225	261	GDB
18	MBP	NA	0.80		208	232	POLYMEROPOULOS,M. ET AL.(1992) HMG 1,658.
18	PACAP	NA	0.78		101	125	GDB
19	N/A	GAAT1C6	0.62		144	152	GDB
19	N/A	GAAA1B03	0.81		204	228	GDB
19	APOC2	MFD 5		0.70	129	155	AM J HUM GEN 44:388-396, 1989
19	ATP1A3	NA	0.57	0.54	164	176	GDB
19	BCL3	NA	0.47		127	139	ST GEORGE-HYSLOP,P.H. ET AL.(1992)NAR 20,927.
19	CEA	MFD113	0.75	0.71	95	111	AM J HUMAN GENET, 44:388-396, 1989
19	CEAII	NA		0.65	104	120	KEIRNAN,E.C.ET AL(1991)NAR,19,3160.
19	D19S112	PCR1	0.86		120	142	GDB
19	D19S112	PCR1	0.86		120	142	JANSEN, G. ET AL. (1992) GENOMICS 13, 509-517.
19	D19S113	MFD111	0.50		92	100	GENOMICS 8:400- , 1990
19	D19S114	MFD112	0.60		85	95	AM J HUMAN GENET, 44:388-396, 1989

CH	LOCUS	ASSAY	HET	PIC	SIZE RANGE MIN	SIZE RANGE MAX	REFERENCE
19	D19S177	MFD120	0.82	0.77	159	177	WEBER, J. PERSONAL COMMUNICATION
19	D19S178	MFD139	0.80	0.74	143	189	AM J HUMAN GENET, 44:388-396, 1989
19	D19S178B	NA	0.75		N/A	N/A	GDB
19	D19S179	MFD176	0.68		100	700	AM J HUMAN GENET, 44:388-396, 1989
19	D19S180	MFD195	0.72	0.71	148	172	AM J HUMAN GENET, 44:388-396, 1989
19	D19S191	NA	0.81		103	125	GENOMICS 8:400- , 1990
19	D19S197	MIT-G116	0.92		153	153	HUMAN GENET 87:401, 1991
19	D19S198	MIT-M198	0.87		115	115	HUMAN GENET 87:401, 1991
19	D19S199	MIT-MS231	0.83		150	150	HUMAN GENET 87:401, 1991
19	D19S200	MIT-MS265	0.88		138	138	HUMAN GENET 87:401, 1991
19	D19S204	ZHC32	0.58		147	172	GDB
19	D19S206	NA	0.93		118	146	PETRUKHIN,K.& GILLIAM,T.(1992) HMG1,454.
19	D19S207	GJ-VSSM2	0.77		135	157	JANSEN,G. ET AL.(1993) 2,333.
19	D19S208	AFM116xc7	0.73		167	175	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
19	D19S209	AFM116xe9	0.78		260	272	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
19	D19S210	AFM119xh6	0.75		165	177	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
19	D19S211	AFM126zc1	0.57		193	207	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
19	D19S212	AFM143xe9	0.68		195	213	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
19	D19S213	AFM158ye3	0.69		174	184	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
19	D19S214	AFM163xf8	0.64		163	183	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
19	D19S215	AFM164za7	0.34		245	263	WEISSENBACH J: NATURE GENETIC, JUNE 1994
19	D19S216	AFM164zb8	0.76		179	191	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
19	D19S217	AFM165xa9	0.77		219	233	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
19	D19S218	AFM207wc3	0.61		240	256	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
19	D19S219	AFM210yg9	0.63		160	190	WEISSENBACH J: NATURE GENETIC, JUNE 1994
19	D19S220	AFM214yf12	0.85		265	283	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
19	D19S221	AFM224ye9	0.87		191	211	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
19	D19S222	AFM234vb2	0.67		233	241	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
19	D19S223	AFM238wf10	0.82		228	246	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
19	D19S224	AFM240vc1	0.80		240	262	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
19	D19S225	AFM248zc1	0.73		170	180	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
19	D19S226	AFM256yc9	0.86		235	263	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
19	D19S228	AFM238xc7	0.64		201	209	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
19	D19S244	MFD236	0.85	0.83	96	144	AM J HUMAN GENET, 44:388-396, 1989
19	D19S245	MFD235	0.68	0.63	187	211	AM J HUMAN GENET, 1993, IN PRESS
19	D19S246	MFD232	0.84	0.82	185	233	AM J HUMAN GENET, 44:388-396, 1989
19	D19S247	MFD219	0.79	0.76	219	275	AM J HUMAN GENET, 44:388-396, 1989
19	D19S248	MFD218	0.60	0.53	131	151	AM J HUMAN GENET, 44:388-396, 1989
19	D19S249	MFD243	0.64	0.57	195	211	AM J HUMAN GENET, 44:388-396, 1989
19	D19S250	MFD242	0.66	0.61	267	291	AM J HUMAN GENET, 44:388-396, 1989
19	D19S251	MFD241	0.66	0.60	228	248	AM J HUMAN GENET, 44:388-396, 1989
19	D19S252	MFD240	0.70	0.66	107	128	AM J HUMAN GENET, 44:388-396, 1989
19	D19S253	MFD239	0.78	0.75	212	244	AM J HUMAN GENET, 44:388-396, 1989
19	D19S254	MFD238	0.78	0.75	110	150	AM J HUMAN GENET, 44:388-396, 1989
19	D19S255	MFD237	0.58	0.52	117	137	AM J HUMAN GENET, 44:388-396, 1989
19	D19S384	MFD283	0.56	0.51	99	107	GENOMICS 8:400- , 1990
19	D19S385	MFD285	0.63	0.60	138	162	GENOMICS 8:400- , 1990
19	D19S386	NA	0.90		129	158	IIAUKE, J. ET AL. (1993) HMG 2, 1981.
19	D19S391	UT486	0.67		124	178	GDB
19	D19S394	UT705	0.92		219	267	GDB
19	D19S404	AFM161xf6	0.75		96	116	WEISSENBACH J: NATURE GENETIC, JUNE 1994
19	D19S405	AFM205yf10	0.66		100	120	WEISSENBACH J: NATURE GENETIC, JUNE 1994
19	D19S406	AFM205zf2	0.81		207	229	WEISSENBACH J: NATURE GENETIC, JUNE 1994
19	D19S407	AFM214yf6	0.51		203	237	WEISSENBACH J: NATURE GENETIC, JUNE 1994
19	D19S408	AFM238vg11	0.53		122	146	WEISSENBACH J: NATURE GENETIC, JUNE 1994
19	D19S409	AFM269xg5	0.79		167	177	WEISSENBACH J: NATURE GENETIC, JUNE 1994
19	D19S410	AFM278xc5	0.43		156	174	WEISSENBACH J: NATURE GENETIC, JUNE 1994
19	D19S411	AFM283yc5	0.73		145	157	WEISSENBACH J: NATURE GENETIC, JUNE 1994
19	D19S412	AFM284yg5	0.81		89	113	WEISSENBACH J: NATURE GENETIC, JUNE 1994
19	D19S413	AFM292wd9	0.77		69	91	WEISSENBACH J: NATURE GENETIC, JUNE 1994
19	D19S414	AFM295xg9	0.82		163	187	WEISSENBACH J: NATURE GENETIC, JUNE 1994
19	D19S415	AFM302yb9	0.81		251	259	WEISSENBACH J: NATURE GENETIC, JUNE 1994
19	D19S416	AFM304xf5	0.75		165	185	WEISSENBACH J: NATURE GENETIC, JUNE 1994
19	D19S417	AFM304zg1	0.71		222	236	WEISSENBACH J: NATURE GENETIC, JUNE 1994
19	D19S418	AFM319vb5	0.82		81	93	WEISSENBACH J: NATURE GENETIC, JUNE 1994
19	D19S419	AFM326xb1	0.76		165	169	WEISSENBACH J: NATURE GENETIC, JUNE 1994
19	D19S420	AFM326xh9	0.55		251	267	WEISSENBACH J: NATURE GENETIC, JUNE 1994
19	D19S421	AFMa123xe5	0.80		188	220	WEISSENBACH J: NATURE GENETIC, JUNE 1994
19	D19S422	AFMa123xe9	0.86		183	199	WEISSENBACH J: NATURE GENETIC, JUNE 1994
19	D19S423	AFMa131wf5	0.48		175	196	WEISSENBACH J: NATURE GENETIC, JUNE 1994

CH	LOCUS	ASSAY	HET	PIC	SIZE ANGE MIN	MAX	REFERENCE
19	D19S424	AFMa132zb9	0.51		141	161	WEISSENBACH J: NATURE GENETIC, JUNE 1994
19	D19S425	AFMa139we9	0.71		252	280	WEISSENBACH J: NATURE GENETIC, JUNE 1994
19	D19S426	MFD318	0.70	0.65	99	113	GENOMICS 8:400- , 1990
19	D19S427	MFD319	0.75	0.72	113	129	GENOMICS 8:400- , 1990
19	D19S428	wg1g5	0.87		301	374	ARMOUR, J. ET AL. (1994) HMG 3, 599-605.
19	D19S431	GATA6D01	0.73		288	304	GDB
19	D19S432	GATA9B02	0.69		184	204	GDB
19	D19S433	GGAA2A03	0.86		199	221	GDB
19	D19S434	GGAT4B07	0.31		264	280	GDB
19	D19S460	SSLP-19	0.83		101	123	GDB
19	D19S47	MFD 9	0.72	0.69	88	106	AM J HUMAN GENET, 1993, IN PRESS
19	D19S48	MFD 10	0.39	0.42	130	144	AM J HUMAN GENET, 1993, IN PRESS
19	D19S49	MFD 11	0.78	0.71	106	122	NAR 18(7):1927, 1990
19	D19S535	SSLP-199	0.85		139	163	GDB
19	D19S536	SSLP-1910	0.76		121	133	GDB
19	D19S582	GATA50C01	0.60		135	135	GDB
19	D19S75	MFD 13	0.60	0.61	136	152	NAR 18(15):4639, 1990
19	D19S76	MFD 37	0.52	0.40	66	70	NAR 18(9):2835, 1990
19	DM	NA	0.73		72	144	GDB
19	EPO	NA	0.74		173	232	GDB
19	EPOR	NA	0.87		236	268	MCDONALD,M. ET AL.(1993) HMG 2,619.
19	HRC	PCR2	0.70		110	120	GDB
19	INSR	INSRE3	0.58		129	151	GDB
19	KLK	NA	0.66		176	202	RICHARDS,R. ET AL.(1991) GENOMICS 11,77-82.
19	LDLR	NA	0.49		106	112	ZULIANI,G & HOBBS,H.H.(1990) NAR18,4300.
19	LIPE	NA	0.73	0.77	150	162	LEVITT,R. ET AL.(1992) HMG 1,139.
19	RFX1	NA	0.72		186	194	GDB
20	N/A	GAAT1F1	0.69		141	157	GDB
20	ADA	ADA8PR		0.77	170	182	ROUSTAN,P. ET AL.(1992) HMG 1,778.
20	CST3	NA		0.64	124	130	GDB
20	CSTP1	NA	0.61	0.58	123	141	GDB
20	D20S100	AFM057xa3	0.77		194	218	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
20	D20S101	AFM058xa1	0.64		136	150	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
20	D20S102	AFM066xh3	0.65		169	177	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
20	D20S103	AFM077xd3	0.71		92	106	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
20	D20S104	AFM102xa7	0.79		144	192	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
20	D20S105	AFM123xh12	0.68		232	246	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
20	D20S106	AFM123yf8	0.71		316	332	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
20	D20S107	AFM142xh4	0.80		262	278	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
20	D20S108	AFM163yh8	0.78		182	208	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
20	D20S109	AFM165xh2	0.89		106	133	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
20	D20S110	AFM182xg9	0.82		139	151	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
20	D20S111	AFM190xg1	0.64		227	235	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
20	D20S112	AFM197xb12	0.82		199	221	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
20	D20S113	AFM205th8	0.50		178	188	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
20	D20S114	AFM210vb4	0.78		253	269	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
20	D20S115	AFM218yg3	0.67		232	238	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
20	D20S116	AFM248td1	0.80		107	117	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
20	D20S117	AFM248yc5	0.85		150	182	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
20	D20S118	AFM260xg5	0.82		169	179	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
20	D20S119	AFM273yh9	0.82		104	118	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
20	D20S120	AFM276xh1	0.87		213	241	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
20	D20S169	AFM189ya7	0.45		193	201	WEISSENBACH J: NATURE GENETIC, JUNE 1994
20	D20S17	NA	0.71		130	140	IWASAKI,N.ET AL(1991)NAR 19,6970.
20	D20S170	AFM193xc9	0.59		230	254	WEISSENBACH J: NATURE GENETIC, JUNE 1994
20	D20S171	AFM046xf6	0.63		123	149	WEISSENBACH J: NATURE GENETIC, JUNE 1994
20	D20S172	AFM211yb8	0.56		162	176	WEISSENBACH J: NATURE GENETIC, JUNE 1994
20	D20S173	AFM218xe7	0.58		137	193	WEISSENBACH J: NATURE GENETIC, JUNE 1994
20	D20S174	AFM224yc1	0.17		277	305	WEISSENBACH J: NATURE GENETIC, JUNE 1994
20	D20S175	AFM234zb12	0.71		166	174	WEISSENBACH J: NATURE GENETIC, JUNE 1994
20	D20S176	AFM238vh6	0.80		167	187	WEISSENBACH J: NATURE GENETIC, JUNE 1994
20	D20S177	AFM238zc11	0.77		94	102	WEISSENBACH J: NATURE GENETIC, JUNE 1994
20	D20S178	AFM240vd6	0.75		244	256	WEISSENBACH J: NATURE GENETIC, JUNE 1994
20	D20S179	AFM240wb8	0.72		136	146	WEISSENBACH J: NATURE GENETIC, JUNE 1994
20	D20S180	AFM240wg1	0.53		151	157	WEISSENBACH J: NATURE GENETIC, JUNE 1994
20	D20S181	AFM240zf4	0.85		156	166	WEISSENBACH J: NATURE GENETIC, JUNE 1994
20	D20S182	AFM242yf8	0.74		197	211	WEISSENBACH J: NATURE GENETIC, JUNE 1994
20	D20S183	AFM248wb9	0.55		253	263	WEISSENBACH J: NATURE GENETIC, JUNE 1994
20	D20S184	AFM248xb9	0.49		246	252	WEISSENBACH J: NATURE GENETIC, JUNE 1994
20	D20S185	AFM262vf9	0.65		196	214	WEISSENBACH J: NATURE GENETIC, JUNE 1994

CH	LOCUS	ASSAY	HET	PIC	SIZE ANGE MIN	MAX	REFERENCE
20	D20S186	AFM120xc7	0.47		113	135	WEISSENBACH J: NATURE GENETIC, JUNE 1994
20	D20S187	AFM262xf9	0.57		155	169	WEISSENBACH J: NATURE GENETIC, JUNE 1994
20	D20S188	AFM288zf5	0.83		141	147	WEISSENBACH J: NATURE GENETIC, JUNE 1994
20	D20S189	AFM292xb5	0.73		285	309	WEISSENBACH J: NATURE GENETIC, JUNE 1994
20	D20S190	AFM294vb5	0.39		235	243	WEISSENBACH J: NATURE GENETIC, JUNE 1994
20	D20S191	AFM294zc5	0.71		201	211	WEISSENBACH J: NATURE GENETIC, JUNE 1994
20	D20S192	AFM299xd1	0.76		287	299	WEISSENBACH J: NATURE GENETIC, JUNE 1994
20	D20S193	AFM308we1	0.79		136	168	WEISSENBACH J: NATURE GENETIC, JUNE 1994
20	D20S194	AFM317te5	0.58		178	228	WEISSENBACH J: NATURE GENETIC, JUNE 1994
20	D20S195	AFM321xc1	0.76		237	261	WEISSENBACH J: NATURE GENETIC, JUNE 1994
20	D20S196	AFM321xd1	0.85		145	176	WEISSENBACH J: NATURE GENETIC, JUNE 1994
20	D20S197	AFM326xd5	0.79		187	205	WEISSENBACH J: NATURE GENETIC, JUNE 1994
20	D20S198	AFM333xe5	0.60		225	243	WEISSENBACH J: NATURE GENETIC, JUNE 1994
20	D20S199	AFMa131wf1	0.67		108	126	WEISSENBACH J: NATURE GENETIC, JUNE 1994
20	D20S200	AFMa141xd9	0.43		253	269	WEISSENBACH J: NATURE GENETIC, JUNE 1994
20	D20S206	GATA11A04	0.63		125	125	GDB
20	D20S211	AFM080ya1	0.69		134	158	WEISSENBACH J: NATURE GENETIC, JUNE 1994
20	D20S212	wg1a3	0.83		414	462	ARMOUR, J. ET AL. (1994) HMG 3, 599-605.
20	D20S25	IP20M61	0.84	0.82	158	190	HAZAN,J. ET AL.(1992) GENOMICS 12,183-189.
20	D20S27	MFD 25	0.65	0.64	128	138	NAR 18(8):2202, 1990
20	D20S41	IP20M57		0.83	128	128	HECHT,J.T. ET AL.(1991)NAR 19,6666.
20	D20S42	62	0.84		129	149	GDB
20	D20S45	IP20M83		0.72	129	129	HECHT,J.T. ET AL.(1991)NAR 19,6666.
20	D20S468	ACT1A04	0.81		194	212	GDB
20	D20S469	GGAA11E12	0.53		210	210	GDB
20	D20S470	GGAA7E02	0.94		266	314	GDB
20	D20S471	GGAA9H11	0.87		290	330	GDB
20	D20S48	IP20M12	0.83	0.81	200	216	HECHT,J.T. ET AL.(1991)NAR 19,6666.
20	D20S50	IP20M29	0.74		116	140	GDB
20	D20S52	IP20M48	0.82	0.79	210	226	HECHT,J.T. ET AL.(1991)NAR 19,6666.
20	D20S54	85	0.75		188	198	GDB
20	D20S55	IP20M3	0.79	0.76	108	124	HAZAN,J. ET AL.(1992) GENOMICS 12,183-189.
20	D20S56	NA	0.69		97	107	GDB
20	D20S57	IP207	0.69	0.64	119	135	HECHT,J.T. ET AL.(1991)NAR 19,6666.
20	D20S58	IP20M14	0.64	0.57	146	152	HAZAN,J. ET AL.(1992) GENOMICS 12,183-189.
20	D20S59	IP2017	0.76	0.73	245	261	HECHT,J.T. ET AL.(1991)NAR 19,6666.
20	D20S60	IP20M1	0.68	0.64	128	142	HECHT,J.T. ET AL.(1991)NAR 19,6666.
20	D20S61	23a/m	0.73		84	96	GDB
20	D20S64	IP20M5	0.88	0.87	216	288	HAZAN,J. ET AL.(1992) GENOMICS 12,183-189.
20	D20S65	MFD134	0.47	0.45	102	114	JONES,M. ET AL.(1992) HMG 1,131-33.
20	D20S66	MFD136	0.78	0.75	129	141	JONES,M. ET AL.(1992) HMG 1,131-33.
20	D20S75	MIT-M127	1.00		173	173	HUMAN GENET 87:401, 1991
20	D20S95	AFM023ta1	0.84		82	100	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
20	D20S96	AFM029yc3	0.82		99	133	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
20	D20S97	AFM036ya3	0.64		262	288	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
20	D20S98	AFM044xb4	0.81		259	275	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
20	D20S99	AFM051xf12	0.50		150	168	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
20	GNAS1	NA	0.58		217	220	GRANQVIST,M.ET AL(1991)NAR 19,4569.
20	NEC2	PC2S2	0.76		178	200	GDB
20	PCK1	NA	0.88		209	225	STOFFEL,M. ET AL.(1993) HMG 2,1-4.
20	PLC1	NA		0.72	156	182	ROTHSCHILD,CB.ET AL(1992)GENOMICS13,560-64.
20	PPGB	PPGBPR		0.65	191	209	ROTHSCHILD,C.ET AL.(1993)AM.J.HUM.GENET.52,110-123.
20	RPN2	NA	0.93		160	184	STOFFEL,M. ET AL.(1992) HMG 1,656.
20	SRC	NA	0.71		193	207	XIANG,K.ET AL(1991)NAR 19,6967.
21	APP	NA	0.81		167	205	TUPLER,R. ET AL.(1993)HMG 2,620.
21	D21S11	NA	0.90	0.82	204	240	SHARMA,V.& LITT,M.(1992) HMG 1,67.
21	D21S120	NA	0.75	0.71	330	330	BURMELSTER,M.ET AL(1990)NAR18,4969.
21	D21S1235	ABM-C75	0.80		104	144	GDB
21	D21S1252	AFM261zg1	0.75		231	251	WEISSENBACH J: NATURE GENETIC, JUNE 1994
21	D21S1253	AFM270xc1	0.58		174	190	WEISSENBACH J: NATURE GENETIC, JUNE 1994
21	D21S1254	AFM276za5	0.79		262	270	WEISSENBACH J: NATURE GENETIC, JUNE 1994
21	D21S1255	AFM283xh9	0.76		112	126	WEISSENBACH J: NATURE GENETIC, JUNE 1994
21	D21S1256	AFM284xe9	0.61		182	190	WEISSENBACH J: NATURE GENETIC, JUNE 1994
21	D21S1257	AFM294we1	0.76		151	175	WEISSENBACH J: NATURE GENETIC, JUNE 1994
21	D21S1258	AFM295yg5	0.65		139	157	WEISSENBACH J: NATURE GENETIC, JUNE 1994
21	D21S1259	AFM326tf1	0.55		208	228	WEISSENBACH J: NATURE GENETIC, JUNE 1994
21	D21S1260	AFMa152wd1	0.63		200	214	WEISSENBACH J: NATURE GENETIC, JUNE 1994
21	D21S1265	MFD338	0.69	0.67	101	133	GENOMICS 8:400- , 1990
21	D21S1268	M815H10	0.74		178	198	GDB
21	D21S1270	GATA8G04	0.86		174	198	GDB

CH	LOCUS	ASSAY	HET	PIC	SIZE ANGE MIN	MAX	REFERENCE
21	D21S13E	NA		0.69	111	115	GUO,Z. ET AL.(1990) NAR 18,4770.
21	D21S1431	ACT2E10	0.56		168	177	GDB
21	D21S1432	GATA11C12	0.64		127	155	GDB
21	D21S1433	GATA31C01	0.83		247	247	GDB
21	D21S1434	GATA43F04	0.67		187	187	GDB
21	D21S1435	GATA49E01	0.67		172	172	GDB
21	D21S1436	GGAA2E02	0.73		160	196	GDB
21	D21S1437	GGAA3C07	0.93		119	143	GDB
21	D21S145	P1.44	0.71		168	180	GDB
21	D21S156	MFD 55	0.92	0.79	77	107	GENOMICS 8:400- , 1990
21	D21S167	NA		0.81	156	182	GUO,Z. ET AL.(1990) NAR 18,4967.
21	D21S168	NA		0.73	104	118	GUO,Z. ET AL.(1990) NAR 18,5924.
21	D21S171	MFD 95	0.66	0.69	111	133	HUMAN GENET 87:401, 1991
21	D21S172	NA		0.58	145	161	SHARMA,V. ET AL.(1992) HMG 1,289.
21	D21S198	NA		0.81	112	128	SHARMA,V. ET AL.(1991) NAR 19,4023.
21	D21S210	NA	0.86		140	190	ANTONARAKIS,S. GENOMICS IN PRESS
21	D21S211	MFD163	0.53	0.48	93	103	GENOMICS 8:400- , 1990
21	D21S212	NA	0.84		414	462	GENOMICS 8:400- , 1990
21	D21S213	JHU21-GT05	0.74		152	164	GDB
21	D21S214	NA	0.82		240	256	GENOMICS 8:400- , 1990
21	D21S215	JHU21-GT14	0.68		168	180	GDB
21	D21S217	21-GT11	0.72		276	286	WARREN,A.ET AL.(1992)GENOMICS 14,818-19.
21	D21S219	NA	0.76		167	181	GOTO,J. ET AL.(1992) HMG 1,782.
21	D21S222	MIT-G121	0.88		131	131	HUMAN GENET 87:401, 1991
21	D21S223	NA	0.80		77	91	ROSEN,D. ET AL. (1992) HMG 1,547.
21	D21S224	NA	0.74		119	137	ROTHSCHILD,C.ET AL.(1993)AM.J.HUM.GENET.52,110-123.
21	D21S228	JHU21-GT25a	0.58		168	174	GDB
21	D21S232	21-GT09	0.68		118	124	WEBER,C. ET AL.(1993) HMG 2,612.
21	D21S235	NA	0.72		134	186	DONALDSON,D. ET AL.(1992) HMG 1,651.
21	D21S236	NA	0.71	0.65	104	128	SHARMA,V. ET AL.(1992) HMG 1,289.
21	D21S258	NA	0.87		184	206	WEHNERT,A. ET AL.(1992)HMG 1,449.
21	D21S259	AFM016xe5	0.80		117	131	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
21	D21S260	AFM147xb12	0.52		267	277	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
21	D21S261	AFM193xf10	0.51		296	304	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
21	D21S262	AFM198tc5	0.67		142	152	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
21	D21S263	AFM211zg9	0.75		175	201	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
21	D21S265	AFM234wa5	0.85		244	258	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
21	D21S266	AFM234xg9	0.60		153	173	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
21	D21S267	AFM238wc3	0.88		175	203	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
21	D21S268	AFM260ze9	0.88		226	250	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
21	D21S269	AFM263xf5	0.73		235	255	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
21	D21S270	AFM031xc5	0.86		199	223	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
21	D21S370	NA		0.73	207	221	GOTO,J. ET AL.(1993) HMG 2,616.
21	D21S416	ABM-C19	0.79		99	129	GDB
21	D21S49	NA	0.70		152	162	BESPALOVA,I. ET AL.(1993)HMG 2,613.
21	D21S65	NA	0.83		184	206	GOTO,J. ET AL.(1992) HMG 1,350.
21	HMG14	NA	0.69	0.67	69	93	POLYMEROPOULOS,M. ET AL.(1991) NAR 19,4306.
21	IFNAR	IOWA 21-07	0.83		462	480	MCINNIS,M. ET AL.(1991) GENOMICS 11,573-576.
21	PFKL	NA	0.70	0.66	129	145	POLYMEROPOULOS,M. ET AL.(1991) NAR 19,2517.
22	CRYB2	NA	0.60		200	212	MARINEAU,C.& ROULEAU,G.(1992) NAR20,1430.
22	CRYB2A	NA	0.75		172	193	BUETOW, K. ET AL. (1993) GENOMICS 18, 329-339.
22	CYP2D	NA	0.80	0.78	98	116	POLYMEROPOULOS,M. ET AL.(1991) NAR 19,3753.
22	CYP2D(q13)	NA	0.70		108	130	TROFATTER,J.A. ET AL. (1991) NAR19,2802.
22	D22S156	MFD 33	0.78	0.63	96	110	NAR 18(7):1927, 1990
22	D22S257	MFD 51	0.67	0.46	125	133	J. WEBER, CEPH, V. 5
22	D22S258	MFD162	0.82	0.78	183	195	J. WEBER, ET AL, GENOMICS, IN PRESS
22	D22S264	NA	0.80		190	210	MARINEAU,C.ET AL.(1992)NAR 20,1430.
22	D22S268	COS75	0.60		244	252	GDB
22	D22S270	MFD204	0.78	0.74	128	148	GENOMICS 8:400- , 1990
22	D22S272	AFM024xc9	0.70		132	150	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
22	D22S273	AFM106xd2	0.73		194	206	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
22	D22S274	AFM164th8	0.78		202	214	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
22	D22S275	AFM164ze3	0.82		160	174	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
22	D22S276	AFM165za5	0.74		241	263	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
22	D22S277	AFM168xa1	0.85		140	170	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
22	D22S278	AFM182xd12	0.77		231	245	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
22	D22S279	AFM205yc11	0.75		249	258	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
22	D22S280	AFM225xf6	0.83		208	220	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
22	D22S281	AFM238wc11	0.83		135	151	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
22	D22S282	AFM261ye5	0.84		144	164	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801

CH	LOCUS	ASSAY	HET	PIC	SIZE MIN	ANGE MAX	REFERENCE
22	D22S283	AFM262vh5	0.89		126	152	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
22	D22S284	AFM078wc5	0.77		86	102	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
22	D22S298	NA	0.74		123	123	RAM,K. ET AL.(1992) NAR 20,1428.
22	D22S299	NA	0.79		192	192	RAM,K. ET AL.(1992) NAR 20,1428.
22	D22S300	42.13	0.87		214	232	GDB
22	D22S300	NA	0.87		214	232	BUETOW, K. ET AL. (1993) GENOMICS 18, 329-339.
22	D22S301	NA	0.71		205	205	RAM,K. ET AL.(1992) NAR 20,1428.
22	D22S302	NA	0.75		218	218	RAM,K. ET AL.(1992) NAR 20,1428.
22	D22S303	NA	0.68		220	220	PORTER,J. ET AL.(1993) GENOMICS 15,57-61.
22	D22S304	NA	0.65		133	133	RAM,K. ET AL.(1992) NAR 20,1428.
22	D22S306	NA	0.61		105	105	RAM,K. ET AL.(1992) NAR 20,1428.
22	D22S307	NA	0.63		136	136	RAM,K. ET AL.(1992) NAR 20,1428.
22	D22S310	NA	0.90		174	196	BUETOW, K. ET AL. (1993) GENOMICS 18, 329-339.
22	D22S311	NA	0.81		262	262	RAM,K. ET AL.(1992) NAR 20,1428.
22	D22S315	AFM183xe9	0.80		177	203	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
22	D22S343	pN92	0.95		169	193	BUETOW, K. ET AL. (1993) GENOMICS 18, 329-339.
22	D22S345	MFD313	0.73	0.69	119	129	GENOMICS 8:400- , 1990
22	D22S351	22TG1	0.76		145	163	GDB
22	D22S418	AFM031yb10	0.72		137	161	WEISSENBACH J: NATURE GENETIC, JUNE 1994
22	D22S419	AFM211yf10	0.34		257	273	WEISSENBACH J: NATURE GENETIC, JUNE 1994
22	D22S420	AFM217xf4	0.73		148	164	WEISSENBACH J: NATURE GENETIC, JUNE 1994
22	D22S421	AFM234vh2	0.54		159	173	WEISSENBACH J: NATURE GENETIC, JUNE 1994
22	D22S422	AFM256vd1	0.56		120	140	WEISSENBACH J: NATURE GENETIC, JUNE 1994
22	D22S423	AFM261xd9	0.80		215	235	WEISSENBACH J: NATURE GENETIC, JUNE 1994
22	D22S424	AFM112yb4	0.58		156	168	WEISSENBACH J: NATURE GENETIC, JUNE 1994
22	D22S425	AFM265yf5	0.64		192	202	WEISSENBACH J: NATURE GENETIC, JUNE 1994
22	D22S426	AFM273vd9	0.88		215	225	WEISSENBACH J: NATURE GENETIC, JUNE 1994
22	D22S427	AFM288we5	0.74		96	110	WEISSENBACH J: NATURE GENETIC, JUNE 1994
22	D22S428	AFM321yb9	0.61		147	155	WEISSENBACH J: NATURE GENETIC, JUNE 1994
22	D22S429	AFM343ye9	0.82		253	259	WEISSENBACH J: NATURE GENETIC, JUNE 1994
22	D22S430	NA	0.68		84	98	GDB
22	D22S442	wg1d5	0.84		210	234	ARMOUR, J. ET AL. (1994) HMG 3, 599-605.
22	D22S444	GGAT3A11	0.56		124	132	GDB
22	D22S445	GGAT3C10	0.77		110	130	GDB
22	D22S446	AFM292va9	0.82		198	232	WEISSENBACH J: NATURE GENETIC, JUNE 1994
22	D22S448	164.1	0.86		186	208	GDB
22	D22S683	GATA11B12	0.94		160	196	GDB
22	D22S684	GATA4F03	1.00		230	230	GDB
22	D22S685	GATA6F05	0.90		172	196	GDB
22	D22S686	GGAA10F06	1.00		180	180	GDB
22	F8VWFP	NA	0.61		329	349	GDB
22	F8VWFPII	PCR2	0.57		130	150	GDB
22	IL2RB	Iowa22-02	0.91		149	163	BREWSTER,E. ET AL.(1992) NAR 19,4022.
22	TOPIP2	NA	0.91		113	155	TROFATTER,J.ET AL.(1992) HMG 1,455.
X		GATA3B02	1.00		N/A	N/A	GDB
X	D18S543	GATA2A12	0.77		230	250	GDB
X	DXS6786	ATA4A02	0.64		227	227	GDB
X	DXS6787	ATA4H10	0.48		261	261	GDB
X	DXS6788	ATA5G11	0.71		178	178	GDB
X	DXS6789	GATA31F01	0.92		148	148	GDB
X	DXS6790	GATA31H06	1.00		290	290	GDB
X	DXS6791	GATA42D03	0.77		166	166	GDB
X	DXS6792	GATA48D12	0.85		240	240	GDB
X	5DMD	NA		0.78	88	108	HUGNOT,J.P.ET AL(1991)NAR 19,3159.
X	ALAS2	NA	0.78		149	167	COX,T. ET AL. (1992) HMG 1,639-641.
X	AR	NA	0.89		195	195	SLEDDENS,H.ET AL(1992) NAR 20,1427.
X	ARA	NA	0.91		261	312	FEENER,ET AL.(1991) AM.J.HUM.GENET. 48,621-627.
X	CD40	HIGM1	0.70		197	231	CUTLER, R. ET AL. (1993) HMG 2, 828
X	CD40	NA	0.70		197	231	GDB
X	CYBB	NA	0.76		154	160	GDB
X	DMD-44	STR44	0.87		174	204	GDB
X	DMD-49	STR49	0.93		227	257	GDB
X	DMD-Y5	5'-5n3ca1	0.76		96	116	GDB
X	DXS1000	AFM248te9	0.35		230	236	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
X	DXS1001	AFM248we5	0.81		197	215	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
X	DXS1002	AFM249yh5	0.71		266	274	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
X	DXS1003	AFM276xf5	0.80		169	195	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
X	DXS101	NA	0.76		185	230	GDB
X	DXS102	CX38.1	0.71		145	163	GDB
X	DXS1036	AFM072zh3	0.82		145	151	WEISSENBACH J: NATURE GENETIC, JUNE 1994

CH	LOCUS	ASSAY	HET	PIC	SIZE MIN	ANGE MAX	REFERENCE
X	DXS1039	AFM119xd6	0.78		89	103	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1043	AFM126zd2	0.69		148	162	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1047	AFM150xf10	0.74		196	210	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1048	AFM151xg11	0.75		162	172	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1053	AFM164zd4	0.50		194	206	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1055	AFM168ya3	0.66		81	93	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1058	AFM200ye7	0.39		275	283	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1059	AFM203yd8	0.83		180	200	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS106	NA	0.81		87	103	FAIRWEATHER,N. ET AL. (1993) HMG 2,607-608.
X	DXS1061	AFM205yd2	0.87		224	242	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1062	AFM207xb8	0.79		222	248	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1065	AFM224zf2	0.76		160	164	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1066	AFM234tf8	0.82		257	269	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1067	AFM234vg7	0.81		214	230	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1068	AFM238yc11	0.82		245	259	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
X	DXS1072	AFM276wa5	0.53		271	285	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1105	AFM263wc1	0.41		208	226	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
X	DXS1106	AFM263we1	0.00		175	185	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
X	DXS1108	SDF	0.75		165	177	GDB
X	DXS1110	NA	0.68		252	268	GDB
X	DXS1111	NA	0.69		119	129	BROWNE,D. ET AL.(1993) HMG 2,611.
X	DXS1113	NA	0.75		154	178	WEBER,C. ET AL.(1993) HMG 2,612.
X	DXS1123	41ADF	0.68		168	178	GDB
X	DXS1126	EAD	0.68		230	252	GDB
X	DXS1191	AFM191za11	0.71		237	245	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1192	AFM196xa1	0.70		121	135	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1193	AFM199wc7	0.85		134	146	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1194	AFM203wa5	0.73		261	283	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1195	AFM207zd6	0.81		235	239	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1196	AFM056yb8	0.78		209	227	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1197	AFM072za5	0.50		240	248	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1199	AFM248wf9	0.75		277	291	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1200	AFM254wh1	0.46		275	281	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1201	AFM256ze5	0.69		267	287	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1202	AFM260ye5	0.60		265	285	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1203	AFM262vg1	0.50		210	220	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1204	AFM106xa3	0.44		237	249	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1205	AFM265va5	0.84		184	198	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1206	AFM269ya5	0.77		167	181	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1209	AFM273zd5	0.75		106	116	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1210	AFM274zb5	0.77		194	206	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1211	AFM276vf9	0.50		159	173	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1212	AFM280vf5	0.76		230	238	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1213	AFM282za9	0.73		230	244	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1214	AFM283wg9	0.81		210	220	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1215	AFM287ze5	0.61		246	250	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1216	AFM287zg1	0.85		244	248	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1217	AFM288ye9	0.55		231	243	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1218	AFM292wb9	0.42		261	275	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1219	AFM297yd1	0.60		230	246	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1220	AFM302xc9	0.58		192	218	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1221	AFM303wd1	0.73		149	161	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1222	AFM308xb9	0.72		234	240	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1223	AFM309yc1	0.75		158	170	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1224	AFM311vf5	0.65		157	167	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1225	AFM311vg5	0.66		194	218	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1226	AFM316yf5	0.72		201	223	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1227	AFM317ye9	0.45		174	186	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1229	AFM337wd5	0.74		202	230	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1230	AFM337zb1	0.62		189	199	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1231	AFM340ye1	0.80		202	208	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1232	AFMa123yb5	0.51		163	197	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1233	AFMa141xe5	0.80		191	197	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1235	DMD-50	0.70		233	251	CLEMENS,ET AL.(1991) AM.J.HUM.GENET. 49,951-960.
X	DXS1237	DMD-45	0.87		154	184	CLEMENS,ET AL.(1991) AM.J.HUM.GENET. 49,951-960.
X	DXS1254	MFD207	0.79	0.76	99	113	GENOMICS 8:400- , 1990
X	DXS1255	MFD279	0.27	0.26	95	115	GENOMICS 8:400- , 1990
X	DXS1275	AFM261zh5	0.75		206	220	WEISSENBACH J: NATURE GENETIC, JUNE 1994
X	DXS1283E	PRGS20	0.88		145	167	GDB
X	DXS1356	wg1e1	0.83		195	233	ARMOUR, J. ET AL. (1994) HMG 3, 599-605.

CH	LOCUS	ASSAY	HET	PIC	SIZE MIN	ANGE MAX	REFERENCE
X	DXS1367	ds-1	0.72		107	135	SCHINDELHAUER, D. ET AL. HMG 3, 1027
X	DXS16	16CA	0.85		88	114	CHANG, Y. ET AL. (1994) HMG 3, 1029.
X	DXS1683	A0563ME	0.67		142	174	GDB
X	DXS1684	NA	0.82		130	148	GONG. W. ET AL. (1994) HMG 3, 1442.
X	DXS178	NA	0.65		174	200	GDB
X	DXS207	PCR1	0.85		115	137	GDB
X	DXS227	NA	0.85		174	194	FOUGEROUSSE,F. ET AL.(1992) HMG 1,64.
X	DXS292	VK14	0.58		66	76	GDB
X	DXS294	NA	0.75	0.75	122	148	GEDEON,A. ET AL.(1991) NAR 19,5087.
X	DXS297	VK23	0.66		179	195	GDB
X	DXS3	NA	0.64		175	181	STANIER,P. ET AL.(1991) NAR 19,4793.
X	DXS337	RX9H6	0.73		139	145	GDB
X	DXS418	P122	0.83		140	158	GDB
X	DXS424	NA	0.83		126	142	HUANG,T. ET AL(1992) GENOMICS 13,375-380.
X	DXS43	43CA	0.86		86	130	CHANG, Y. ET AL. (1994) HMG 3, 1029.
X	DXS441	NA	0.76		173	189	RAM,K. ET AL.(1992) NAR 20,1428.
X	DXS443	RX324	0.60		204	210	GDB
X	DXS451	KQST80	0.80		182	204	GDB
X	DXS453	MFD 66	0.72		160	183	NAR 18(13):4037, 1990
X	DXS454	MFD 72	0.75		144	152	NAR 18(15):4635, 1990
X	DXS458	MFD 79	0.58		178	190	NAR 18(15):4635, 1990
X	DXS538	NA	0.72		154	184	BROWNE,D. ET AL.(1991) NAR 19,1161.
X	DXS548	RS46	0.65		190	206	RIGGINS, GJ. (1992) AM J MED GENET 44, 237-243.
X	DXS556	NA	0.73		176	192	THISELTON,D. ET AL.(1993)HMG 2,613.
X	DXS559	NA	0.63		230	248	ROUSTAN,P. ET AL.(1992) HMG 1,778.
X	DXS571	NA	0.46		130	148	CURTIS,A. ET AL.(1992) HMG 1,776.
X	DXS573	NA	0.72		137	145	ROUSTAN,P. ET AL.(1993) HMG 2,92.
X	DXS7	NA	0.00		157	167	MOORE,B. ET AL.(1992)NAR 20,929.
X	DXS730	MIT-MS21	0.63		192	192	HUMAN GENET 87:401, 1991
X	DXS731	MIT-MS266	0.74		100	100	HUMAN GENET 87:401, 1991
X	DXS737	MIT-MS120	0.60		167	167	HUMAN GENET 87:401, 1991
X	DXS738	MIT-E114	0.71		144	144	HUMAN GENET 87:401, 1991
X	DXS85	85CA	0.79		0	0	CHANG, Y. ET AL. (1994) HMG 3, 1029.
X	DXS981	STRX1	0.86		182	199	GDB
X	DXS983	AFM078za1	0.63		173	183	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
X	DXS984	AFM105xc5	0.72		154	184	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
X	DXS985	AFM112xf2	0.60		133	139	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
X	DXS986	AFM116xg1	0.76		149	173	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
X	DXS987	AFM120xa9	0.84		206	224	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
X	DXS988	AFM123xd4	0.62		134	144	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
X	DXS989	AFM135xe7	0.82		173	199	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
X	DXS990	AFM136yc7	0.76		172	180	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
X	DXS991	AFM151xf6	0.82		266	290	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
X	DXS992	AFM184xg5	0.87		201	211	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
X	DXS993	AFM203wf4	0.79		292	312	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
X	DXS994	AFM205wd2	0.00		212	220	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
X	DXS995	AFM207zg5	0.61		193	199	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
X	DXS996	AFM212xe5	0.82		153	171	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
X	DXS997	AFM217xa5	0.65		109	117	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
X	DXS998	AFM224zg11	0.57		113	119	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
X	DXS999	AFM234yf12	0.75		260	276	WEISSENBACH,J ET AL.(1992) NATURE 359:794-801
X	DYS-II	NA	0.82	0.77	214	228	FEENER,ET AL.(1991) AM.J.HUM.GENET. 48,621-627.
X	DYS-III	NA	0.54	0.58	219	225	FOUGEROUSSE,F. ET AL.(1992) NAR 20,1165.
X	F8C	NA	0.69		133	149	GDB
X	FMR-1	NA	N/A		N/A	N/A	PERGOLIZZI,R.G., ET AL. (1992)LANCET 339,2271-72.
X	HPRT	NA		0.70	151	163	HEARNE,C.& TODD,J. (1991) NAR 19,5450.
X	HPRTB	NA	0.75		263	299	EDWARDS,A. ET AL.(1992)GENOMICS 12,241-253.
X	KAL	NA	0.61	0.72	179	187	BOULOUX,P.ET AL(1991)NAR 19,5453.
X	MAOA	NA	0.75		285	388	HINDS,H. ET AL.(1992) GENOMICS 13,896-97.
X	MAOB	NA	0.64	0.73	285	388	GRIMSBY,J. ET AL.(1992)NAR 20,924.
X	PFC	NA	0.65		224	224	COLEMAN,M.ET AL.(1991) GENOMICS 11,991-996.
Y	DYS388	ATA10F11	N/A		127	127	GDB
Y	DYS389	GATA30F10	N/A		248	256	GDB
Y	DYS390	GATA31E10	N/A		205	221	GDB
Y	DYS391	GATA32C10	N/A		285	293	GDB

A genetic map of the mouse with 3,012 simple sequence length polymorphisms

William F. Dietrich, Joyce C. Miller, Robert G. Steen,
Mark Merchant, Deborah Damron, Robert Nahf, Diane C. Joyce,
Michael Wessel, Robert D. Dredge, Andre Marquis, Lincoln D. Stein,
Nathan Goodman, David C. Page, and Eric S. Lander

Whitehead Institute/MIT Center for Genome Research, Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge MA 02142

Abstract

We have constructed a genetic map of the mouse genome containing 3,012 SSLPs. The map provides an average spacing of 0.5 cM between markers, corresponding to about 1 Mb. Approximately 88% of the genome lies within 1.1 cM of a marker and 98% lies within 2.2 cM. The markers have an average polymorphism rate of 50% in crosses between laboratory strains. The markers are distributed in a relatively uniform fashion across the genome, although some deviations from randomness can be detected. In particular, there is a significant underrepresentation of markers on the X chromosome. This map represents the halfway point toward our goal of developing a mouse genetic map containing 6,000 SSLPs.

Introduction

Dense linkage maps are an invaluable tool for genetic and genomic analysis. They facilitate high resolution genetic mapping and positional cloning of monogenic traits, allow genetic dissection of polygenic traits, permit fine-structure linkage disequilibrium studies, assist in evolutionary comparisons, and provide an ordered scaffold on which complete physical maps of genomes can be assembled. The power of genetic maps increases with their density. For key organisms such as the human and mouse, extremely dense genetic maps are essential.

The discovery of simple sequence length polymorphisms (SSLPs) or microsatellites has greatly accelerated genetic map construction[1-4]. SSLPs occur at high frequency throughout mammalian genomes, tend to be highly polymorphic, are easily assayed by the polymerase chain reaction (PCR), and can be disseminated simply by publishing the sequence of their PCR primers. Several projects are underway to build dense SSLP maps of the human genome, notably by Weissenbach and colleagues[3,5].

For the past three years, the Whitehead Institute/MIT Center for Genome Research (CGR) has been developing an SSLP map of the mouse[4,6-9]. In 1992, we reported an initial map consisting of 317 markers[4]. In October 1993, we reported a 1518 marker map, integrated into the mouse gene map[8,9]. Here, we report the construction of a genetic map of the mouse genome containing 3,012 SSLPs. Markers are distributed at an average spacing of 0.5 cM, corresponding to about 1 Mb. The map represents the halfway point toward our goal of developing a mouse genetic map containing 6,000 SSLPs.

Construction of genetic map

The mouse genetic map was constructed essentially as previously described[4]. Briefly, random clones containing the simple sequence repeat $(CA)_n$ were identified by oligonucleotide hybridization from total mouse genomic libraries with size-selected inserts and their DNA sequence was determined by single-pass automated sequencing. In addition, mouse DNA sequences containing a variety of simple sequence repeats were identified in known gene sequences, primarily from GenBank. A PCR assay encompassing each simple sequence repeat was designed based on computer analysis. The PCR assays were tested for polymorphism on 12 inbred mouse strains. Those that defined different alleles in the OB and CAST strains–about 92% of the total–were genotyped in 46 progeny from an (OB x CAST) F2 intercross. Because the cross involves 92 meioses, there is a crossover every 1.1 cM on average and markers can be ordered to this resolution. (See Materials and Methods for further details.)

The genetic map is shown in Figure 1. The map contains 3012 SSLPs, of which 2798 were derived from anonymous clones and 214 were taken from known gene sequences reported in GenBank and elsewhere. The map densely covers all 20 mouse chromosomes and has a total genetic length of 1405 cM (Table 1). Because the markers were genotyped in an F2 intercross, the map represents sex-averaged genetic distance. The observed genetic distances agree reasonably well with previous consensus estimates based on data from various mouse crosses.

A full description of the markers–including primer sequences, complete locus sequence, allele sizes in characterized inbred strains, and genotypes in the cross-would require more than 200 journal pages and is thus omitted. Instead, this information can be obtained for any subset of markers by sending queries to an automatic electronic mail server maintained by CGR. To obtain a query form, send electronic mail consisting of the single word "help" to genome_database@genome.wi.mit.edu. Over the Internet, queries are typically answered in under two minutes.

Our SSLP map has recently been integrated with the mouse gene map[8,9]. Over 250 SSLPs were genotyped in a B6 x (B6 x SPRET) backcross in which Copeland, Jenkins and colleagues have genetically mapped RFLPs for about 1000 genes.

Distribution of genetic markers

A key issue in evaluating a map is the distribution of markers throughout the genome. Analysis of our initial 317 marker mouse map suggested that SSLPs were distributed in a relatively uniform fashion. With more than 3000 markers on the current map, it is possible to reinvestigate this question with greater precision.

There are a variety of ways to study whether the distribution of markers is uniform. One approach is to ask whether the observed number of markers on each chromosome agrees with expectation assuming that markers are uniformly distributed with respect to cytogenetic length[10]. For the autosomes, the chromosomal distribution of the random markers agrees remarkably well

with expectation (Table 2). There are no statistically significant deviations (after accounting for multiple hypothesis testing).

By contrast, the X chromosome shows a clear deficit of random markers (Table 2). In examining the proportion of markers from the X chromosome, a small correction is required inasmuch as the first 12% of the random markers were isolated from male DNA[4] while the remaining 88% were isolated from female DNA. After adjusting for this slight systematic underrepresentation, the X chromosome contains only 54% as many markers as expected under the assumption of a uniform distribution across cytogenetic length. Possible explanations for the striking deficit include: (1) a lower density of $(CA)_n$ repeat sequences on the X chromosome or (2) a lower rate of polymorphism among $(CA)_n$ repeats present on the X chromosome. (The trivial explanation that our "female" DNA source was actually from a male was excluded on two grounds. We confirmed the sex of the DNA sources by using PCR assays for the mouse *Zfx* and *Zfy* loci on the X and Y chromosomes, respectively[11]. We also noted that the proportion of markers on the X chromosome was two-fold higher among those markers isolated from the female DNA than the known male DNA.)

Another way to study the distribution of genetic markers is to examine the occurrence of clusters of crossovers and clusters of markers in the map. In our data, the position of every marker relative to every crossover can be identified. By ordering all crossovers (occurring in any of the meioses studied) and all markers relative to one another, the map of each chromosome can be reduced by a long string of the form "mmmmccmmmcc. . . ", where each m denotes the occurrence of a marker and each c denotes the

occurrence of a crossover (in one of the meioses studied). The string above, for example, indicates a succession of a block of four markers which showed no recombination in the meioses studied, an interval of two crossovers, a block of three markers which showed no recombination, and so on. Runs of many consecutive *c*'s correspond to large genetic intervals, while runs of many consecutive *m*'s correspond to large blocks of recombinationally unseparated markers.

If genetic markers are uniformly distributed with respect to crossovers, then the string should correspond to tossing a coin with probability π_m of being "m" and probability π_c (= 1-π_m) of being "c". Here, $\pi_m = M/(M+C)$ where M is the total number of markers and C is the total number of crossovers. The expected proportion of genetic intervals containing ≥ i consecutive crossovers is easily seen to be π_c^i (i.e., the probability that an m is followed by at least i consecutive c's). The distribution of the length of the longest genetic interval (i.e., the longest run of c's) can also be calculated[12] (see Methodology). Similarly, the expected proportion of blocks containing ≥ i recombinationally unseparated markers is π_m^i and the distribution of the longest such block can be calculated. To avoid bias due to the distribution of gene sequences, the analysis was performed using only the randomly generated markers.

With a single exception, the observed distribution of interval lengths fits expectation reasonably well (Table 3 and Figure 2). The longest run of consecutive crossovers has expected length 5.9, with a 95% confidence interval of roughly 5.0 – 8.5. The single outlier is the distalmost interval on Chromosome 19 between *D19Mit33* and *D19Mit6*, which had 11 crossovers or about 13 cM. The probability that such a large interval would occur by chance

anywhere in the map is only 0.003. (Genotypes were carefully reconfirmed to ensure that the crossovers did not result from mistyping.) The data suggest the possibility of a recombinational hotspot near the telomere of Chromosome 19, at least in (CAST x B6) F2 crosses. Interestingly, there does not appear to be enhanced recombination in a (SPRET x B6) x B6 backcross: Eicher and Shown[13] reported that the interval *D19Mit1–D19Mit33–D19Mit6* measured only 5 cM. This might be due to a difference between the strains or to enhanced recombination being present only in male meiosis (which contributes to recombination frequency in F2 intercrosses but not in *M. spretus* backcrosses in which only the female parent segregates for polymorphisms).

The distribution of the number of markers occurring between consecutive crossovers shows some modest evidence of clustering (Table 4 and Figure 3). There are significantly more occurrences of two consecutive crossovers without an intervening marker than expected by chance (485 vs. 425.3 ± 17.0; Z-score = 3.5). These data are consistent with the presence of recombinational hotspots in some regions. Adjusting for this excess, the remainder of this distribution is not a bad fit to expectation. The largest block of recombinationally unseparated markers would be expected to contain about 18 markers, with a 95% confidence interval of 15 – 25. In fact, the largest observed block is 24 which falls within the expected range.

The map appears to provide convenient entry points for nearly the entire genome. Approximately 88% of the map lies within 1.1 cM of a marker and 98% lies within 2.2 cM. The coverage is only slightly less than the expectation

for randomly spaced markers, which is 91% and 99%, respectively (based on expectations in Tables 3 and 4).

Polymorphism among mouse strains

SSLPs are particularly useful for mouse genetics because they are highly variable even among inbred laboratory mouse strains, making it possible to genotype virtually any cross of interest[4]. For the SSLPs in the map, allele sizes were determined in 12 inbred strains (10 laboratory strains, which are derived from *M. m. domesticus* and*M. m. musculus* progenitors; the different subspecies *M. m. castaneus*; and the different species *M. spretus*). The SSLPs in the map are necessarily variant between OB and CAST, since this was a prerequisite for genetic mapping in the cross.

The average number of alleles per SSLP was 6.0 ± 1.0 (Figure 4). Conveniently, over 75% of the pairwise allele difference are ≥ 4 basepairs – making it possible to score the difference on high percentage agarose gels. The polymorphism rate between laboratory strains and the different species *M. spretus* or the different subspecies *M. m. castaneus* was about 94%, while the polymorphism rate among laboratory strains averaged about 50% (Table 5). In only five of 45 pairwise comparisons among the 10 laboratory strains was the polymorphism rate below 44%: C3H-DBA (38%), C3H-BALB (37%), C3H-A (33%), BALB-A (32%), and OB–B6 (11%). (The last case is expected since OB and B6 are a congenic pair, witrh OB having been derived by repeated backcrossing to B6 with selection for the ob mutation.)

Interestingly, the polymorphism rate among laboratory strains was not constant across chromosomes (Table 6). The most extreme deviation was for the X chromosome, which showed a polymorphism rate of only 29% as compared to 50% for the autosomes. In addition, Chromosome 10 showed a significantly lower rate of polymorphism (35%). The rate of polymorphism between laboratory strains and *M. spretus* or *M. m. castaneus* did not vary significantly across chromosomes.

Discussion

The 3,012 marker genetic map of the mouse constructed here constitutes the densest SSLP map constructed in any organism to date. The total genetic length has not grown significantly with the addition of the last 1500 markers, suggesting that the map covers essentially the entire mouse genome. Interestingly, the genetic length of 1405 cM measured in our (CAST x OB) F2 intercross is significantly larger than the length of 1224 cM in a (SPRET x B6) backcross. (For this comparison, the genetic length of the SPRET x B6 backcross was recalculated using the Kosambi map function. The corresponding lengths are 1437 and 1344 cM with Haldane's map function[8].) The discrepancy is more striking than it may appear, since the F2 intercross reflects sex-averaged genetic distance while the backcross measures female genetic distance, which is generally thought to be substantially larger than in males. The difference may reflect crossover-suppression caused by local inversions between laboratory mouse and the evolutionary more distant *M. spretus*[8].

By a number of tests, the markers appear to be relatively uniformly distributed across the genome, although some modest evidence of clustering is present. There is only one suprisingly large gap, a 12 cM interval at the distal end of chromosome 19. Increased recombination in subtelomeric regions has been suggested for some human chromosomes[14]. More generally, there is an slight overall excess clustering of crossovers, which could reflect non-uniformity in the distribution of recombination or $(CA)_n$ repeats with respect to physical distance. Recombinational hotspots and coldspots are certainly known to exist in many organisms including the mouse[15], but the relative uniformity of marker distribution indicates that their effect is not dramatic on maps of the density and resolution reported here. Studies involving much denser maps may reveal greater clustering of recombination at a finer level, while studies involving many more meioses might reveal greater clustering of markers.

An unexpected observation was the nearly two-fold underrepresentation of markers on the X chromosome. The deficit could be due either to a deficit of $(CA)_n$ repeats on the X chromosome or a lower polymorphism rate among those $(CA)_n$ repeats on the X chromosome. In principle, these alternatives could be distinguished by determining the chromosomal distribution of the $(CA)_n$ repeats that were not polymorphic between OB and CAST.

It is also striking that the rate of polymorphism among laboratory strains was significantly lower on the X chromosome than for other chromosomes, although there was no difference for the rate of polymorphism between laboratory strains and the more distant CAST or SPR. In the human, the X chromosome has been reported to have a three-fold lower rate of RFLP

polymorphism[16]. The effect has been attributed to the different genetic and population genetic forces acting on the X chromosome as compared to the autosomes. For example, the mutation rate is thought to be higher in the male germline than the female germline. Since X chromosomes pass through only 2/3 as often as do autosomes, the mutation rate may be correspondingly lower. Also, the fact that X chromosomes function in the haploid state in males implies that selection acts differently, which may diminish polymorphism. Our data suggest that the presence of reduced polymorphism on the X chromosome may be general, at least in mammals.

The 3,012-marker SSLP map should facilitate a wide range of biological studies. For initial genetic mapping studies, one can select about 100 markers spaced at 15 cM intervals. It may be convenient to use polymorphisms that are easily resolved on agarose gels (about 75% of the total). To map a polygenic or quantitative trait[17], one would genotype each progeny for each marker, a task that might take a few months or less. To map a monogenic trait, one can proceed even more rapidly by using 'phenotyping pooling'[18] in which one initially genotypes only two samples—containing pooled DNA from affected progeny and unaffected progeny, respectively. The two samples should show similar proportions of the two parental alleles at markers unlinked to the trait, but quite different proportions for linked markers. In this manner, one can initially localize a trait with only about 200 PCR reactions, a task that can be accomplished in a few days. Once initial linkage is detected, individual progeny should be genotyped using all markers in the region to identify the closest flanking markers. Since a typical gene should lie at an average distance of 500 kb of a marker and since YAC lbraries with average insert size of 700 kb are available[19,20], chromosomal walking to the

gene should be rapid. In addition to its application in positional cloning, the map should be valuable for evolutionary studies as well[21].

For the purpose of constructing a physical map of the mouse genome with overlapping YACs, an even denser genetic map would be desirable. With a map consisting of 6,000 SSLPs, the average spacing between markers would be 500 kb and the typical gene would be at an average distance of 250 kb, both distances being smaller than the average size of current YACs. Given the presence of 50,000-100,000 $(CA)_n$ repeats in the mouse genome and the availability of streamlined methods for genetic map construction, such a goal should be feasible[19].

Methodology

Construction of genetic map. Briefly, (1) sequences containing simple sequence repeats (almost all $(CA)_n$) were obtained, either through sequencing of genomic clones that hybridize to $(CA)_{15}$ or $(GT)_{15}$, or by searching sequence databases; (2) PCR primers flanking the simple sequence repeat were selected; (3) the PCR assays were used to characterize allele sizes in twelve mouse strains: C57BL/6J-ob/ob, C57BL/6J, DBA/2J, A/J, C3H/HeJ, BALB/cJ, AKR/J, NON/Lt, NOD/MrkTacBr, LP/J (all laboratory strains, derived from *M. m. domesticus* and *M. m. musculus*), SPRET/Ei (a strain of the species *M. spretus*), and CAST/Ei (a strain of the subspecies *M. m. castaneus*); (4) for those assays detecting variation between OB and CAST, 46 progeny from an (OB x CAST)F2 intercross were genotyped; and (5) genetic maps were constructed by using the MAPMAKER computer package[22], incorporating a mathematical error-checking procedure[23]. These steps were performed essentially as previously described, with the following modifications. The OB and B6 strains are a congenic pair, with OB having been constructed by repeated backcrossing to B6 with selection for the *ob* mutation.

Genomic Libraries. The short-insert total genomic libraries were constructed by using a variety of different procedures: complete single digestion with *MboI, AluI, HaeIII,* and complete triple digests using *AluI, HaeIII,* and *RsaI.* Digests were fractionated on 4% NuSieve GTG agarose and fragments between 200 and 500 bp were selected for ligation into M13mp19. (Some libraries were also prepared by ligation into the plasmid pcDNAII, but this vector was eventually abandoned in favor of M13mp19 due to the superior sequence quality obtained from the single stranded template.) All libraries were transformed into XL1-Blue cells (Stratagene). DNA was prepared from

M13 clones by using a magnetic bead miniprep, essentially as previously described[24].

Length Screen of Clones: In some proportion of clones, the $(CA)_n$ repeat is too close or too far from the cloning site to allow PCR primers to be selected on both sides. To avoid sequencing such clones, a preliminary screening step was used to determine the size of the insert and the position of the repeat relative to the vector sequence. Miniprepped DNA was diluted 30-fold in distilled, de-ionized H_2O, and 1 µl of this dilution was used as the template in a 15 µl PCR reaction using AmpliTaq DNA polymerase (Perkin-Elmer Cetus) set up according to the manufacturer's specifications. Three PCR reactions were performed on each miniprep: (1) with primers flanking the M13 cloning site ("Forward": 5'-TGTAAAACGACGGCCAGT-3' and "Reverse": 5'-CAGGAAACAGCTATGACC-3'); (2) with the "Forward" primer and a primer complementary to a $(CA)_n$-repeat (5'-CCCGGATCC$(GT)_9$-3'); and (3) with the "Forward" primer and a primer complementary to a $(GT)_n$-repeat (5'-CCCGGATCC$(CA)_9$-3'). Reaction 1 is designed to measure the length of the insert, while Reaction 2 or 3 is designed to measure the distance from the Forward primer to the repeat. Reaction 1 and a pool of reactions 2 and 3 were electrophoresed on a 2% Metaphor agarose gel (FMC Bioproducts). Clones with insert size less than 700 bp and with the repeat within a range of 50 – 500 bp from the Forward primer were sequenced. Length screening was carried out in high throughput in 96-well microtiter plates.

Duplicate Checking. To avoid mapping previously encountered simple sequence repeats, a computer program was used to compare newly determined DNA sequences to previously sequenced clones. The proportion of duplicates remained in the range of 5-10% throughout the project, owing to

periodic substitution of new libraries constructed with different restriction enzymes.

Genotyping: To genotype F2 progeny for SSR polymorphisms, PCR reactions were performed with one radioactively labeled primer and one unlabelled primer and the products were visualized upon autoradiography of polyacrylamide gels. Primers were end-labeled with [γ-^{32}P]ATP (RediVue, Amersham) according to standard protocols[25]. A 20 ng aliquot of genomic DNA was amplified in a 10 µl PCR reaction using AmpliTaq DNA polymerase (Perkin-Elmer Cetus) according to manufacturer's specifications. The primer concentrations were 75 nM end-labeled forward primer, and 75 nM unlabeled reverse primer. The reactions were overlaid with 40 µl of light mineral oil (Sigma). Reactions were amplified on a TC1600 thermal cycler (Intelligent Automation Systems, Cambridge, MA) using the following protocol: 30 cycles of 92 °C for 30 seconds, 55 °C for 30 seconds, and 72 °C for 30 seconds. Gels and autoradiography were as previously described (Dietrich, *et al.* 1992).

Analysis of clusters of crossovers and markers. As noted in the text, the assumption that markers are randomly distributed with respect to genetic distance implies that the sequence of markers and crossovers occurring in the map should follow the expected behavior of coin flips. The expected behavior of head runs in coin flipping has been well studied[12]. If the probability of heads is p, the expected proportion of tails followed by at least i consecutive heads is p^i. If R_n denotes the longest run of consecutive heads when the coin is flipped n times, the expected value of R_n is $\mu = \log_{1/p}((n-1)(1-p)+1)$ and the

distribution of R_n is given approximately by $\text{Prob}(R - \mu > t) \approx \exp(-p^t)$. These formulas are used in computing the expectations in Tables 5 and 6.

Acknowledgments

We thank laboratory aides Loden Wangchuk, Dawa Tsering and Gail Farino for excellent technical assistance. We thank Robert Williams and Richelle Cutler of the University of Tennessee and Lois Maltais of the Jackson Laboratory for their help in ascertaining the official mouse gene names for SSLPs developed from gene sequences in GENBANK. This work was supported in part by grants from the National Center for Human Genome Research and the Markey Foundation (to E.S.L.).

References

1. Weber, J.L. & May, P.E. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* 44, 388-396 (1989).

2. Love, J.M., Knight, A.M., McAleer, M.A., & Todd, J.A. Towards construction of a high-resolution map of the mouse genome using PCR analysed microsatellites. *Nucleic Acids Res.* 18, 4123-4130 (1990).

3. Weissenbach, J., et al. A second-generation linkage map of the human genome. Nature 359, 794-801 (1992).

4. Dietrich, W.F. et al. A genetic map of the mouse suitable for typing intraspecific crosses. *Genetics* 131, 423-447 (1992).

5. Weissenbach, J. et al. (submitted).

6. Dietrich, W.F. *et al.* in *Genetic Maps 1992* (ed O'Brien, S.) 4.110-4.142 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1992).

7. Miller, J.C., *et al.* in Genetic Variants and Strains of the Laboratory Mouse, 3rd ed. (ed. Lyons, M.F. & Searle, A.) (Oxford Univ. Press, New York , 1994) (in press).

8. Copeland, N.G., *et al.* A genetic linkage map of the mouse: Current applications and future prospects. *Science* 262, 57-66 (1993).

9. Copeland, N.G., *et al.* Genome Maps IV: The Mouse. *Science* 262, 67-82 (1993).

10. Evans, E. in Genetic Variants and Strains of the Laboratory Mouse, 2nd ed., (ed. M.F. Lyon and A. Searle) 576-578 (Oxford University Press, New York, 1989).

11. Mardon, G., *et al.* Mouse Zfx protein is similar to Zfy-2: each contains an acidic activating domain and 13 zinc fingers. Mol Cell Biol 10, 681-8 (1990).

12. Arratia, R., Goldstein, L. & Gordon, L. Poisson Approximation and the Chen-Stein Method. *Stat. Sci.* 5, 403-434 (1990).

13. Eicher, E.M. & Shown, E.P. Molecular markers that define the distal ends of mouse autosomes 4, 13, 19 and the sex chromosomes. *Mammal. Genome.* 4, 226-229 (1993).

14. Tanzi, R.E., et al. A genetic linkage map of human chromosome 21: analysis of recombination as a function of sex and age. Am J Hum Genet 50, 551-8 (1992).

15. Bryda, E.C., De Pari, J.A., Sant'Angelo, D.B,, Murphy, D.B., & Passmore, H.C. Multiple sites of crossing over within the Eb recombinational hotspot in the mouse. Mamm Genome 2, 123-9 (1992).

16. Hofker, M.H., et al. The X chromosome shows less genetic variation at restriction sites than the autosomes. Am J Hum Genet 3, 438-51 (1986).

17. Lander, E.S. & Botstein, D. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121, 185-199 (1989).

18. Michelmore, R.W., Paran, I., & Kesseli, R.V. Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. Proc Natl Acad Sci USA 88, 9828-32 (1991).

19. Larin, Z., Monaco, A.P. & Lehrach, H. Yeast artificial chromosome libraries containing large inserts from mouse and human DNA. *Proc. Natl. Acad. Sci.*, 88, 4123 (1991).

20. Kusumi, K., Smith, J.S., Segre, J.A., Koos, D.S. & Lander, E.S. Construction of a large-insert yeast artificial chromosome (YAC) library of the mouse genome. *Mammalian Genome* 4, 391-392 (199 3).

21. Bruford, M.W. & Wayne, R.K. Microsatellites and their application to population genetic studies. *Curr. Opin. Genet. Devel.* 3, 939-944 (1993).

22. Lander, E.S., Green, P., Abrahamson, J., Barlow, A., Daly, M., Lincoln, S. and Newburg, L. MAPMAKER: An interactive computer package for constructing genetic linkage maps of experimental and natural populations. Genomics 1, 174-181 (1987).

23. Lincoln, S.E. & Lander, E.S. Systematic detection of errors in genomic linkage data. *Genomics* 14, 604-610 (1992).

24. Hawkins, T. M13 single-strand purification using a biotinylated probe and streptavidin coated magnetic beads. DNA Seq 3, 65-9 (1992).

25. Sambrook, J., Fritsch, E.F. and Maniatis, T. Molecular Cloning: A Laboratory Manual, Ed. 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1989).

Figure Legends

Fig. 1 SSLP genetic map of the mouse. For the 99% of markers that were developed at the Whitehead Institute/MIT Center for Genome Research, formal locus names have been abbreviated. For example, the locus D7Mit3 is simply denoted by 3 on chromosome 7. For loci developed elsewhere, the laboratory designation is retained (e.g., *D4Nds1* is denoted Nds1). For loci developed from genes for which an gene symbol has been assigned by the mouse nomenclature committee, the gene symbol is given in parentheses to aid in correlation with the mouse gene map. Linkage groups are represented by lines, with the centromere at the top. Each linkage group consists of markers that are linked to each other by a LOD score of at least 5. The type face of marker names indicates the statistical support for the genetic order shown. Markers whose order relative to the rest of the map is supported by a LOD of 2.5 or higher are indicated in bold face type; these are referred to as "framework" markers. Markers whose order is supported by a LOD between 1.0 and 2.5 are indicated in plain face type. Markers in italics have a LOD in support of order of 1.0 or less; these markers represent ambiguities with regard to the flanking markers only, but their placement LOD score relative to the rest of the map is greater than 2.5. Ambiguities can arise either because genotypes were not obtained for a few progeny or because the marker is dominant rather than codominant (about 5% of the total). Distances in centiMorgans between markers are indicated to the left of the line and were calculated using Kosambi's map function. Where multiple markers did not recombine in the 92 meioses studied, they are listed together in a block.

Fig. 2 Observed proportion of genetic intervals in the map having ≥ i crossovers compared to the expected proportion of π_c^i (where $\pi_c = 0.32$). Data are plotted on a logarithmic scale, for which the expected data fall on a line. Observed data are plotted as points. Solid line contains expected values, with dotted lines indicating upper and lower confidence intervals corresponding to 2 standard deviations.

Fig. 3 Observed proportion of blocks in the map containing ≥ i recombinationally unseparated markers compared to the expected proportion of π_m^i (where $\pi_m = 0.68$). Data are plotted on a logarithmic scale, for which the expected data fall on a line. Observed data are plotted as points. Solid line contains expected values, with dotted lines indicating upper and lower confidence intervals corresponding to 2 standard deviations.

Fig. 4 Histogram showing number of distinct allele sizes among 12 strains characterized for 3012 SSLPs reported here.

Table 1. Genetic Markers and Genetic Length by Chromosome

Chromosome	Number of Markers	Number of Random Markers	Number of Genes	'Consensus' Genetic Length[a]	Observed Genetic Length[b]
1	236	220	16	98	114.4
2	223	208	15	107	95.5
3	160	149	11	100	67.8
4	177	171	6	81	74.7
5	174	162	12	93	84.4
6	169	153	16	74	63.9
7	175	160	15	89	69.4
8	161	155	6	81	75.3
9	158	141	17	70	71.1
10	142	136	6	78	73.6
11	172	149	23	78	84.8
12	140	130	10	68	61.9
13	157	152	5	72	62.6
14	139	126	13	53	65.9
15	131	123	8	62	63.7
16	97	96	1	59	54.8
17	130	114	16	53	51.1
18	110	105	5	57	40.1
19	65	62	3	42	57.6
X	96	86	10	88	72.3
Total	3012	2798	214	1503	1404.9

a. Based on 'consensus' genetic map in Encyclopedia of the Mouse Genome (1993).
b. Distance between most proximal and most distal markers in the map reported here.

Table 2. Distribution of Random Markers based on Cytogenetic Length of Chromosomes

Chrom.	Number of Random Markers[b]	Based on cytogenetic length[a] % of total length	Expected Number of Markers[c]	Z-score[d]
Autosomes only				
1	220	7.7%	208.4 ± 13.9	0.84
2	208	7.4%	201.2 ± 13.6	0.50
3	149	6.4%	173.4 ± 12.7	-1.91
4	171	6.3%	170.5 ± 12.6	0.04
5	162	6.1%	164.4 ± 12.4	-0.19
6	153	5.9%	160.1 ± 12.3	-0.58
7	160	5.5%	150.2 ± 11.9	0.82
8	155	5.3%	143.8 ± 11.7	0.96
9	141	5.1%	138.6 ± 11.5	0.21
10	136	5.1%	137.2 ± 11.4	-0.10
11	149	5.0%	136.6 ± 11.4	1.09
12	130	5.2%	141.2 ± 11.6	-0.97
13	152	4.7%	126.8 ± 11.0	2.29
14	126	4.8%	129.1 ± 11.1	-0.28
15	123	4.3%	117.2 ± 10.6	0.55
16	96	4.1%	110.3 ± 10.3	-1.39
17	114	4.1%	111.7 ± 10.3	0.22
18	105	4.1%	112.3 ± 10.4	-0.70
19	62	2.9%	79.0 ± 8.8	-1.94
Total	2712	100.0%	2712.0	

Table 2 (continued)

Autosomes vs. X[f]

Autosomes	2712	93.7%	2640.7 ± 12.0	5.94
X	86	6.3%	159.3 ± 12.0	-5.94
Total	2798	100.0%	2798.0	

a. Cytogenetic length taken from Evans (1989).
b. Only random markers are considered to avoid biases in chromosomal distribution of known genes.
c. Mean ± standard deviation.
d. Z-score = (observed-expected)/standard deviation.
e. None of the Z-scores are significant at the p=0.05, after correction for multiple testing (Bonferroni correction for 19 tests).
f. Expectation reflects the fact that 12% of the random markers were derived from male DNA (thus underrepresentingthe X chromosome by a factor of two) and 88% from female DNA. Z-score is significant at $p < 0.0001$.

Table 3. Distribution of number of crossovers between consecutive random markers[a]

Crossovers per Interval	Observed		Expected[b]			P(longest run ≥ n)[b]
	Number	Percentage	Number		Percentage	
0	1928	69.9%	1878.1 ±	24.5	69.7%	100%
1	542	19.7%	598.7 ±	21.6	21.1%	100%
2	185	6.7%	190.9 ±	13.3	6.4%	100%
3	61	2.2%	60.8 ±	7.7	1.9%	100%
4	26	0.9%	19.4 ±	4.4	0.6%	100%
5	8	0.3%	6.2 ±	2.5	0.2%	95%
6	5	0.2%	2.0 ±	1.4	0.1%	61%
7	1	<0.1%	0.6 ±	0.8	<0.1%	26%
8	0	0.0%	0.2 ±	0.4	<0.1%	9%
9	0	0.0%	0.1 ±	0.3	<0.1%	3%
10	0	0.0%	0.0 ±	0.1	<0.1%	1%
11	1	<0.1%	0.0 ±	0.1	<0.1%	0.3%
Total	2757					

a. Only random markers are considered to avoid biases in distribution of known genes.
b. See methodology concerning calculation.

Table 4. Distribution of number of random markers occurring between consecutive crossovers[a]

Markers per Block	Observed		Expected[b]		P(longest run ≥ n)[b]
	Number	Percentage	Number	Percentage	
0	485	36.4%	425.3 ± 17.0	31.9%	100%
1	284	21.3%	289.7 ± 15.1	21.7%	100%
2	170	12.7%	197.3 ± 13.0	14.8%	100%
3	105	7.9%	134.4 ± 11.0	10.1%	100%
4	80	6.0%	91.6 ± 9.2	6.9%	100%
5	58	4.3%	62.4 ± 7.7	4.7%	100%
6	44	3.3%	42.5 ± 6.4	3.2%	100%
7	36	2.7%	29.0 ± 5.3	2.2%	100%
8	23	1.7%	19.7 ± 4.4	1.5%	100%
9	14	1.0%	13.4 ± 3.6	1.0%	100%
10	13	1.0%	9.2 ± 3.0	0.7%	100%
11	6	0.4%	6.2 ± 2.5	0.5%	100%
12	8	0.6%	4.2 ± 2.1	0.3%	100%
13	4	0.3%	2.9 ± 1.7	0.2%	100%
14	0	0.0%	2.0 ± 1.4	0.1%	98%
15	2	0.1%	1.3 ± 1.2	0.1%	94%
16	0	0.0%	0.9 ± 1.0	0.1%	85%
17	0	0.0%	0.6 ± 0.8	<0.1%	73%
18	0	0.0%	0.4 ± 0.7	<0.1%	59%
19	0	0.0%	0.3 ± 0.5	<0.1%	45%
20	1	0.1%	0.2 ± 0.4	<0.1%	34%
21	0	0.0%	0.1 ± 0.4	<0.1%	24%
22	0	0.0%	0.1 ± 0.3	<0.1%	17%
23	0	0.0%	0.1 ± 0.2	<0.1%	12%
24	1	0.1%	0.0 ± 0.2	<0.1%	8%
Total	1334				

a. Only random markers are considered to avoid biases in distribution of known genes.
b. See methodology concerning calculation.

Table 6. Polymorphism rate for 3,012 markers by chromosome

Chromosome	Among lab strains[a,b]	Lab strains vs. SPR or CAST[c]
1	54%	94%
2	45%	93%
3	50%	94%
4	53%	94%
5	47%	96%
6	47%	94%
7	45%	93%
8	42%	93%
9	50%	94%
10	35%	96%
11	54%	95%
12	49%	92%
13	47%	94%
14	48%	94%
15	51%	93%
16	43%	95%
17	57%	92%
18	51%	95%
19	49%	93%
X	29%	94%
Genome-wide	49%	94%

a. Pairwise comparisons of OB, B6, DBA, A C3H, BALB, AKR, NON, NOD, and LP.
b. Standard error of the mean for each chromosome depends on number of markers studied, but is approximately 0.6% for typical chromosomes and 0.1% for the genome-wide average.
c. Standard error of the mean for each chromosome depends on number of markers studied, but is approximately 1.7% for typical chromosomes and 0.4% for the genome-wide average.

Table 5. Rate of polymorphism for 3,012 markers among 12 mouse strains[a,b]

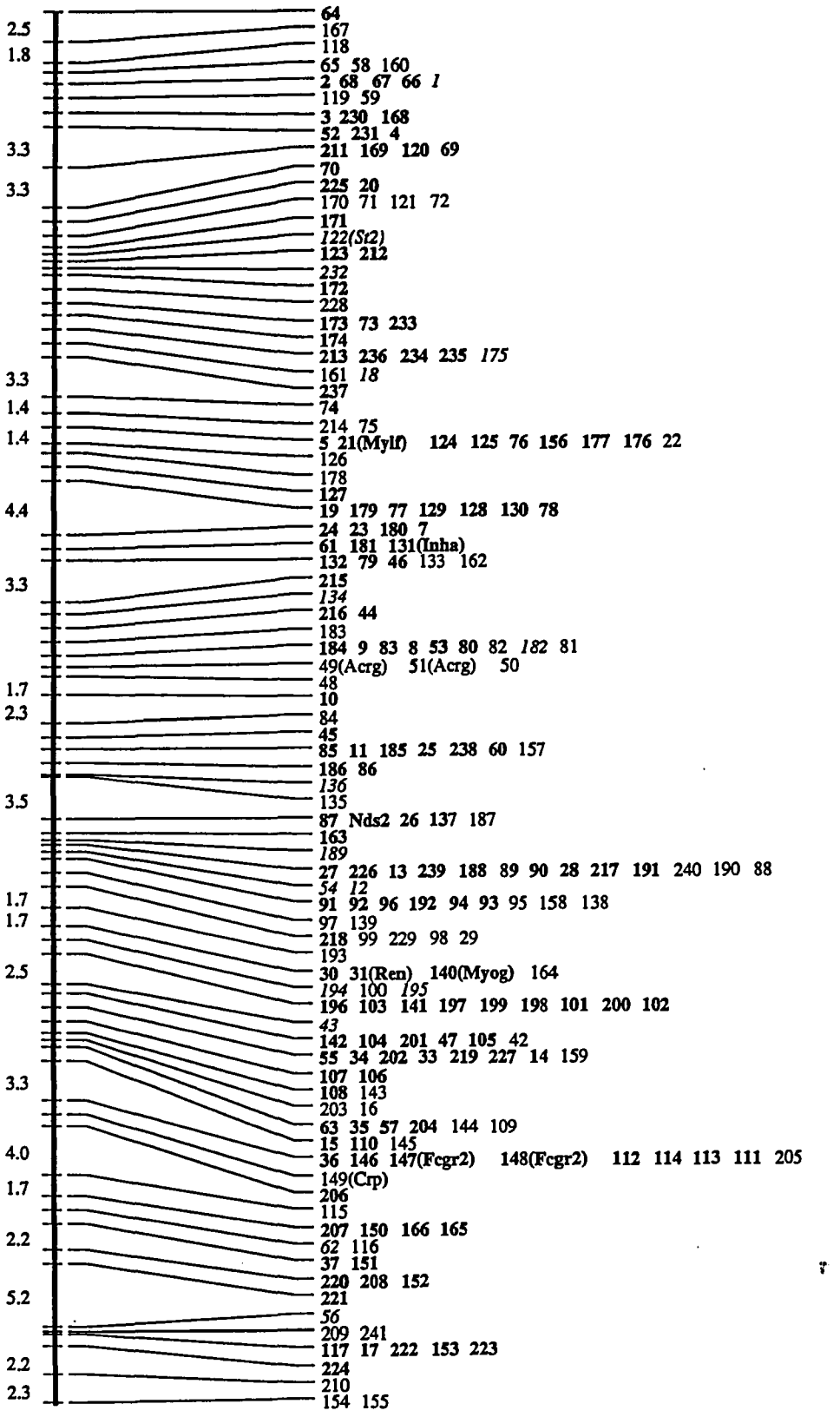
	OB	B6	DBA	A	C3H	BALB	AKR	NON	NOD	LP	SPR	CAST
OB	-											
B6	10%	-										
DBA	54%	51%	-									
A	55%	53%	47%	-								
C3H	55%	52%	38%	33%	-							
BALB	53%	51%	46%	32%	37%	-						
AKR	54%	52%	48%	45%	44%	44%	-					
NON	55%	52%	52%	48%	48%	48%	49%	-				
NOD	56%	53%	49%	51%	49%	50%	49%	45%	-			
LP	56%	54%	51%	53%	50%	49%	52%	51%	52%	-		
SPR	93%	93%	92%	93%	93%	93%	93%	92%	92%	93%	-	
CAST	100%	98%	94%	94%	95%	95%	95%	95%	95%	94%	95%	-

a. Strains designations are: OB = C57BL/6J-ob/ob, B6 = C57BL/6J, DBA = DBA/2J, A = A/J, C3H = C3H/HeJ, BALB = BALB/cJ, AKR = AKR/J, NON = NON/Lt, NOD = NOD/MrkTacBr, LP = LP/J, SPR = SPRET/Ei, CAST = CAST/Ei.

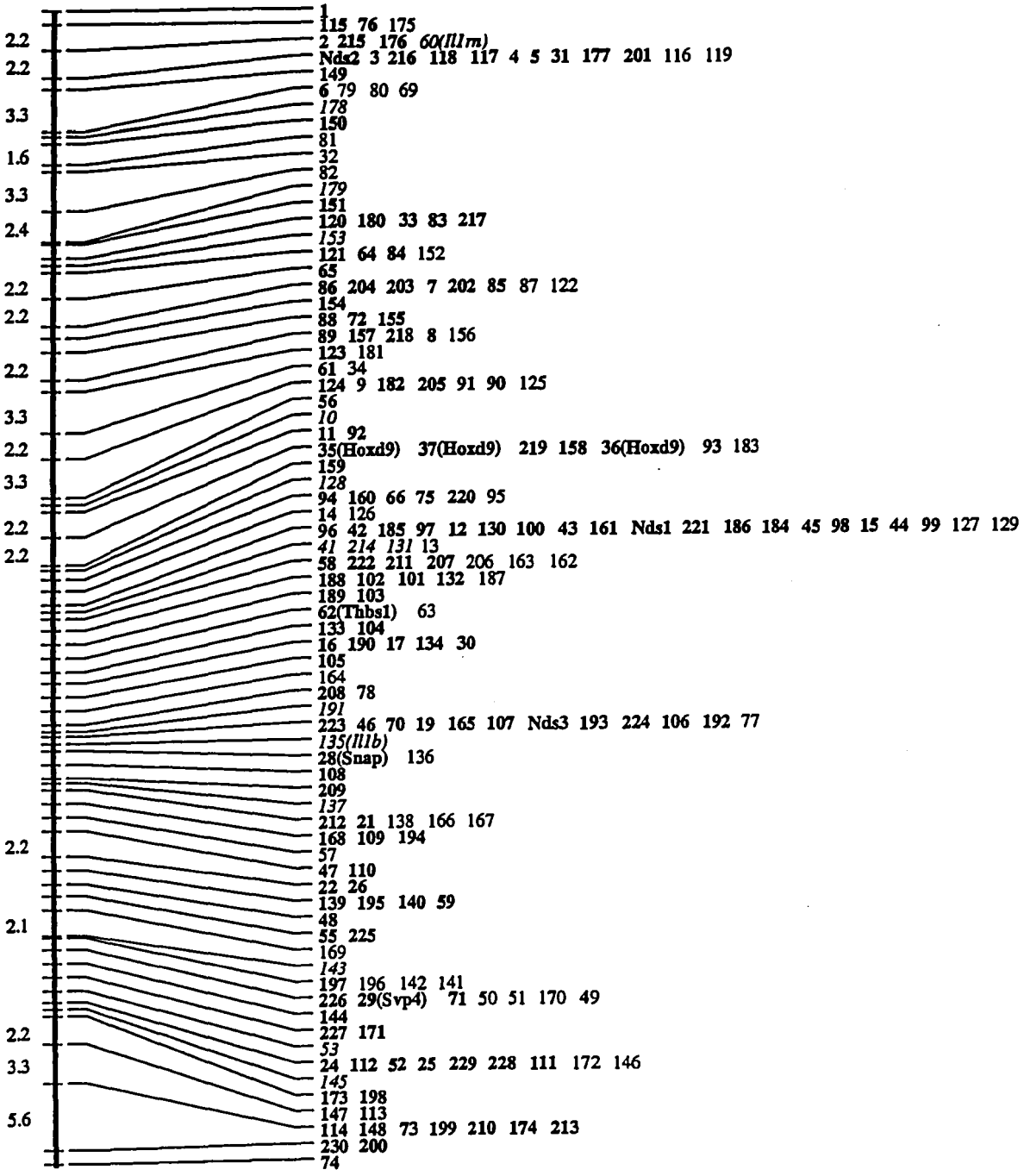
b. Standard error of the mean is approximately 0.9% for rates near 50% and 0.4% for rates near 95%.,

Chromosome 1

Figure 1 consists of twenty (20) page

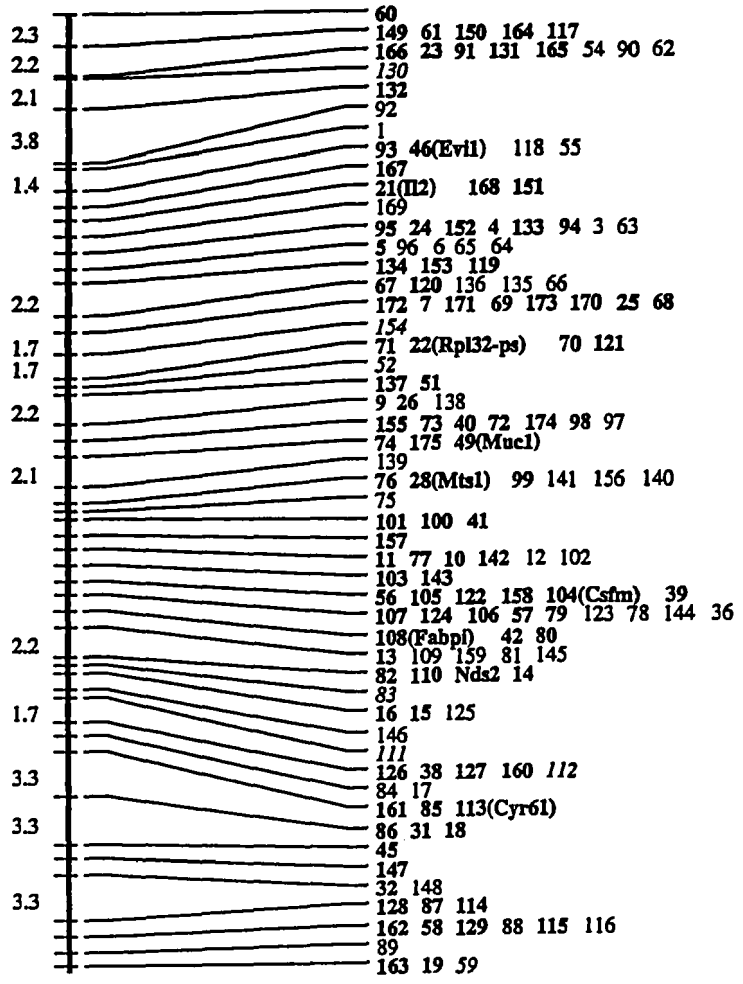


Chromosome 2



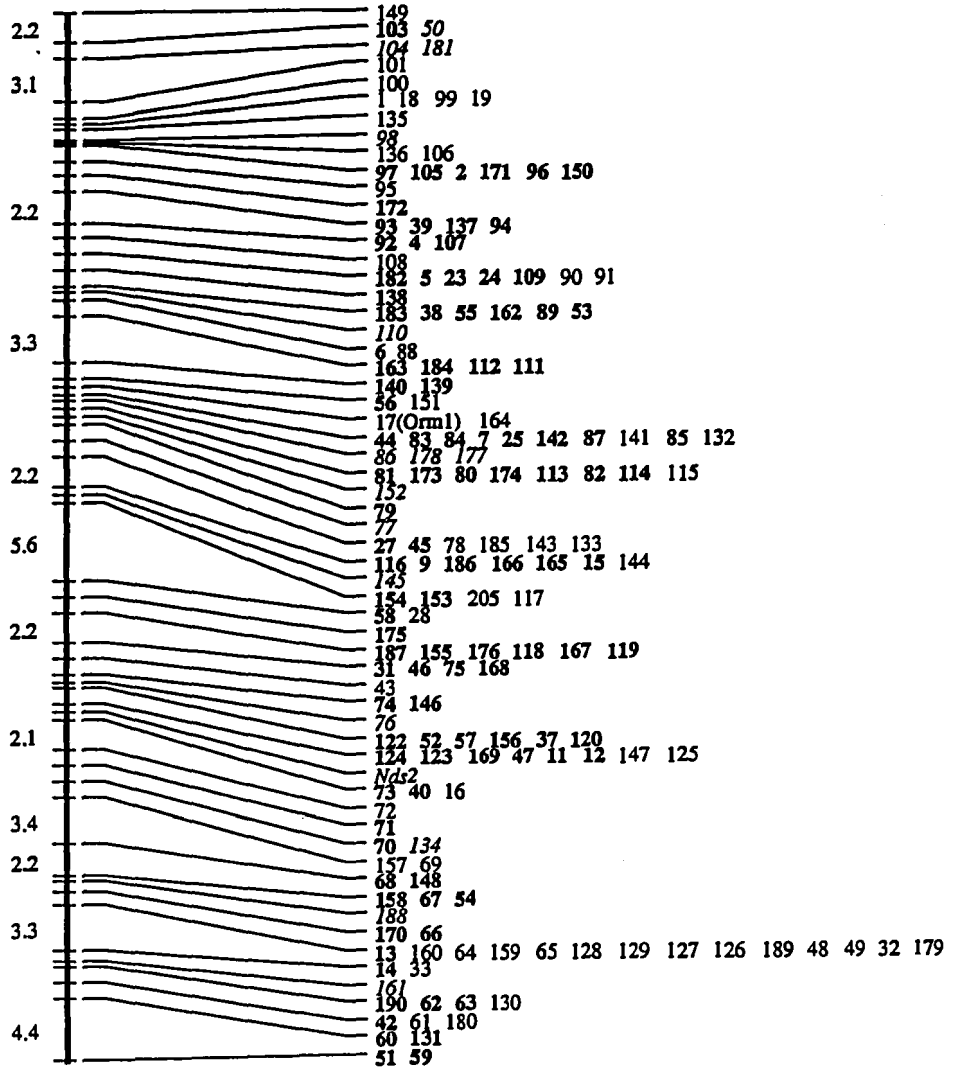
5.4 cM

Chromosome 3

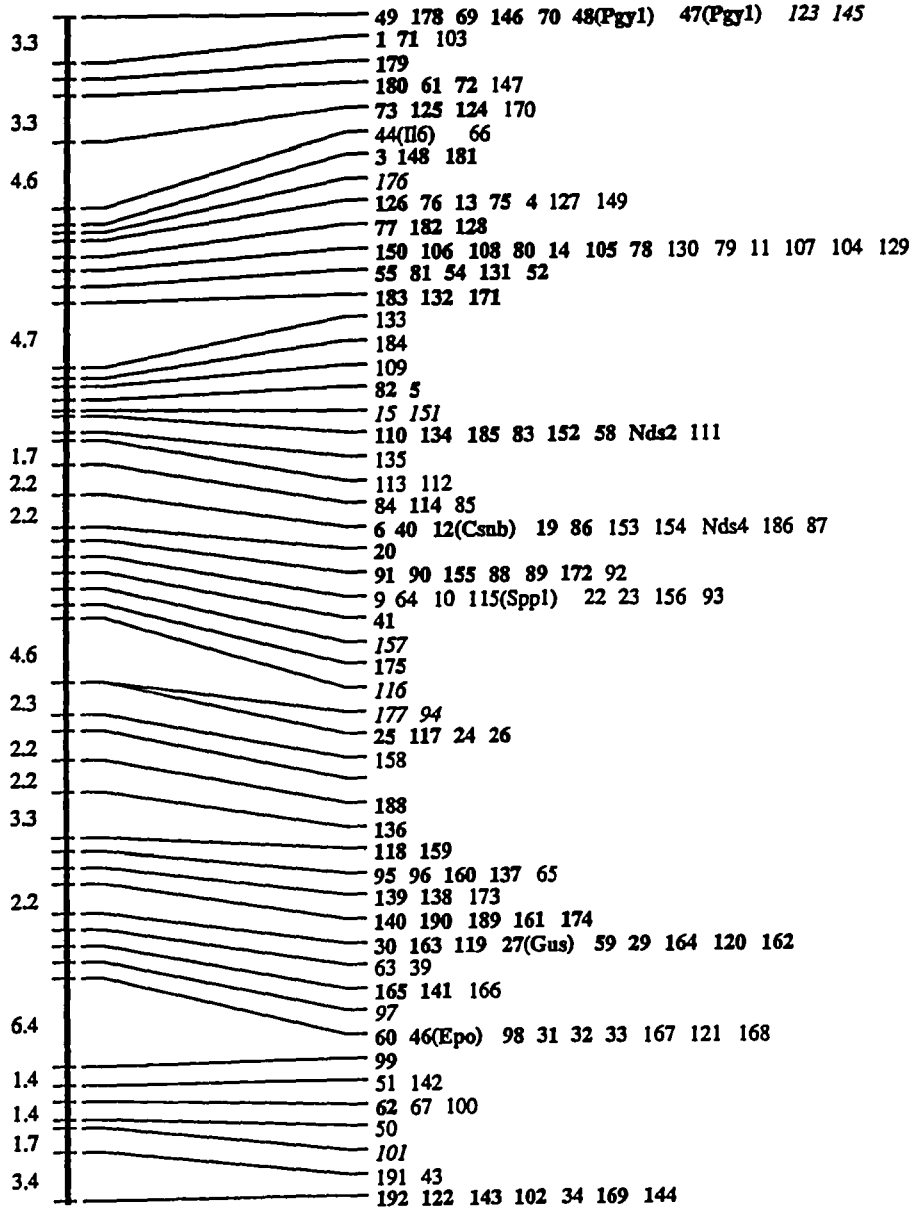


5.4 cM

Chromosome 4

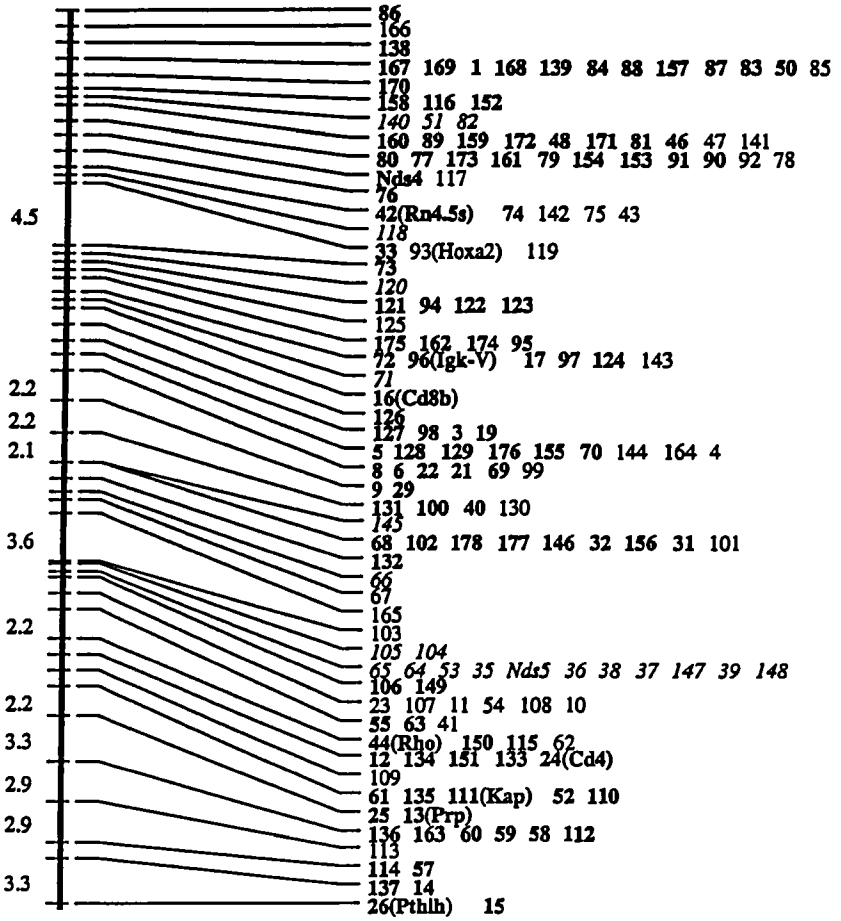


Chromosome 5



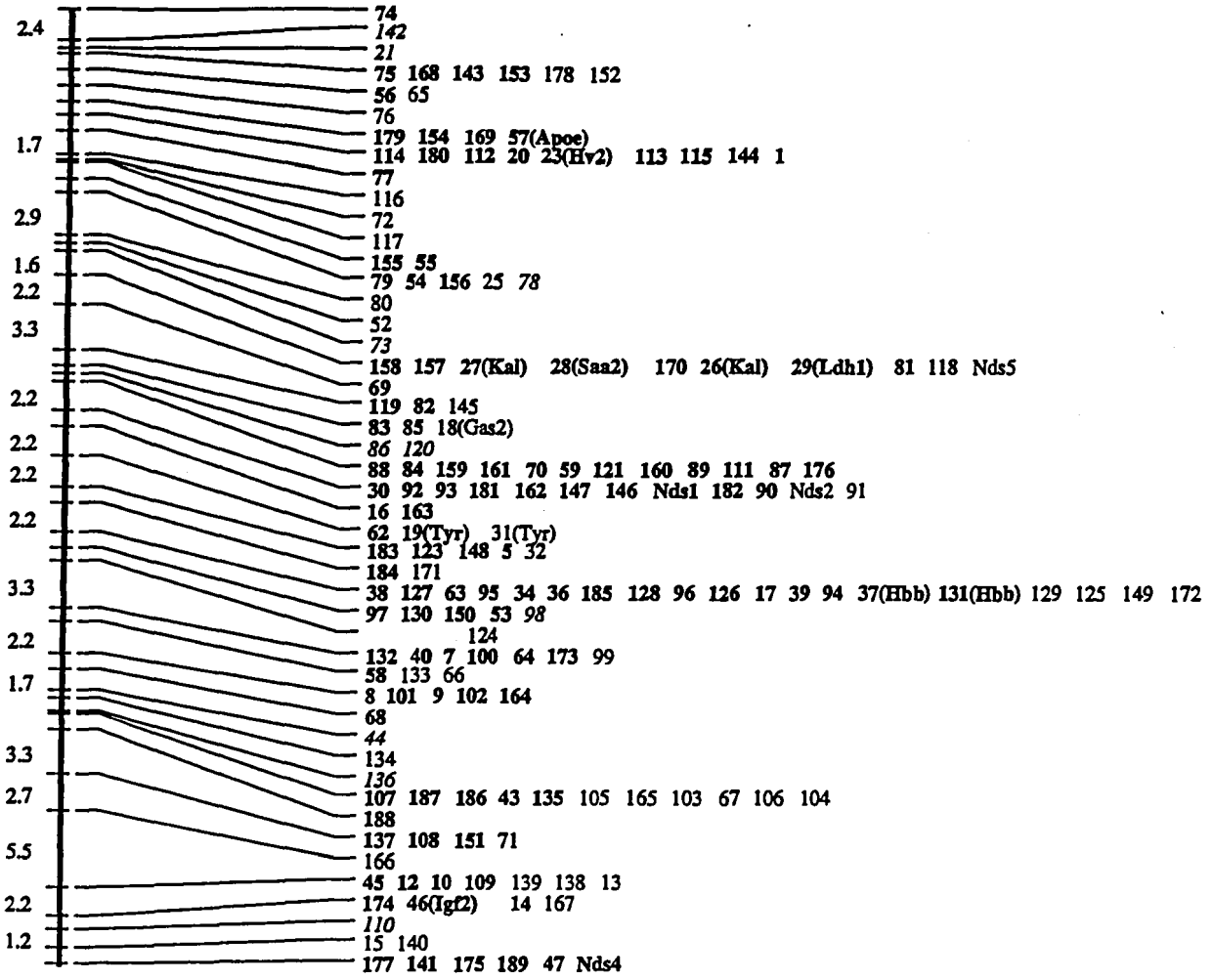
5.4 cM

Chromosome 6

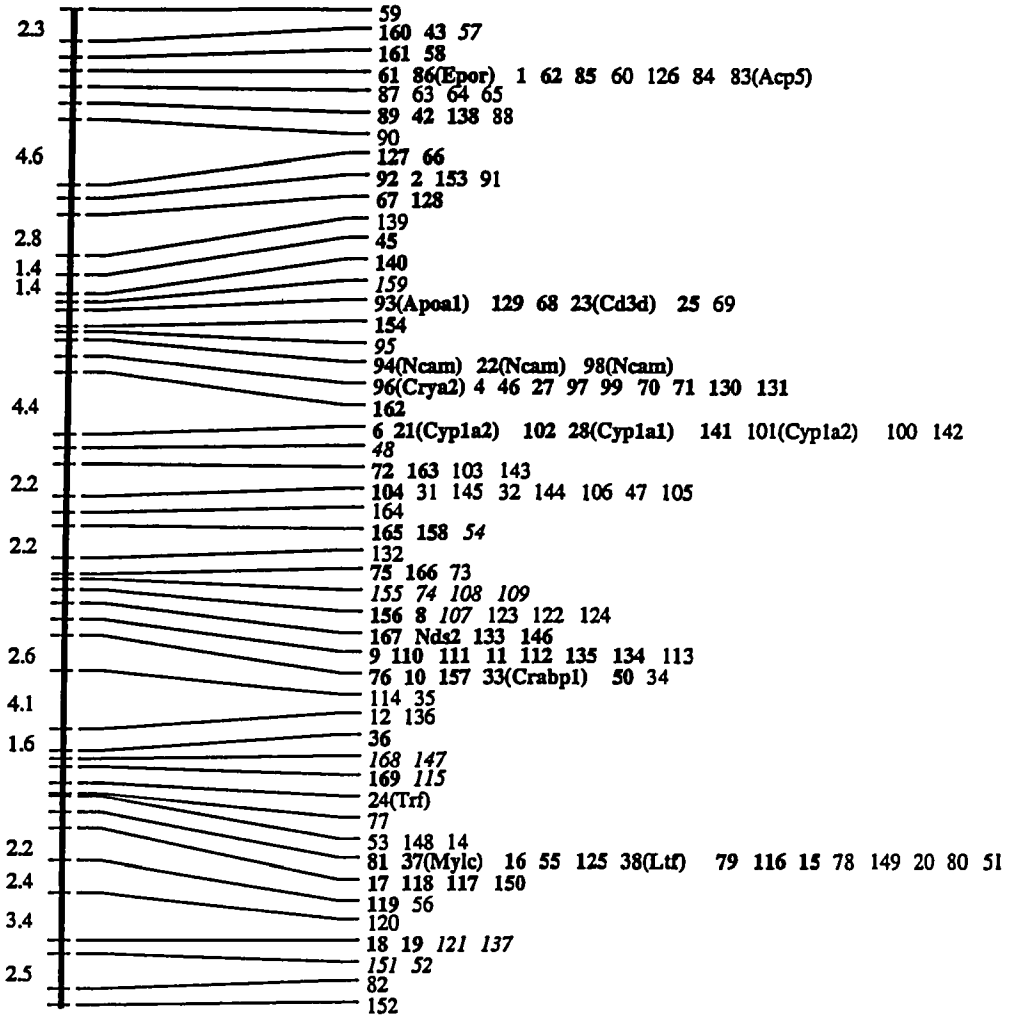


5.4 cM

Chromosome 7

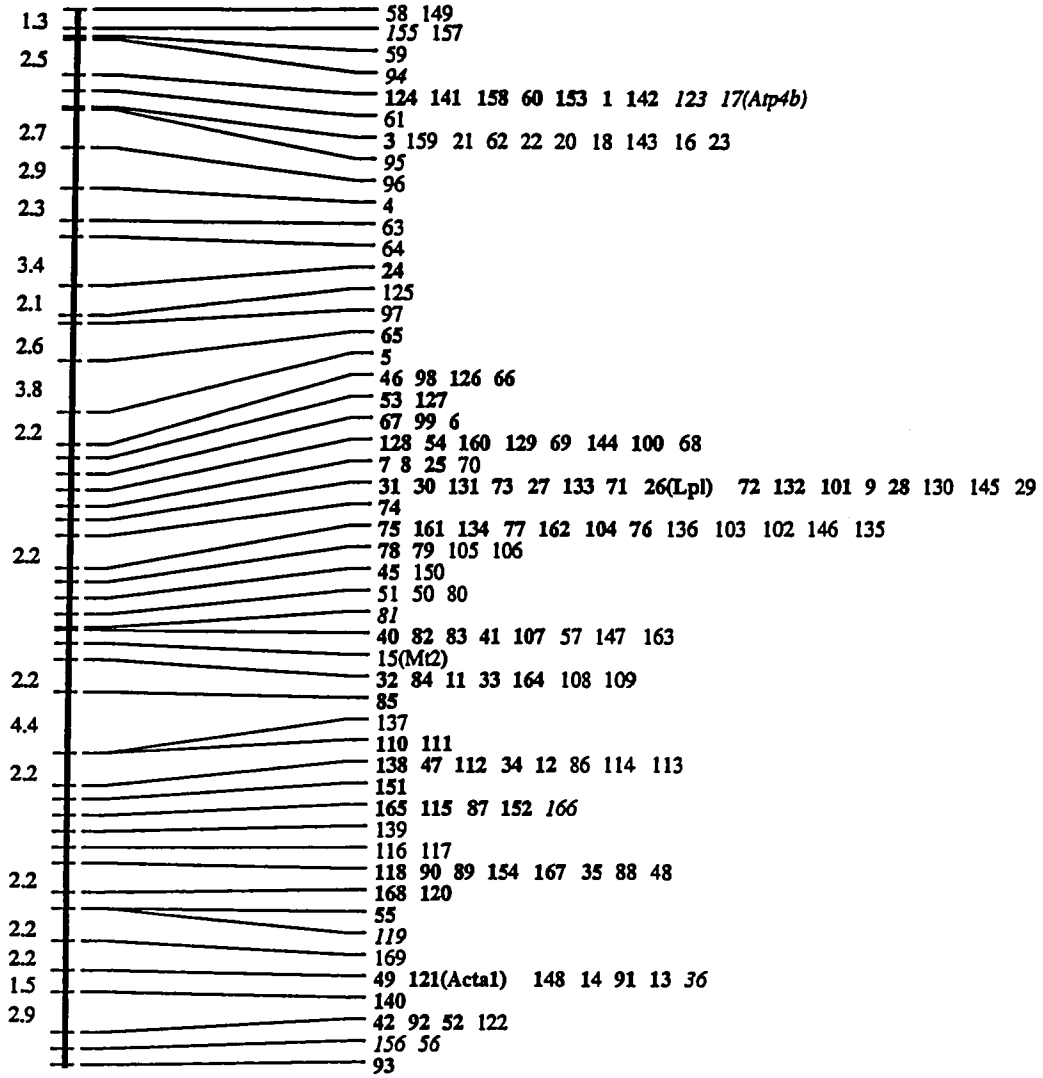


Chromosome 9



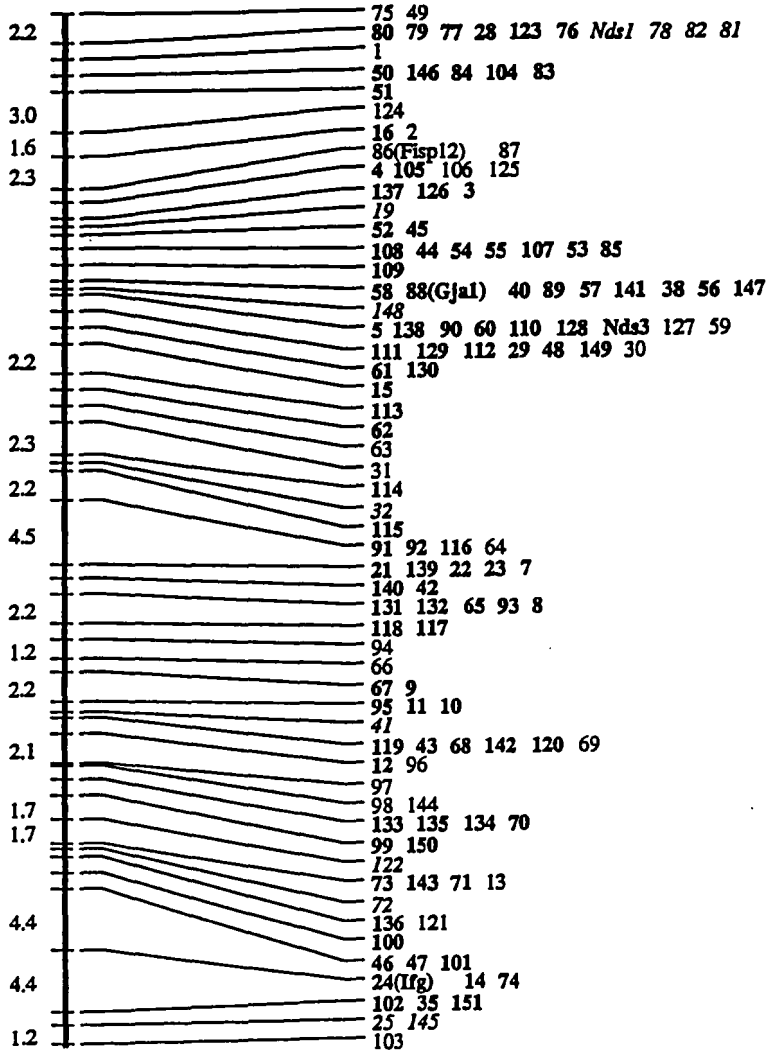
5.4 cM

Chromosome 8



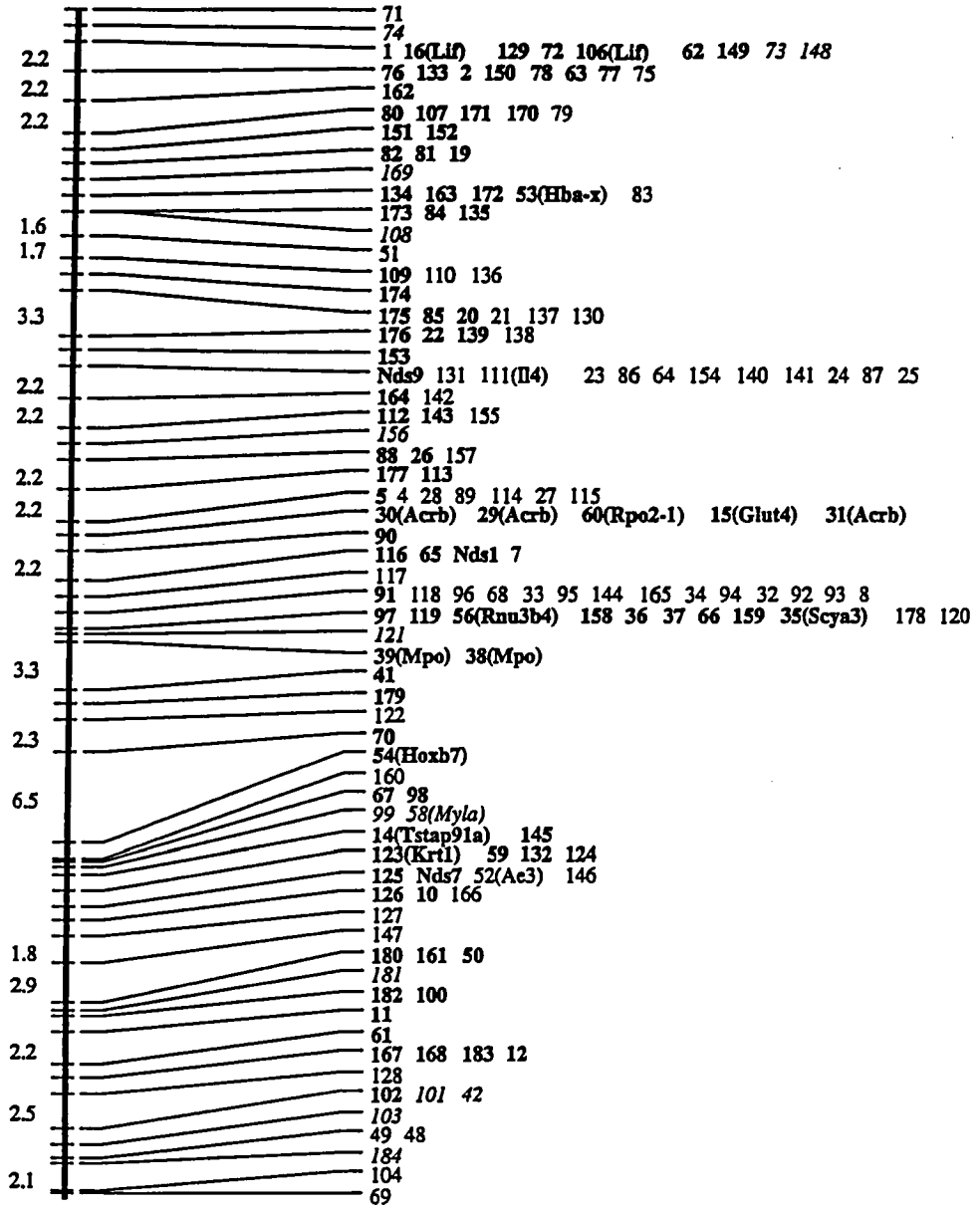
5.4 cM

Chromosome 10



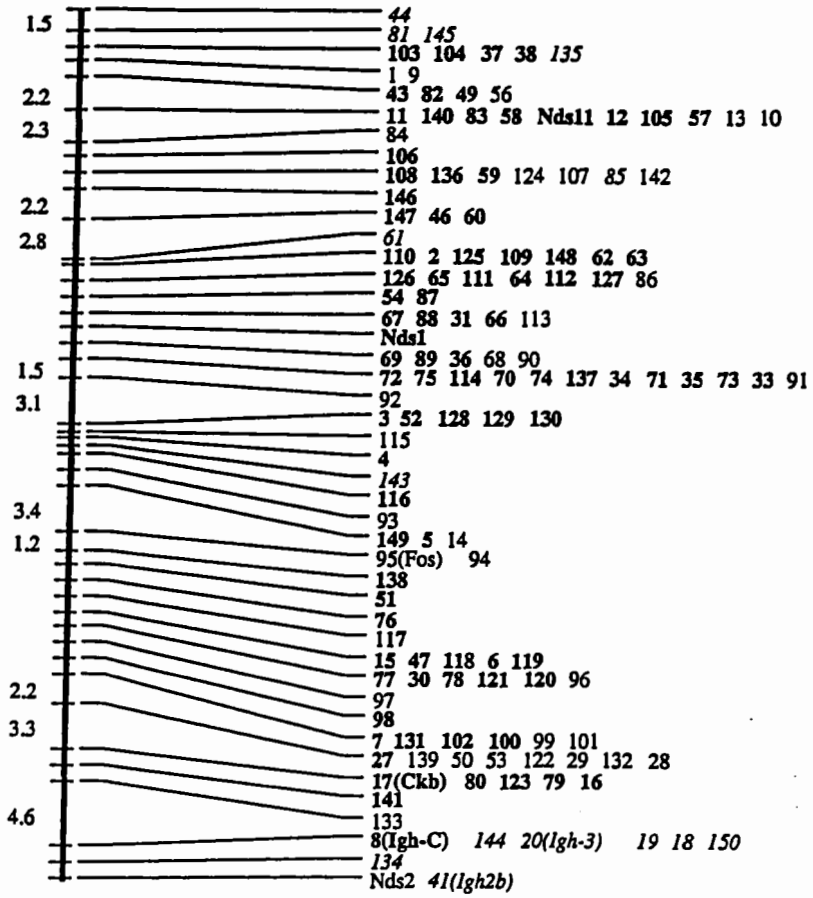
5.4 cM

Chromosome 11



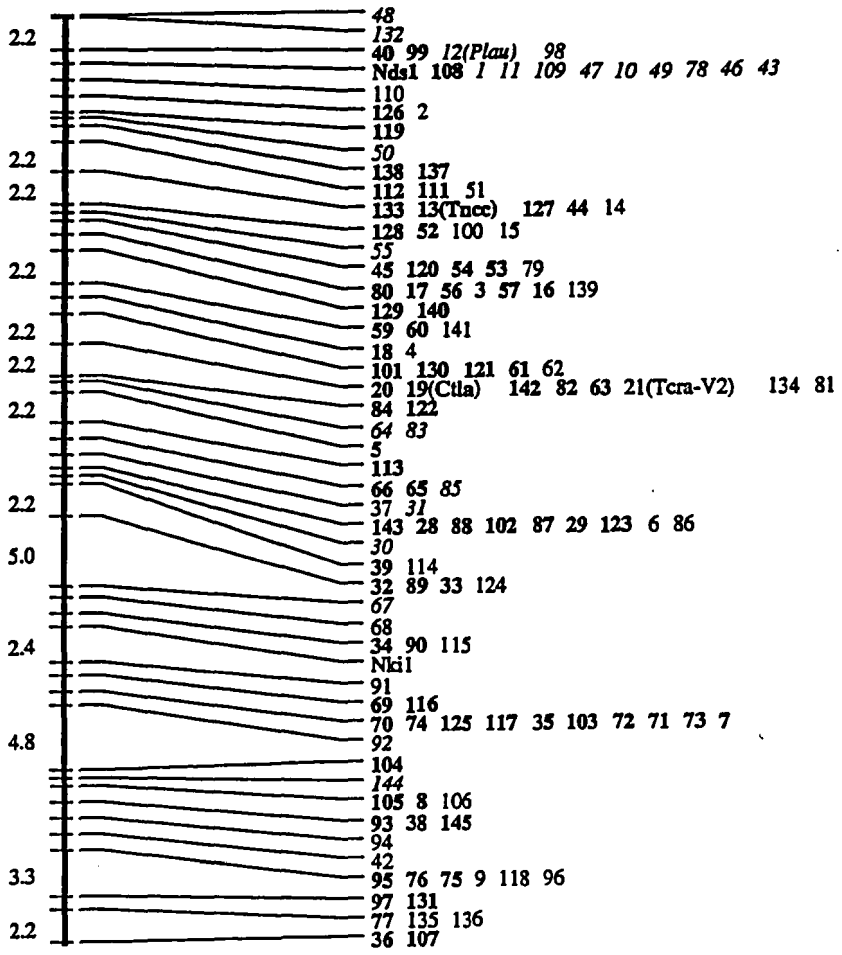
5.4 cM

Chromosome 12



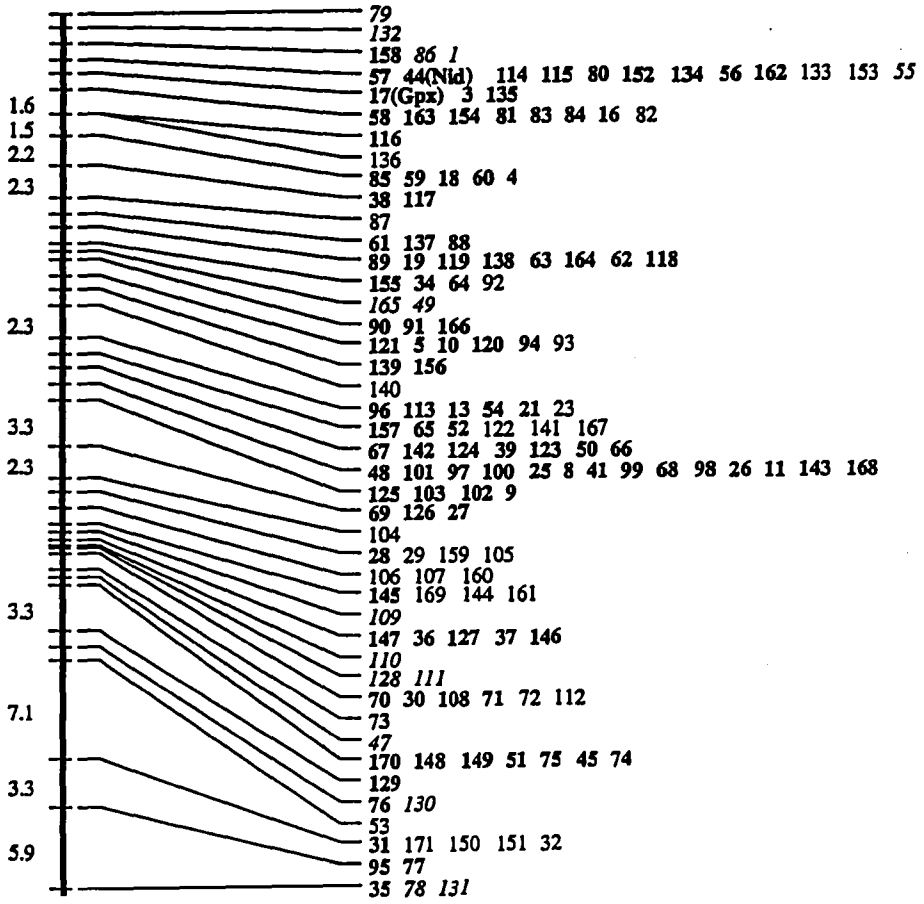
5.4 cM

Chromosome 14



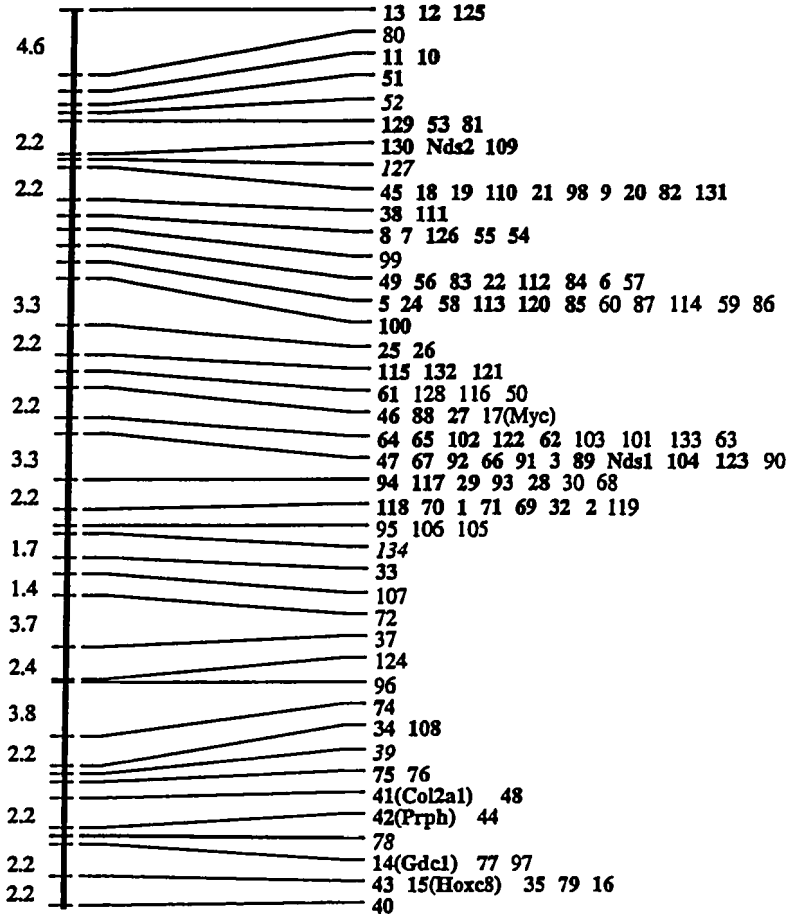
5.4 cM

Chromosome 13



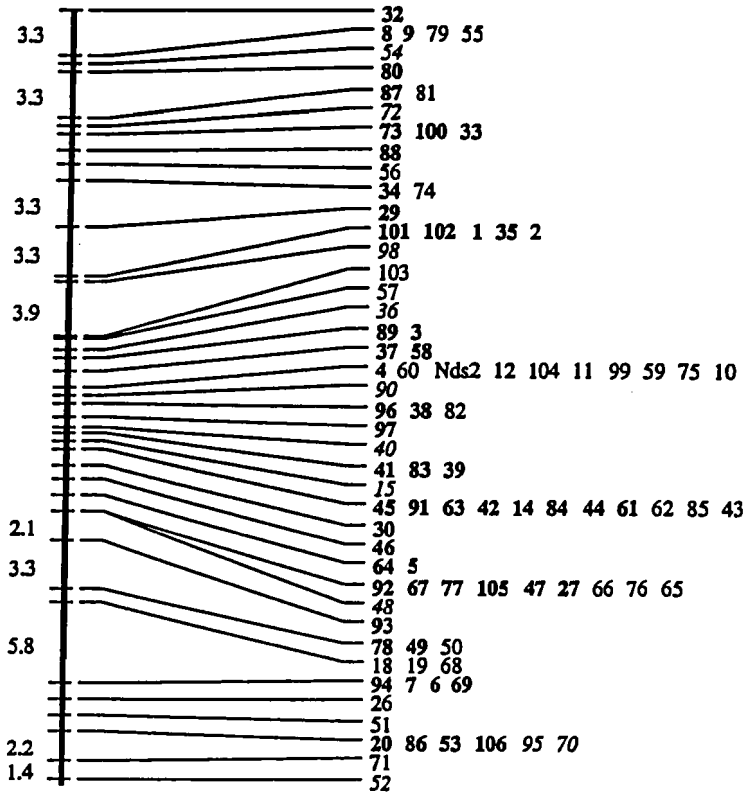
5.4 cM

Chromosome 15



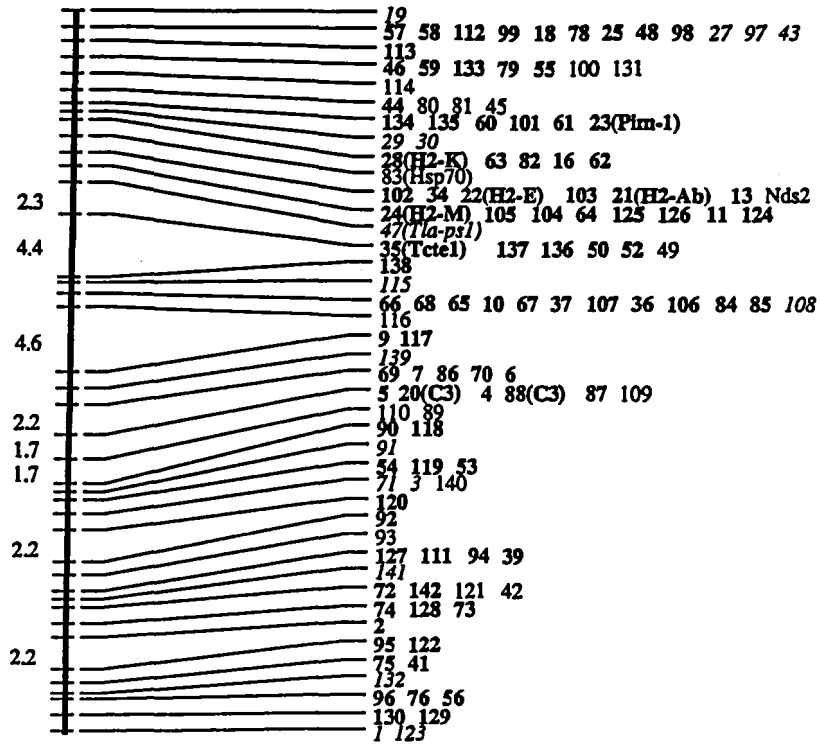
5.4 cM

Chromosome 16



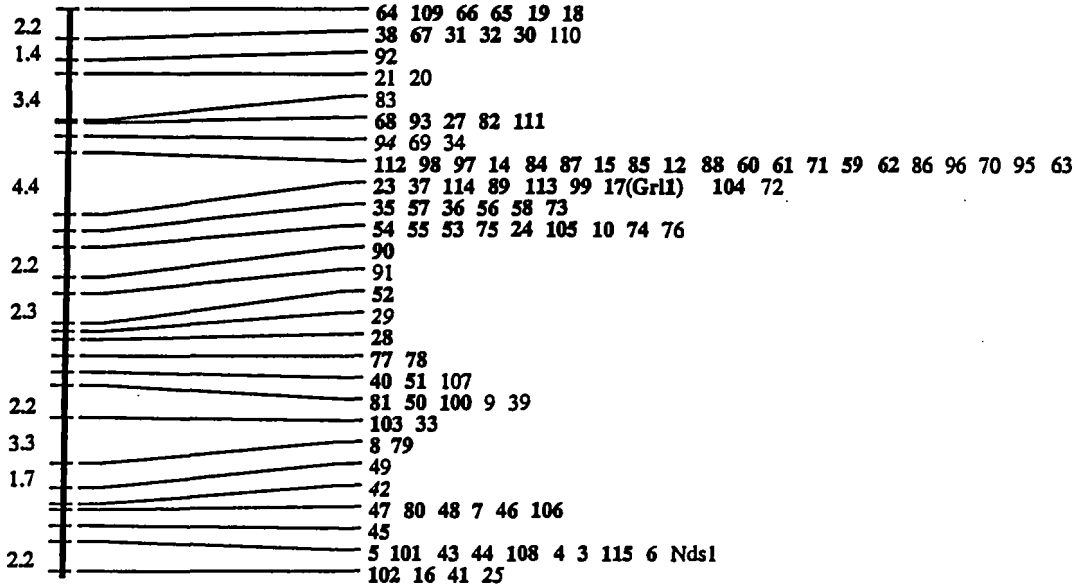
5.4 cM

Chromosome 17



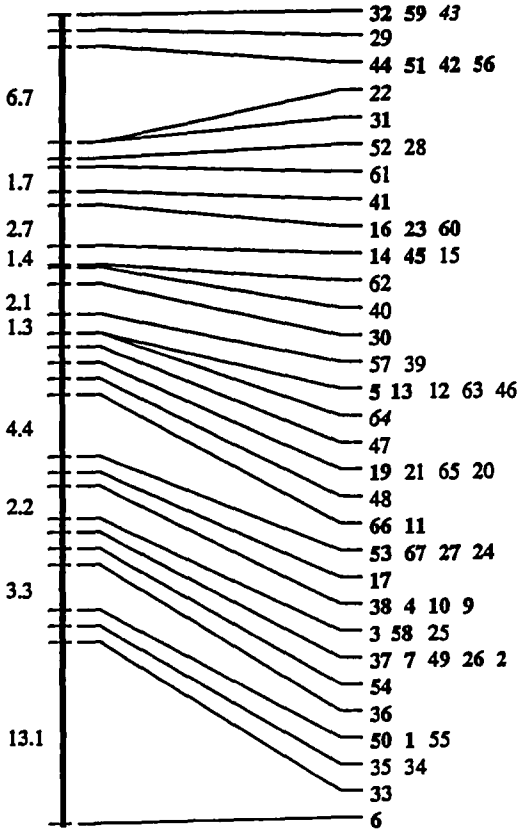
5.4 cM

Chromosome 18



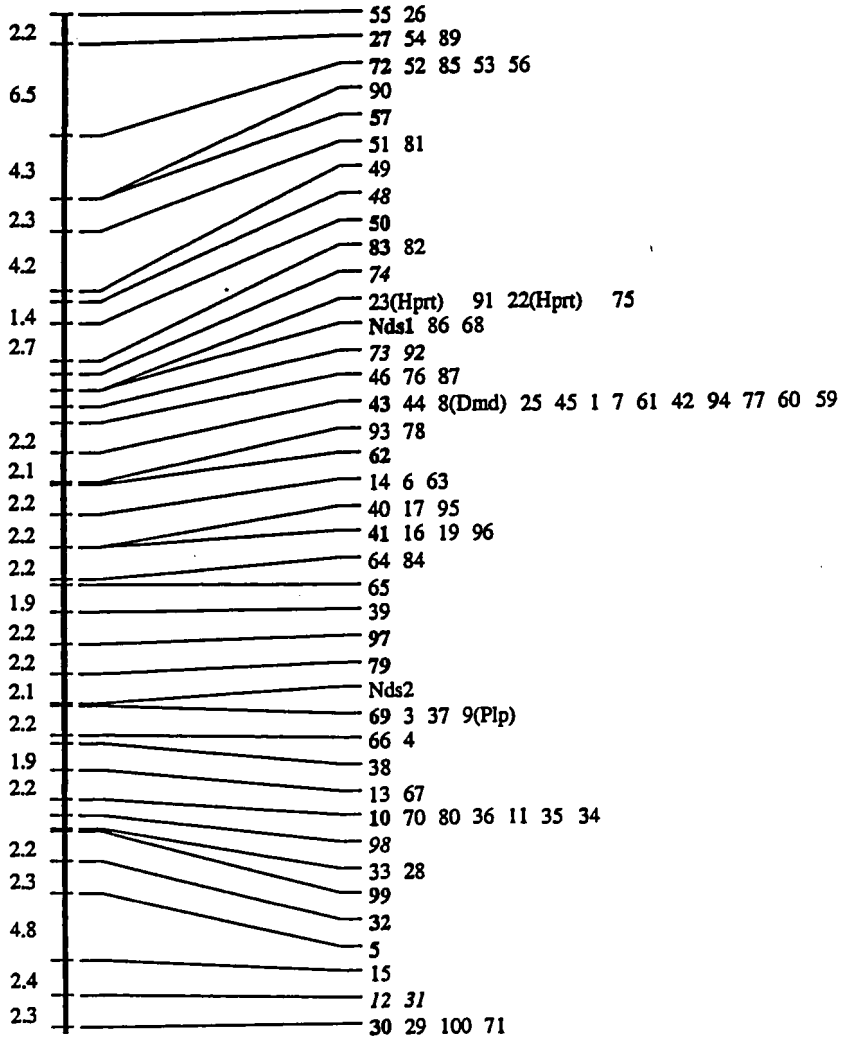
5.4 cM

Chromosome 19

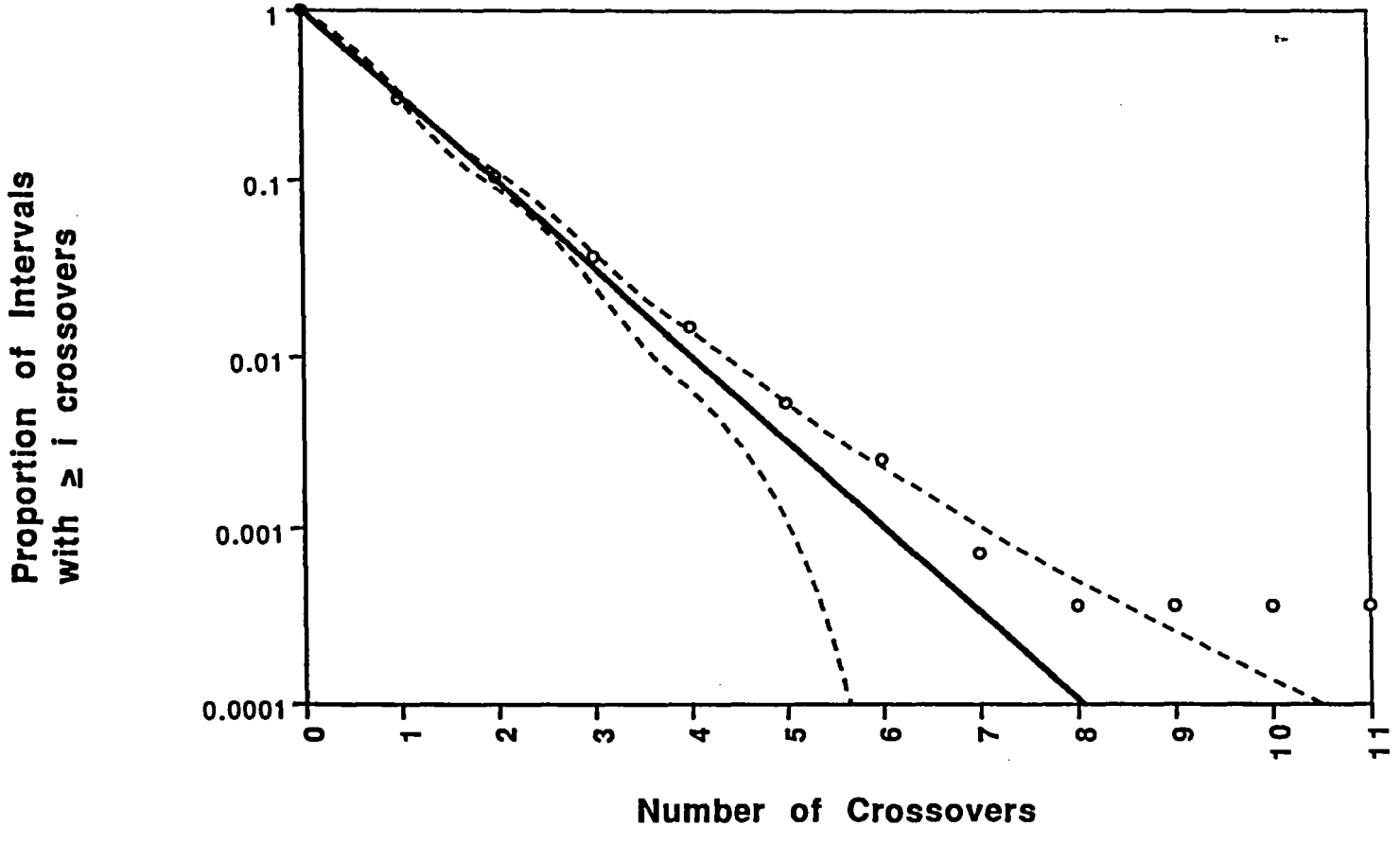


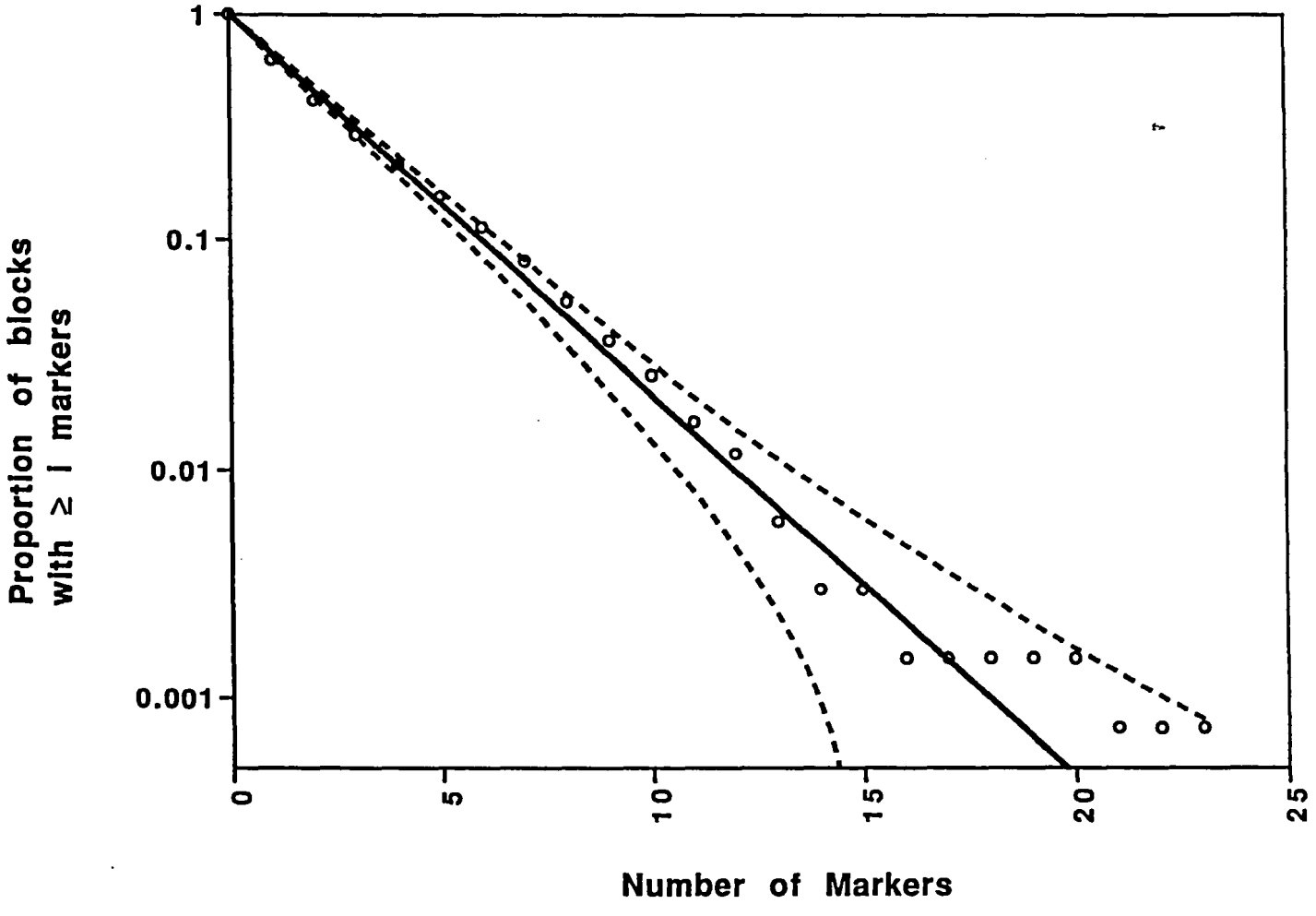
5.4 cM

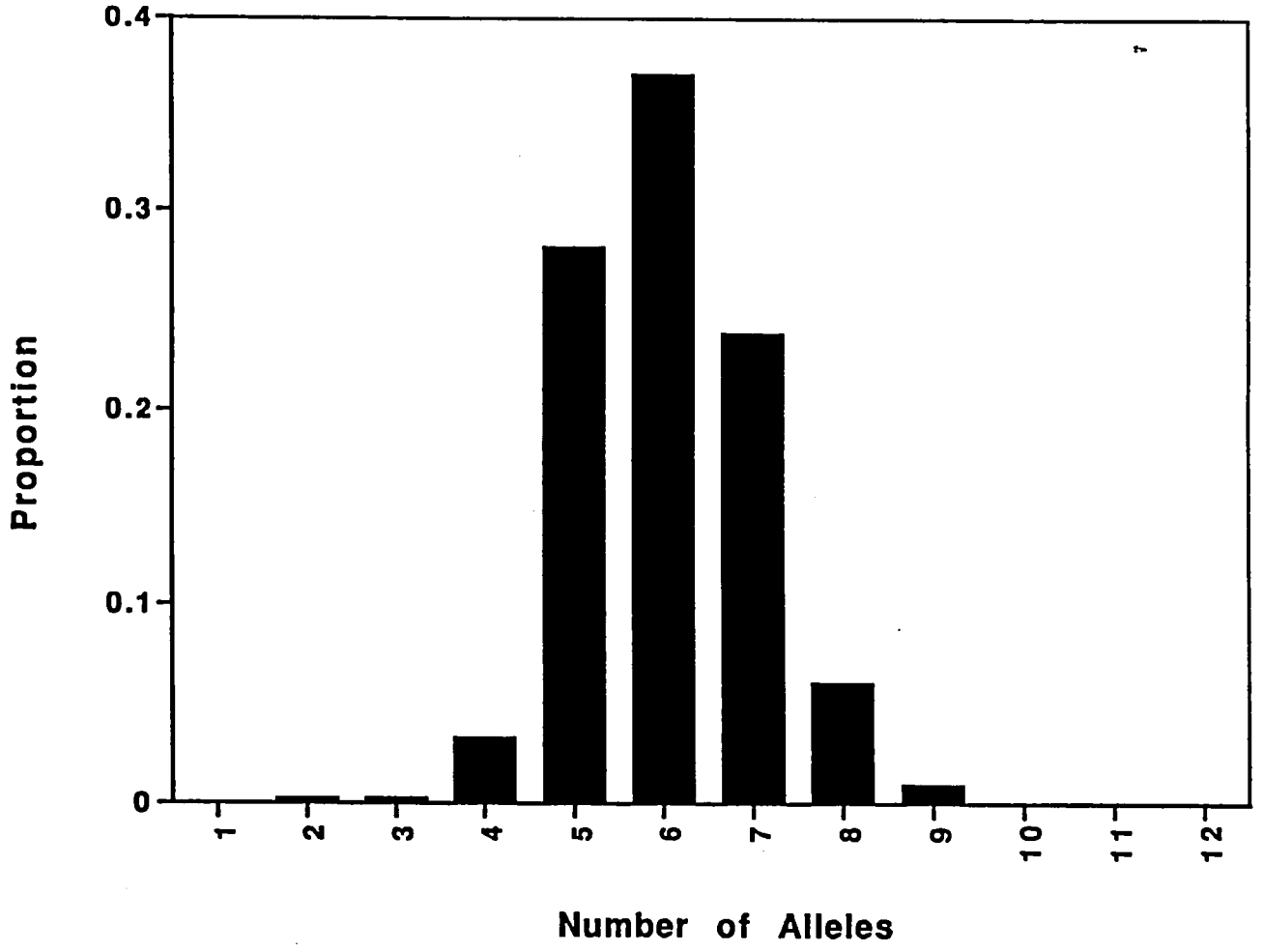
Chromosome X



5.4 cM


Proportion of Intervals with ≥ i crossovers
1
0.1
0.01
0.001
0.0001
0
1
2
3
4
5
6
7
8
9
10
11
Number of Crossovers


Proportion of blocks
with ≥ l markers
1
0.1
0.01
0.001
0
5
10
15
20
25
Number of Markers


0.4
0.3
0.2
0.1
0
Proportion
1
2
3
4
5
6
7
8
9
10
11
12
Number of Alleles

GENOMICS 13, 622–629 (1992)

Isolation and Chromosomal Assignment of 100 Highly Informative Human Simple Sequence Repeat Polymorphisms

THOMAS J. HUDSON,*·† MARCY ENGELSTEIN,* MATTHIAS K. LEE,* ELIZABETH C. HO,* MARC J. RUBENFIELD,*·[1] CHRISTOPHER P. ADAMS,* DAVID E. HOUSMAN,*·† AND NICHOLAS C. DRACOPOLI*·[2]

*Center for Genome Research, and †Center for Cancer Research, Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139

Received January 29, 1992; revised March 2, 1992

One hundred highly informative simple sequence repeat (SSR) polymorphisms have been isolated and mapped to specific human chromosomes by somatic cell hybrid analysis. These markers include 97 $(CA)_n$, 2 $(AGAT)_n$, and a single $(AACT)_n$ repeat. All the SSRs have heterozygosities > 0.50 and can be amplified using identical PCR conditions. At least one SSR was detected on every chromosome, except for chromosomes 22 and Y. The frequency of $(CA)_n$ repeats on each chromosome was proportional to the relative chromosomal length, except for chromosome 15, on which a substantial excess of markers was identified. © 1992 Academic Press, Inc.

INTRODUCTION

The development of genetic maps of eukaryotic genomes has always been limited by the availability of markers. The earliest maps of human autosomes were developed using a variety of phenotypic markers, including blood group antigens, serum protein polymorphisms, and erythrocyte isozymes (Mohr, 1954; Giblett, 1969). These markers were generally quite uninformative and required a diverse range of biochemical and immunological techniques for their analysis. This situation changed rapidly after Botstein *et al.* (1980) observed that DNA sequence polymorphisms provided an enormous, untapped source of variation in eukaryotic genomes. Markers based on restriction site polymorphisms proliferated rapidly during the early 1980's and were soon supplemented by the more informative minisatellites (Jeffreys *et al.*, 1985) or variable number of tandem repeat (VNTR) polymorphisms (Nakamura *et al.*, 1987). Low-resolution maps of the human genome (Donis-Keller *et al.*, 1987) have been completed using RFLPs and minisatellites, but these maps are limited in their usefulness because of the relatively low informativeness of most RFLPs ($H < 0.50$), and the nonrandom distribution of the minisatellites in the telomeric regions of human chromosomes (Royle *et al.*, 1988). The recent discovery that simple sequence repeats (SSRs), or microsatellites, are highly informative has provided another source of DNA polymorphisms (Weber and May, 1989). SSRs are widely dispersed throughout eukaryotic genomes (Hamada *et al.*, 1982; Stallings *et al.*, 1991), highly polymorphic (Weber and May, 1989; Weber, 1990), and easily typed using the polymerase chain reaction (PCR). These characteristics make them ideal markers for the construction of high-resolution maps of the human genome (White *et al.*, 1990; Dracopoli *et al.*, 1991). In this article, we report the development of 100 highly informative markers ($H > 0.50$) in the human genome that can all be analyzed using identical PCR conditions.

METHODS AND MATERIALS

Construction of small insert genomic libraries. Human female genomic DNA was digested with *Rsa*I and *Hae*III. Fragments in the 300- to 500-bp range were isolated after electrophoresis in a 1% low-melting-temperature agarose gel and purified by phenol extraction. The size-fractionated human genomic DNA was ligated into 10 multiplex sequencing vectors (Church and Kieffer-Higgins, 1988) and into M13mp19. Both the multiplex plasmid and M13mp19 vector DNA were prepared by digestion with *Sma*I and dephosphorylated by treatment with calf intestinal alkaline phosphatase (Sambrook *et al.*, 1989).

Library screening for clones containing $(CA)_n$ and tetranucleotide repeats. The multiplex plasmid libraries were plated at low density (100–500 colonies per 150-mm plate) on LB plates with tetracycline (50 μg/ml). Colonies were replicated onto nylon filters and autoclaved for 3 min before prehybridization at 65°C in 1.0 *M* sodium phosphate (pH 7.2), 0.5 *M* EDTA, 20% SDS, 10% BSA (Church and Gilbert, 1984) for 1–4 h. Labeling reactions using T4 polynucleotide kinase were carried out as described in Sambrook *et al.* (1989), and the unincorporated [γ-^{32}P]ATP was removed on a Nuctrap probe purification column (Stratagene Inc., La Jolla, CA). The filters were hybridized overnight at 65°C in the same buffer after the addition of 1×10^6 cpm/ml of end-labeled $(CA)_{15}$ oligonucleotides. The filters were washed at room temperature in two changes of 1× SSC/0.1% SDS for 30–60 min and then for 60 min in two changes of the same buffer at 65°C. The washed filters were exposed to X-ray film overnight, and positive colonies were picked onto a secondary plate for rescreening.

The M13mp19 libraries were also plated out at low density on LB plates by standard methods (Sambrook *et al.*, 1989). Replica filters of

[1] Present address: Collaborative Research, Inc., 2 Oak Park, Bedford, MA 01730.

[2] To whom correspondence should be addressed.

0888-7543/92 $5.00

the M13mp19 plaques were hybridized with both the $(CA)_{15}$ and its reverse complement $(GT)_{15}$ oligonucleotides under the same conditions as the multiplex library colony lifts. Replica filters of the M13mp19 library were also hybridized at 50°C to a pool of labeled oligonucleotides containing tetranucleotide repeat sequences ($(AATT)_8$, $(AAAT)_8$, $(AACT)_8$, $(AAGT)_8$, $(AGAT)_8$, and $(ACAT)_8$) and washed in 1× SSC/0.1% SDS at a maximum temperature of 50°C.

DNA sequencing. M13mp19 phage clones were prepared by standard methods, and DNA was purified by phenol extraction (Sambrook *et al.*, 1989) or on Qiagen columns (Qiagen Inc.) and sequenced using the *Taq* dye primer cycle sequencing kit on the ABI 373A DNA sequencing system (Applied Biosystems Inc.). All M13mp19 clones are designated 'MH' (Table 3). The multiplex vector clones were prepared and sequenced by standard multiplex sequencing methods (Church and Kieffer-Higgins, 1988) and are designated MS (Table 3). All clones with any other designation are multiplex vector clones sequenced using a custom-labeled primer with the *Taq* dye primer cycle sequencing kit on the ABI 373A DNA sequencing system or by manual dideoxy sequencing with Sequenase (US Biochemicals Inc.).

Primer selection. The sequence data were analyzed using PRIMER (M. J. Daly, S. Lincoln, and E. S. Lander, unpublished). PRIMER is a computer program for selecting PCR primer pairs to amplify regions of genomic DNA flanking specified target sequences, such as $(CA)_n$ repeats. PRIMER analyzes potential primer sequences on each side of the target, calculates annealing temperatures, determines homology to *Alu* or LINE repetitive elements, determines whether primers have significant complementarity to themselves or each other, and determines the total size and GC content of the PCR product.

Primer pairs were selected using the automatic function of PRIMER, where the optimal oligonucleotide T_m was set at 60°C, and the range for the PCR product size was set from 100 to 250 bp. In some cases alternate primers were selected because of the presence of an *Alu* repeat close to the SSR that prevented the selection of primer pair within the default size range.

The PRIMER program can be obtained directly over the Internet by using anonymous ftp to GENOME.WI.EDU and copying the program from the folder DISTRIBUTION/PRIMER.0.4. The program may also he obtained from Dr. Eric S. Lander at the Whitehead Institute for Biomedical Research (9 Cambridge Center, Cambridge, MA 02142).

PCR typing. PCR was performed using a single [γ-^{32}P]ATP end-labeled primer under conditions described previously (Dracopoli and Meisler, 1990), except that all reactions were carried out in 96-well plate format. The 20-μl PCR reactions contained 100 ng template DNA, 2 pmol of the end-labeled forward primer, 8 pmol unlabeled forward primer, 10 pmol reverse primer, 0.2 m*M* dNTPs, and 0.5 U of *Taq* DNA polymerase. To increase specificity, 0.2 μl Perfect Match polymerase enhancer (Stratagene, La Jolla, CA) was added to the amplifications containing somatic cell hybrid template DNA. PCR reactions were carried out in Perkin-Elmer/Cetus 9600 thermal cyclers using the following cycling conditions: Initial denaturation at 94°C for 5 min, followed by 30 cycles of 94°C for 10 s, 55°C for 30 s, and 72°C for 30 s, and a final extension of 5 min at 72°C. The PCR reactions were then mixed with an equal volume of deionized formamide containing bromophenol blue (0.25%) and xylene cylanol (0.25%), denatured at 100°C for 5 min, rapidly cooled on ice and loaded onto 6% denaturing polyacrylamide gels. The gels were run for 3 h at 120 W, covered with a thin plastic film, and exposed to X-ray film at −70°C for 2–24 h.

Somatic cell hybrid panels. The analysis of the somatic cell hybrid panels was completed in two phases. Preliminary assignments were made by analysis of the 18 samples in the NIGMS human/rodent somatic cell hybrid mapping panel 1 (Taggart *et al.*, 1985; Mohandas *et al.*, 1986). These assignments were independently confirmed by analysis of the appropriate monochromosomal hybrids from the NIGMS human/rodent somatic cell hybrid mapping panel 2.

Panel 1 consists of 15 somatic cell hybrids derived from the fusion of human male fibroblast cells (IMR-91) with mouse B-82 cells. These 15 hybrids, derived from the same human parental cell line, are supplemented with two mouse/human monochromosomal hybrids for chromosomes 16 and X and a single Chinese hamster/human monochromosomal hybrid for chromosome 9. These three monochromosomal hybrids are all derived from a different human parental cell line.

Panel 2 consists of 23 hybrids containing a single human chromosome and one hybrid containing human chromosome 1 and X. The monochromosomal hybrids that contain human chromosomes 9 and 16 are the same as those in the NIGMS panel 1.

Preliminary chromosomal assignments for each SSR were obtained by comparing the distribution of human-specific PCR products with the pattern of human chromosomal distribution in the NIGMS panel 1. The assignments based on the analysis of panel 1 were all independently confirmed by analysis of the appropriate monochromosomal hybrids from NIGMS panel 2. PCR amplification of hybrid cell DNAs resulted in specific amplification of the predicted human band and in some cases the amplification of additional human and rodent bands. Preliminary chromosomal localizations were only determined for those systems that generated a distinct human PCR product with identical allele sizes to the parental IMR-91 cell line in NIGMS panel 1. In most cases, complete concordance was detected between the presence of the human PCR product and a specific chromosome in panel 1. Complex systems with multiple nonspecific bands that often resulted in discordancies were abandoned at this point. However, a few systems with relatively clean amplification of the human bands from the hybrid cell DNA gave minor discrepancies resulting in their tentative assignment to >1 chromosome. These systems were subsequently resolved by the analysis of the appropriate monochromosomal hybrids or they were also abandoned.

Mendelian inheritance of SSR polymorphisms. The Mendelian inheritance of each of the SSRs was determined by analysis of two extended families containing a total of 34 individuals.

RESULTS

$(CA)_n$ Repeats

Sequence data were obtained from 417 short insert genomic clones containing a $(CA)_n$ repeat. These sequences included perfect, imperfect, and compound dinucleotide repeats (Weber, 1990). The repeat element lengths are defined as the longest, uninterrupted sequence of $(CA)_n$ dinucleotides. A total of 265 (64%) sequences contained a $(CA)_n$ repeat, where $n \geq 14$, 115 (28%) sequences contained a $(CA)_n$ repeat, where $n < 14$, and 37 (9%) did not contain a $(CA)_n$ repeat within the region that was sequenced. One additional $(CA)_n$ repeat, D7S466, was derived from an M13mp19 library containing genomic DNA from the hybrid cell line A9/1492-37, which was previously thought to be monochromosomal for chromosome 1p (Dracopoli *et al.*, 1988). The complete nucleotide sequences for all the $(CA)_n$ repeats in Table 3 have been deposited in GenBank and the Accession Nos. are listed in Table 3.

Tetranucleotide Repeats

Sequence data were obtained for eight clones containing tetranucleotide repeats after hybridization of a pool of six oligonucleotides to the low-density filters used for the isolation of $(CA)_n$ repeats from the M13mp19 library. The complete nucleotide sequences for the tetranucleotide repeats in Table 3 have been deposited in GenBank and the Accession Nos. are listed in Table 3.

TABLE 1

Chromosomal Assignment of 124 $(CA)_n$ Repeat Blocks Derived from a Genomic Library of Human Female DNA

Chromosome	Relative length[a]	No. $(CA)_n$ expected ± 1 SD[b]	No. $(CA)_n$	z-score[c]
1	8.0	10.0 ± 3.0	7	−1.0
2	7.6	9.5 ± 3.0	10	0.2
3	6.5	8.1 ± 2.7	10	0.7
4	6.0	7.4 ± 2.6	7	−0.2
5	5.8	7.2 ± 2.6	8	0.3
6	5.6	7.0 ± 2.6	8	−0.4
7	5.1	6.3 ± 2.4	6	−0.1
8	4.7	5.8 ± 2.4	5	−0.3
9	4.6	5.7 ± 2.3	9	1.4
10	4.4	5.4 ± 2.3	4	−0.6
11	4.4	5.4 ± 2.3	8	1.1
12	4.4	5.5 ± 2.3	6	0.2
13	3.6	4.4 ± 2.1	2	−1.1
14	3.4	4.2 ± 2.0	4	−0.1
15	3.3	4.1 ± 2.0	12	4.0
16	3.2	4.0 ± 2.0	2	−1.0
17	3.1	3.8 ± 1.9	4	0.1
18	2.8	3.5 ± 1.8	1	−1.3
19	2.5	3.1 ± 1.8	4	0.5
20	2.4	3.0 ± 1.7	1	−1.2
21	1.8	2.2 ± 1.5	1	−0.8
22	1.9	2.4 ± 1.5	0	−1.6
X	4.9	6.0 ± 2.4	5	−0.4

[a] The relative chromosomal length was obtained from Ott (1985).
[b] Calculated assuming that the frequency of $(CA)_n$ repeats is proportional to chromosome length.
[c] z-score, (observed − expected)/standard deviation.

Primer Selection

The 265 sequences containing an uninterrupted run of ≥14 dinucleotides and the eight tetranucleotide repeats with more than eight uninterrupted repeats were selected for further analysis. Primer pairs were chosen from 166 (63%) of the 265 $(CA)_n$ sequences analyzed by PRIMER. Primer pairs were not chosen from 99 (37%) of the sequences because the $(CA)_n$ repeat was too close to the cloning site in 55 (21%) clones, was flanked by an *Alu* repeat in 22 (8%) clones, or was flanked by a region of low G/C content in 5 (2%) clones or because the sequence was ambiguous or incomplete in 17 (6%) clones.

Five of the eight tetranucleotide repeats were located adjacent to *Alu* repetitive elements. Primers were selected from the three sequences without *Alu* and were only selected from 2/5 sequences containing the *Alu* repeat.

Chromosomal Assignments

A total of 124 (75%) of the 166 $(CA)_n$ repeats and 4 tetranucleotide repeats were mapped to specific chromosomes. Chromosomal assignments were not determined for 42 (25%) of these $(CA)_n$ repeats because they either had very low heterozygosities or because they amplified poorly. At least one $(CA)_n$ repeat was detected on every chromosome except for 22 (Table 1). Since the library was constructed from female genomic DNA, no markers were expected on the Y chromosome. The observed assignment of the $(CA)_n$ repeats was compared to the expected distribution if $(CA)_n$ repeats are randomly distributed in the human genome. The expected frequencies for each chromosome were calculated by assuming that the frequency of $(CA)_n$ repeats is proportional to the relative chromosomal length (Table 1). The observed frequency of $(CA)_n$ repeats agrees with that predicted by the model of random distribution, except on chromosome 15, which has a substantial excess of markers. The data describing the chromosomal assignment of each of the 100 markers in Table 3 have been deposited in the Genome Database (GDB).

Heterozygosity

The 166 $(CA)_n$ repeats and 4 tetranucleotide repeats were tested for heterozygosity against 24 unrelated CEPH parents (12 male and 12 female). The heterozygosity of the 5 $(CA)_n$ repeats assigned to the X chromosome were reanalyzed on a panel of 24 unrelated females from the CEPH reference families. Heterozygosities were determined for 136 (82%) of the $(CA)_n$ repeats and 4 (80%) tetranucleotide repeats. Thirty (18%) of the $(CA)_n$ repeats and 1 (20%) tetranucleotide repeat were abandoned for a variety of reasons, including poor amplification, extreme stuttering, or the amplification of multiple secondary bands. The heterozygosity varied between 0.00 and 1.00 in the screening panel of 24 individuals (Fig. 1). Eighty percent of the $(CA)_n$ repeats had heterozygosities > 0.50, and 49% had heterozygosities > 0.70 (Table 2). The data describing the polymorphism at each of the 100 markers in Table 3 have been deposited in the GDB.

Mendelian Inheritance of SSR Polymorphisms

Mendelian inheritance of all the SSRs listed in Table 3 was tested by coamplifying pairs of markers in the

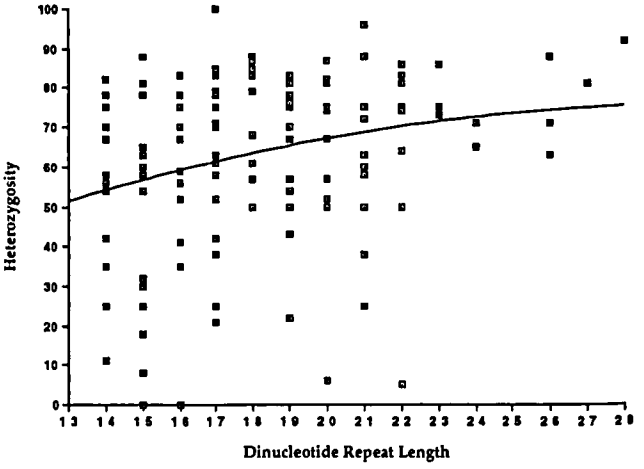


FIG. 1. Heterozygosity at 136 human $(CA)_n$ repeats as a function of the maximum length of the uninterrupted dinucleotide repeat in the sequenced alleles. The relationship between heterozygosity and dinucleotide repeat length was estimated by fitting a second-order polynomial curve to the data using an iterative least-squares method.

TABLE 2

Heterozygosity at 136 Human $(CA)_n$ Repeats

Heterozygosity	Number of SSRs	Cumulative %
1.00–0.90	3	2.2
0.89–0.80	25	20.6
0.79–0.70	39	49.3
0.69–0.60	18	62.5
0.59–0.50	24	80.1
0.49–0.40	6	84.6
0.39–0.30	7	89.7
0.29–0.20	7	94.9
0.19–0.10	2	96.3
0.09–0.00	5	100.0

same 20-μl PCR reaction. Marker pairs were selected for coamplification so that the PCR product sizes differed by >50 bp. In almost all cases, the pairs of SSRs coamplified without problems, although the relative signal strength was often reduced. Mendelian inheritance for the 100 SSRs listed in Table 3 was observed in 2 extended families containing DNA samples from 34 individuals.

DISCUSSION

We describe the development of a cohesive panel of human genetic markers in which every SSR can be analyzed using identical PCR conditions. This panel includes highly informative markers from every chromosome, except 22 and Y. The primer sequences of the 100 SSRs with confirmed chromosomal assignments and with heterozygosities ≥ 0.50 are given in Table 3. Most of these highly informative markers are being typed on the CEPH reference families by several collaborating laboratories.

Analysis of $(CA)_n$ repeats derived from mouse genomic libraries (Love *et al.*, 1990; Dietrich *et al.*, 1992) and RFLPs defined by probes derived from a human genomic library (Donis-Keller *et al.*, 1987) demonstrates that total genomic libraries are a relatively unbiased source of clones. In contrast, the isolation of markers from libraries derived from somatic cell hybrids is often complicated by the nonrandom distribution of clones, and by contamination with other human chromosomal DNA or rodent DNA (Dracopoli *et al.*, 1988; Hazan *et al.*, 1992; Kwiatkowski *et al.*, 1992). $(CA)_n$ repeats have been reported to be relatively uniformly distributed throughout eukaryotic genomes (Hamada *et al.*, 1982; Hamada and Kakunaga, 1982; Stallings *et al.*, 1991), and linkage mapping of human $(CA)_n$ repeats (Decker *et al.*, 1992; Dracopoli *et al.*, 1991; Wilkie *et al.*, 1992) and mouse $(CA)_n$ repeats (Dietrich *et al.*, 1992) derived from total genomic libraries have not identified any evidence of clustering. The frequencies of $(CA)_n$ repeats on each chromosome were, with a single exception, not significantly different from that expected by a model of random distribution (Table 1). However, the detection of 12 $(CA)_n$ repeats on chromosome 15 is significantly greater than that predicted for a chromosome of this size. The reason for the excess of $(CA)_n$ repeats on this chromosome is not apparent, and previous studies have not demonstrated an abnormal number of $(CA)_n$ repeats on chromosome 15. Analysis of the single-copy sequence flanking each $(CA)_n$ repeat on chromosome 15 demonstrates that 11 of the 12 different clones are unique. Therefore, the excess of markers on this chromosome is not due to the biased amplification of a single clone in the multiplex plasmid library.

The heterozygosities observed at $(CA)_n$ repeats have been shown to increase with the length of the dinucleotide repeats (Weber, 1990). Since we were attempting to identify highly informative markers, PCR assays were only developed for the repeats with at least 14 uninterrupted dinucleotides in the sequenced allele. The distribution of heterozygosities for these 136 $(CA)_n$ repeats (Fig. 1) is very similar to that described by Weber (1990). There is a gradual increase in the average heterozygosity with increasing length of dinucleotides, but the range of heterozygosity remains wide in the interval from 14 to 22 uninterrupted dinucleotides, which includes the great majority of $(CA)_n$ repeats identified in this study (Fig. 1). A total of 67 (49%) of these SSRs have heterozygosities > 0.70 (Table 2) and are therefore suitable for inclusion in the "index maps" of human chromosomes.

Weber (1990) has estimated that there are approximately 35,000 $(CA)_n$ repeat blocks with >12 uninterrupted dinucleotides in the human genome. Between 5 and 10% of human genomic DNA is recovered in the 300- to 500-bp fraction after digestion with restriction enzymes with four-base recognition sites that do not contain a CpG sequence. Therefore, the *Hae*III and *Rsa*I fractions should each contain approximately 3500 $(CA)_n$ repeats with >12 uninterrupted dinucleotides. In the absence of extensive cloning biases, it should be possible to isolate many more $(CA)_n$ repeats and other SSRs from these libraries without encountering many duplicate clones. Only 7 duplicated sequences were identified after analysis of the first 417 sequences. Five of these were sequentially numbered clones that presumably resulted from duplicate picking of the same colonies. The sixth duplicated sequence consisted of 2 independently derived clones that mapped on chromosome 15. The seventh duplicated sequence consisted of two independently derived clones that mapped on chromosome 1. Although the sequence flanking the dinucleotide repeat was identical, the two clones contained a $(CA)_{19}$ and $(TG)_{17}$ repeat, demonstrating that they were derived from different alleles in the heterozygous genomic DNA used to construct the library. Surprisingly, this $(CA)_n$ repeat has also been isolated a third time from the A9/1492-37 hybrid library. At this time, it is not possible to systematically compare these 100 highly informative SSRs with those developed in other laboratories because the sequence data for most of the clones are not available in GenBank. However, it is unlikely that many SSRs have been duplicated because of the very large number of $(CA)_n$ repeats in the human genome, and because the

TABLE 3

Description of 100 Highly Informative Human SSRs

Locus	Marker	GenBank Accession No.	Primer sequences (5′–3′)	bp	Repeat	Chromosome	H
D1S159	MIT-MX4	M87710	TCCTTTACATAAATCATTGTCGTG CGACTCTGCATTACCTTGATAGC	147	(CA)19	1	0.67
D1S160	MIT-MS48	M87711	GGTGAAACTAACACTCAACCTGG GCATCTAGCAAACAGCATGTG	150	(CA)19	1	0.72
D1S161	MIT-E112	M87712	CAGGCTTCCAGTTGTCTTCC CTTCCTGGAATCCAGATGGA	159	(CA)17	1	0.84
D1S162	MIT-MS154	M87713	GGGGGAAGAAGTCCGAGTAG ATAAGGGGAACAGGTCTGGG	134	(GT)22	1	0.91
D1S163	MIT-MS217	M87714	TCTTCGTGTGTGGAACCGTA GCGAGAAATGAACTTGGCTC	200	(GT)18	1	0.68
D1S164	MIT-MS165	M87715	TATTTGGGGCAATAAATCAACC CTCAGCTCGTTCATTAAATCCC	229	(GT)20	1	0.83
D2S93	MIT-G105	M87720	ATGGTGTCATGGTGTTTTGTG GCACATTAAAAATTGCAAAATG	146	(GT)16	2	0.83
D2S94	MIT-MS153	M87721	AGCCTTGGGGAAAACTGG AACTGGCACAAAGATGCTCC	150	(GT)17	2	0.75
D2S95	MIT-A119	M87722	GACAGAGCAACACCCCAACT TCATCACTCACCCAGACCAA	146	(GT)17	2	0.85
D2S96	MIT-N118	M87723	TTCCCCTGGTTCTCTCCC GATCTGCTAGAATGAAGAAAACACA	178	(GT)14	2	0.78
D2S97	MIT-MS211	M87724	GAAAAAGCAGAGAAAAAGACCG TCAAGGGAAAAAACAGCGTT	105	(GT)15	2	0.81
D2S98	MIT-MS222	M87725	TGTTCAGGATAAATGTACCCCC GAGCACAGAGGCAGGAAGTC	131	(CA)24	2	0.71
D2S99	MIT-F6	M87726	ACTGCTATTCACAGTTCAGGGA TTTCTGGAAGGTTCTTCAGAGC	192	(TG)23	2	0.73
D2S100	MIT-MH105	M87719	AGGCTCTTGCCATTCTGAAA GTATGTCAACCATCCTCTTCCA	143	(CA)16	2	0.52
D3S1209	MIT-MS24	M87727	GCTCTTCCTCTCCCTGCC TACAAGGGGTGGGAGGTACA	156	(CA)22	3	0.75
D3S1210	MIT-MS140	M87728	GGGCTATTTTGCAACTTACTCG ATCCTGATGGCAATATGAAATG	157	(CA)17	3	0.71
D3S1211	MIT-I106	M87729	CAGGGCTTGTGGGATTAGAA ATTTCAGATTTCAGGACAAGGG	181	(CA)15	3	0.88
D3S1212	MIT-E109	M87730	GGTACTTTCCACCTAGTCAAAACA TGTAGGGTTTGCAATGTCCA	193	(CA)16	3	0.75
D3S1214	MIT-E144	M87731	TCTCCACTTTTCCACCCTA TTCGGTCAGGAGCTGCTG	151	(CA)15	3	0.60
D3S1215	MIT-MS207	M87732	ATATTTCAGCGTGTGAGATACACA CATCTCACTCTGGAAGAGAAAATG	101	(CA)16	3	0.78
D3S1216	MIT-K117	M87733	CTACTGAGGGATGTTGATGGC TTGTTTAAGCCATTCAGTCTATGG	170	(GT)23	3	0.86
D3S1217	MIT-F8	M87734	TGACAAGTTTAAAGGGTCCCA TGTCAAAGTCCCCTTCCTTG	190	(GT)17	3	0.83
D3S1227	MIT-MS238	M87735	TAACAGGAGGAATTTTTCTTCTGG GCAAACTGGGTCCTACCCTT	149	(GT)17	3	0.61
D4S243	MIT-MH34	M87736	TCAGTCTCTCTTTCTCCTTGCA TAGGAGCCTGTGGTCCTGTT	173	(AGAT)10	4	0.67
D4S244	MIT-MS176	M87737	CGTTTAAGGCCACTTTGCTT AAAATTGCAAGAAGGCTAATGG	148	(CA)18	4	0.83
D4S245	MIT-N133	M87738	TGCAAGTAAACAGTGACCAAAG TTTGGATATTTGCATTCAAAAA	107	(CA)16	4	0.52
D4S246	MIT-MS205	M87739	TGAATATCCCAGCTTTAGAAAAGA CCAGCTGTCACTGAGTCAGTT	163	(GT)14	4	0.67
D4S247	MIT-MS240	M87740	AATGAGTGGGAAGGTTGCAG TATACCTATTTCCAGGCATAAGCA	173	(GT)14	4	0.82
D4S250	MIT-N136	M87686	TGGACTTGAACTAGTTCTCCAGC AGGTTCTCCAGAGAAACAAAACC	215	(CA)17	4	0.83
D5S349	MIT-A127	M87741	ATTTGGTTTCCATAGAATCTGAGA TTACACCCACCAGATTAAGCG	140	(CA)27	5	0.81
D5S350	MIT-MS131	M87742	CTCACTCACTTCTCTCTCTGCG TTCAAGCGCGAGAAGAATTT	136	(GT)18	5	0.61
D5S351	MIT-I105	M87743	ACCAGTCTATGGCAACACAGC GATGAGCATTGCCACTTTAGC	197	(CA)17	5	0.75
D5S352	MIT-MS158	M87744	CCACCGCAGCCAGCTAAT GAGGTGGGTAGATTACTTGAGTCC	149	(GT)21	5	0.96
D5S353	MIT-MH98	M87745	ATACACTGGAAATCCACATTGTG ATCCCACACACAGTGCAGAA	133	(GT)18	5	0.83
D5S354	MIT-MH96	M87746	CCGAATTGGTCTATAGGTACGC TCTCATATTGAAGCACAGAAAAAA	135	(GT)19	5	0.76
D5S355	MIT-MH91	M87747	GATGTCTTTCCATTGTCTTCTGG ATAGAAAACCCAGCAAGATAAACA	194	(GT)14	5	0.67

TABLE 3—*Continued*

Locus	Marker	GenBank Accession No.	Primer sequences (5′–3′)	bp	Repeat	Chromosome	H
D6S220	MIT-G119	M87748	ACCAGGCACCCAAAACTGT TTGGGATTTTGGCTGAGATC	175	(GT)18	6	0.68
D6S224	MIT-MS135	M87749	AAAAACTAGCCCAGCTCGGT AGAAGTTGTCTGCCCATTTCA	245	(GT)16	6	0.59
D6S225	MIT-E116	M87750	ACAGTTTCTGGATCAGCTATTTCC AGAAGAGCTGATAATGTTGAGTGC	132	(CA)18	6	0.79
D6S226	MIT-MS236	M87751	CCAGTCACCCTGCTAACCAT CCGCAGCAAATTTATGTCAG	206	(CA)14	6	0.70
D7S460	MIT-MH26	M87752	AATACCCCAAGGGGTGGTAA CATTGATGAACAGTTCAAGCA	188	(AACT)8	7	0.95
D7S461	MIT-MS97	M87753	GGGAAACTCAAAGGCAGAGA AAACTCAGCATTGTCCTGCC	177	(CA)20	7	0.87
D7S462	MIT-MS262	M87754	CCACAAGCCTGGATTTGG GTCTTCTACTTGGTTTGCACAGG	150	(GT)17	7	0.52
D7S463	MIT-G111	M87755	AGTTGGAGCCCAGCTTGC GCAGTCTTCCCTCCTCAGC	159	(CA)19	7	0.70
D7S466	MIT-COS43	M87756	TGGGCCACCATAGTATAGCC TGACCACACGTCCACTGG	244	(GT)23	7	0.83
D8S205	MIT-MS45	M87757	ATTCTCTCGTCCTTTCTTGGG ACAAATTTCCGTGTAAGGTAGAGG	108	(GT)19	8	0.78
D8S206	MIT-MS61	M87758	GAAAACCATGGCTGGGTG ACATGCATTAGCACTACCATGC	127	(GT)16	8	0.67
D8S207	MIT-MS142	M87759	TCTCCCATAGGGTCCTGTCA ACACATACCATTGGAAACTCCC	142	(GT)20	8	0.74
D8S208	MIT-MS91	M87760	AGCCTCATCACTCCCCTCTT ACAGCAGCGATAAGGAATTAGC	155	(CA)20	8	0.75
D9S129	MIT-MS47	M87761	TTCAGAAATTCTTTCTTCTGCTTG TAATGAAACAAATCTTGAAACCAA	135	(GT)16	9	0.67
D9S130	MIT-E117	M87762	AACTGGAGGGGGAAAATACG TTGTGTAGGCAAACTTGAGTGC	184	(GT)15	9	0.58
D9S131	MIT-MS202	M87763	CCAGCGTGGCATGTCTCT TCATCTCATAACCCTCTGACCC	100	(GT)17	9	0.83
D9S132	MIT-G115	M87764	CTCGGGACTTTTCAGCCTC TGAAAGTTGAATGATGTTCACG	156	(GT)14	9	0.75
D9S133	MIT-MS67	M87765	ATCGCAAACTGCTTCCTTTT TCACAAGGCCCATTACAGTG	150	(GT)15	9	0.63
D9S135	MIT-MS93	M90976	TTTCAGGTCTCTTTTCTGTAGGG CCAACAGAAAAGATACCAGAAATC	99	(GT)21	9	0.63
D10S172	MIT-MX5	M87670	TATTTTGATTTTGCATGTGTGAA TTTGCTGAGAGAGTAGATTTCAGG	151	(GT)22	10	0.64
D10S173	MIT-MX6	M87671	GCTGATTTTTCCTGCTGGTC TGTTTCTGAAGCATTTTCCTTG	155	(CA)19	10	0.81
D10S174	MIT-MS88	M87672	TTCTGTGAGCATCTGTACAGCT TATGAGCCATCCCTAGCCAG	150	(CA)20	10	0.75
D10S175	MIT-MS77	M87673	CCATAGCCATTCTTCCTCCA GTGGTGACAACTCTGAGCCA	233	(CA)17	10	0.70
D11S861	MIT-A136	M87674	CTGAAACCAAGTGAAAAGGAGA AAGCTCCATTGTCTTCTGGC	154	(CA)16	11	0.70
D11S862	MIT-MS7	M87675	TACCATATTAAATCACCCACATGG GGAATCAGGATTCAAACTCTGG	152	(CA)16	11	0.83
D11S863	MIT-MS20	M87676	GCAACATGGTAAGAGTCCAGC GTAGCTGGGACCAGAGGGAT	133	(CA)15	11	0.65
D11S865	MIT-E137	M87677	CTTTTTGTTGCCCATTGCTT GTAAAGGATAAAAACTCTGCAGGC	170	(CA)22	11	0.81
D11S866	MIT-A106	M87678	TCTGCTCTTCCAGGTGCC TGCAGGCTGGTCTTCAGTC	178	(CA)17	11	0.78
D12S68	MIT-G117	M87679	TCCTGGAACCTTGCTTGC CCACCTCCAAGTCATCCAGT	201	(GT)21	12	0.50
D12S69	MIT-MS6	M87680	GGATCATAATGTAGGTTCACCTCC GCAAGACCCCCATCTCTACA	214	(GT)16	12	0.56
D12S70	MIT-MS54	M87681	TGTAGAAAAGAGAATGATGATGCC TTACTATCTACCCCCTGACTGACA	177	(GT)21	12	0.72
D12S71	MIT-MS159	M87682	CTGGTACCCCTTGTCAGCAT TTAAAGCTGGGGAGAGGGAT	179	(GT)15	12	0.78
D12S72	MIT-MS263	M87683	CATCATCCCATGGTCGAAG GAGAGTAGGTTCCTTATCCTGGG	159	(GT)22	12	0.83
D12S74	MIT-A118	M87684	CAACTCTTCCCATTTTTGAAGG GGGCTTGGAAAACCAGCT	175	(CA)19	12	0.54
D13S115	MIT-MS34	M87685	TGTAAGGAGAGAGAGATTTCGACA TCTTAGCTGCTGGTGGTGG	169	(CA)19	13	0.82
D14S57	MIT-MS16	M87687	TGATACCTATGCAAGTATGTTTGC GTGGATGAGAGTTTGATTGTGTG	141	(GT)17	14	0.63
D14S58	MIT-MS162	M87688	ATTTCTCTCATTAAATCTGCTCCG TCAGCCCAGGAGGTTGAG	205	(GT)26	14	0.63

TABLE 3—*Continued*

Locus	Marker	GenBank Accession No.	Primer sequences (5'–3')	bp	Repeat	Chromosome	H
D14S59	MIT-MH90	M87689	CTTTTGCTTTCCCAGGAGG GTGCTGATGCAAGAAGGTGA	107	(GT)18	14	0.88
D15S97	MIT-MS14	M87695	CTTCTAGCCTCAGGTTCCCC ATTCACTTTTCAAACCACCCC	159	(GT)22	15	0.75
D15S98	MIT-MS112	M87696	CATGTGAAACTGCAAAAGCTG AAAAGTCGCATTTGGTCGTT	152	(CA)20	15	0.81
D15S99	MIT-MS149	M87697	GGGTGAGAAGTTTGTCCTTCC ATGCAGAGCCCCTGTGAG	135	(GT)19	15	0.50
D15S100	MIT-MS164	M87690	CTTTCCAATTCACCCCCAC ATCCAGCTCCCCCAAATATT	183	(GT)17	15	0.79
D15S101	MIT-MS178	M87691	GCTGCATTCCAGCCTAAAAG GCCCACTAGTTTGAGACAGAGA	104	(CA)19	15	0.83
D15S102	MIT-N130	M87692	TCAATAACTCCATTGCTCAGTCC TAGGGCCAATGGAGAGAGC	217	(CA)18	15	0.85
D15S103	MIT-G113	M87693	GTTCAAAAGGCTCATCATGTAGC TTCCTGAGTCCTTATAGCTTCCA	233	(CA)17	15	0.58
D15S104	MIT-M131	M87694	ATGCAGAGCCCCTGTGAG CCACTCTGTAGGGTTTCACTCC	170	(CA)18	15	0.83
D16S310	MIT-MH20	M87698	GGGCAACAAGGAGAGACTCT AAAAAAGGACCTGCCTTTATCC	162	(ATAG)12	16	0.67
D16S312	MIT-I103	M87699	CTAGGGGGACACACGCAC CTGTCCACCCTGTGACCC	176	(CA)23	16	0.75
D16S313	MIT-MS79	M87700	TGACTCTTTTCGTCAACACAGC TATTCAAATACAAGCAACCGTG	141	(CA)20	16	0.57
D17S581	MIT-MS52	M87701	CAAGAGTGGAAATTGACCTCG TAAGATTTTCTCTCAGAGTGCACC	155	(CA)16	17	0.75
D17S582	MIT-MS105	M87702	GCTGTTTCAAAAGGCTTTGG GGAGGGACTCAGTCAGAAAGG	123	(GT)15	17	0.63
D17S583	MIT-N127	M87703	TTCCCTTAAGCCCAGGAGTT CCACGTCAAATACTTGGGCT	178	(CA)14	17	0.54
D17S584	MIT-MS246	M87704	CATATCAATGCACTGATAAAACCC GGTGGGGAGATGGTTTCTCT	139	(GT)20	17	0.52
D18S36	MIT-MS156	M87705	TTCAGTTTTTCACATGCATAAAA TCTTCCATTGATCCCAAAATG	146	(GT)21	18	0.75
D19S197	MIT-G116	M87706	TCCTCAGCAGGTCTCCTCAT CACCAGGGAAATGCCAAG	153	(GT)28	19	0.92
D19S198	MIT-M133	M87707	GAAAGTGTCCACAACGGTAGC AAAGGGGTTAGAGGAGAAGGC	115	(CA)18	19	0.87
D19S199	MIT-MS231	M87708	TGGATTTGTGCTTTATTTTACATT GAAAAAGTTGAATCAACCCTGG	150	(CA)17	19	0.83
D19S200	MIT-MS265	M87709	GGAGTTGCTTCGCTTGTAGC TAAAAAGAAGATAGCCATGTGAGG	138	(GT)26	19	0.88
D20S75	MIT-M127	M87717	GCTGAAATGGGAGGATCG GCTGCAGTGAGACATGATCA	173	(CA)17	20	1.00
D21S222	MIT-G121	M87718	CACTTTCTGCAGAGGAAAAGC ATCTTCCCAATTTTACAATGATCT	131	(CA)21	21	0.88
DXS730	MIT-MS21	M87766	CCTGATCATCCCTGAGGCTA TTAGTCCCAGGGGGGAGAG	192	(GT)21	X	0.63
DXS731	MIT-MS266	M87767	CTCACCATTGGGTCTTCATACA TATGATAGGCATGAATTGTGTCTG	100	(CA)22	X	0.74
DXS737	MIT-MS120	M87768	GAGGAAAAAATCAAGTCCATGT TGTCTAGTGTATGACTCATCTCCA	167	(CA)21	X	0.60
DXS738	MIT-E114	M87769	CCAGCAATAACCATAAGTAAAACC AATGTGTTGTTGTATTCACCTTGC	144	(CA)17	X	0.71

Note. The chromosomal assignment for each locus was determined by somatic cell hybrid analysis. Heterozygosities were determined after analysis of 24 unrelated genomic DNA samples. The length of the PCR product of the sequenced allele is defined in base pairs, and the number and type of repeated elements on the sequenced strand is defined. The primer sequences all have a T_m between 58 and 62°C and may be amplified using identical PCR conditions described in the text. The complete sequence data for each locus have been deposited in GenBank and may be accessed using the GenBank Accession No.

libraries used by Weber (1990) to isolate many $(CA)_n$ repeats were constructed from different restriction digests. Similar analyses of 303 $(CA)_n$ repeats from an *Mbo*I-digested and size-fractionated mouse genomic library identified only 6 duplicates (Dietrich *et al.*, 1992). Although there are approximately 50% more $(CA)_n$ repeats in the mouse genome (Stallings *et al.*, 1991), these data suggest that the independent isolation of identical human $(CA)_n$ repeats in different laboratories will be a rare event that can be quickly resolved as the repeats are added to the human genetic map.

ACKNOWLEDGMENTS

We thank George Church for help with multiplex sequencing. This work was supported by National Institutes of Health (NIH) Grant

RO1-HG00395 (to N.C.D.) and NIH Center for Genome Research Grant P50-HG00098. Thomas Hudson is a recipient of a Clinician-Scientist Award from the Medical Research Council of Canada.

REFERENCES

Botstein, D., White, R. L., Skolnick, M. H., and Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction length polymorphisms. *Am. J. Hum. Genet.* **32:** 314–331.

Church, G. M., and Gilbert, W. (1984). Genomic sequencing. *Proc. Natl. Acad. Sci. USA* **81:** 1991–1995.

Church, G. M., and Kieffer-Higgins, S. (1988). Multiplex DNA sequencing. *Science* **240:** 185–188.

Decker, R. A., Moore, J., Ponder, B., and Weber, J. L. (1992). Linkage mapping of human chromosome 10 microsatellite polymorphisms. *Genomics* **12:** 604–606.

Dietrich, W., Katz, H., Lincoln, S. E., Shin, H.-S., Friedman, J., Dracopoli, N. C., and Lander, E. S. (1992). A genetic map of the mouse suitable for typing intraspecific crosses. *Genetics,* in press.

Donis-Keller, H., Green, P., Helms, C., Cartinhour, S., Weiffenbach, B., Stephens, K., Keith, T. P., Bowden, D. W., Smith, D. R., Lander, E. S., Botstein, D., Akots, G., Rediker, K. S., Gravius, T., Brown, V. A., Rising, M. B., Parker, C., Powers, J. A., Watt, D. E., Kauffman, E. R., Bricker, A., Phipps, P., Muller-Kahle, H., Fulton, T. R., Ng, S., Schumm, J. W., Braman, J. C., Knowlton, R. G., Barker, D. F., Crooks, S. M., Lincoln, S. E., Daly, M. J., and Abrahamson, J. (1987). A genetic linkage map of the human genome. *Cell* **51:** 319–337.

Dracopoli, N. C., and Meisler, M. H. (1990). Mapping the human amylase gene cluster on the proximal short arm of chromosome 1 using a highly informative $(CA)_n$ repeat. *Genomics* **7:** 97–102.

Dracopoli, N. C., O'Connell, P., Elsner, T. I., Lalouel, J.-M., White, R. L., Buetow, K. H., Nishimura, D. Y., Murray, J. C., Helms, C., Mishra, S. K., Donis-Keller, H., Hall, J. M., Lee, M. K., King, M.-C., Attwood, J., Morton, N. E., Robson, E. B., Mahtani, M., Willard, H. F., Royle, N. J., Patel, I., Jeffreys, A. J., Verga, V., Jenkins, T., Weber, J. L., Mitchell, A. L., and Bale, A. E. (1991). The CEPH consortium linkage map of human chromosome 1. *Genomics* **9:** 686–700.

Dracopoli, N. C., Stanger, B. Z., Ito, C. Y., Call, K. M., Lincoln, S. E., Lander, E. S., and Housman, D. E. (1988). A genetic linkage map of 27 loci from PND to FY on the short arm of human chromosome 1. *Am. J. Hum. Genet.* **43:** 462–470.

Giblett, E. R. (1969). "Genetic Markers in Human Blood," Blackwell, Oxford, UK.

Hamada, H., and Kakunaga, T. (1982). Potential Z-DNA forming sequences are highly dispersed in the human genome. *Nature* **298:** 396–398.

Hamada, H., Petrino, M. G., and Takunaga, T. (1982). A novel repeated element with Z-DNA forming potential is widely found in evolutionary diverse eukaryotic genomes. *Proc. Natl. Acad. Sci. USA* **79:** 6465–6469.

Hazan, J., Dubay, C., Pankowiak, M.-P., Becuwe, N., and Weissenbach, J. (1992). A genetic linkage map of human chromosome 20 composed entirely of microsatellite markers. *Genomics* **12:** 183–189.

Jeffreys, A. J., Wilson, V., and Thein, S. L. (1985). Hypervariable 'minisatellite" regions in human DNA. *Nature (London)* **314:** 67–73.

Kwiatkowski, D. J., Henske, E. P., Weimer, K., Ozelius, L., Gusella, J. F., and Haines, J. (1992). Construction of a GT polymorphism map of human 9q. *Genomics* **12:** 229–240.

Love, J. M., Knight, A. M., McAleer, M. A., and Todd, J. A. (1990). Towards construction of a high resolution map of the mouse genome using PCR analyzed microsatellites. *Nucleic Acids Res.* **18:** 4123–4130.

Mohandas, T., Heinzmann, C., Sparkes, R. S., Wasmuth, J., Edwards, P., and Lusis, A. J. (1986). Assignment of human 3-hydroxyl-3-methylglutaryl coenzyme A reductase gene to q13–q23 of chromosome 5. *Somat. Cell Mol. Genet.* **12:** 89–94.

Mohr, J. (1954). "A Study of Linkage in Man," Munksgaard, Copenhagen.

Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E., and White, R. (1987). Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science* **235:** 1616–1622.

Ott, J. (1985). "Analysis of Human Genetic Linkage," The Johns Hopkins Univ. Press, Baltimore.

Royle, N. J., Clarkson, R. E., Wong, Z., and Jeffreys, A. J. (1988). Clustering of hypervariable minisatellites in the proterminal regions of human autosomes. *Genomics* **3:** 352–360.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). "Molecular Cloning: A Laboratory Manual," Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Stallings, R. L., Ford, A. F., Nelson, D., Torney, D. C., Hildebrand, C. E., and Moyziz, R. K. (1991). Evolution and distribution of $(GT)_n$ repetitive sequences in mammalian genomes. *Genomics* **10:** 807–815.

Taggart, R. T., Mohandas, T. K., Shows, T. B., and Bell, G. I. (1985). Variable number of pepsinogen genes are located in the centromeric region of human chromosome 11 and determine the high frequency of electrophoretic polymorphisms. *Proc. Natl. Acad. Sci. USA* **82:** 6240–6244.

Weber, J. L. (1990). Informativeness of human $(dC\text{-}dA)_n \cdot (dG\text{-}dT)_n$ polymorphisms. *Genomics* **7:** 524–530.

Weber, J. L., and May, P. E. (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* **44:** 388–396.

White, R. L., Lalouel, J.-M., Nakamura, Y., Donis-Keller, H., Green, P., Bowden, D. W., Mathew, C. G. P., Easton, D. F., Robson, E. B., Morton, N. E., Gusella, J. F., Haines, J. L., Retief, A. E., Kidd, K. K., Murray, J. C., Lathrop, G. M., and Cann, H. M. (1990). The CEPH consortium primary linkage map of human chromosome 10. *Genomics* **6:** 393–412.

Wilkie, P. J., Krizman, D. B., and Weber, J. L. (1992). Linkage map of human chromosome 9 microsatellite polymorphisms. *Genomics* **12:** 607–609.

CGR Newsletter

Whitehead Institute/MIT Center for Genome Research

July 1994 Issue 1

The Whitehead Institute/MIT Center for Genome Research (CGR) opened in July, 1993 at One Kendall Square in Cambridge, supported by a grant from the National Center for Human Genome Research. The major goals of the Center are: (1) to construct genetic and physical maps of the mouse genome; (2) to create a physical map of the human genome; and (3) to ensure that these resources are distributed to the scientific community in a timely and convenient fashion. These maps and others like them provide crucial infrastructure for the study of mammalian genetics and should be valuable in the identification of disease genes.

In order to accomplish its goals, CGR is organized in project teams and cores. At present, they are: Mouse, Human, Informatics, Sequencing and YAC.

This is the first issue of our Center newsletter, which is planned to appear quarterly. Its goal is to provide up-to-date information about progress and to describe resources available from the Center. We solicit your suggestions about ways to improve the utility of this newsletter in subsequent editions.

MOUSE GENETIC MAP REACHES 4537 SSLPs

The Mouse Genome Mapping Project aims at building genetic and physical maps covering the entire mouse genome. Such maps should make it rapid and simple for any mouse geneticist to map monogenic or polygenic traits and to obtain cloned DNA spanning the region of interest. In this way, tedious mapping and walking efforts should become unnecessary.

The first step is the construction of a high density genetic map consisting of 6,000 simple sequence length polymorphisms (SSLPs). These markers are easily typed by PCR and have a high polymorphism rate in both interspecies and intraspecies crosses. The vast majority of these markers are anonymous CA_n-repeat loci, isolated and sequenced from a whole genomic library.

(continued on page 2)

HUMAN PHYSICAL MAP TOPS 3419 STSs

The Human Genome Mapping Project aims at constructing a physical map of the human genome by the strategy of STS content mapping. The goal is to screen a total of 10,000 sequence tagged sites (STSs) by mid-1996 on 25,000 YACs from the CEPH mega-YAC library (average size 1 Mb), in order to identify the YACs containing each STS. In addition, the STSs will be screened on a panel of 'whole-genome' radiation hybrids to provide an important measure of top-down order. The STSs will consist of about 5,000 genetically mapped simple sequence length polymorphisms (SSLPs) from other groups (principally, Jean Weissenbach's group at Genethon and the CHLC consortium) and 5,000 completely random STSs developed at CGR.

(continued on page 2)

MOUSE GENETIC MAP

(continued from page 1)

The genetic markers are all genotyped in a single (OB x CAST) F_2 intercross with 46 progeny, for a total of 92 informative meioses. With this cross, genetic markers can thus be placed in "bins" with an average size of 1.1 centiMorgans (cM). The genetic markers are also tested in twelve inbred mouse strains, to determine their allele sizes.

As of the July 1994 data release, the genetic map contains 4537 SSLPs—with an average spacing of 0.30 cM or 660 kb. The SSLP map is also being closely integrated into the mouse gene map. Approximately one-quarter of the SSLP markers will be genotyped in DNAs from a (B6 x SPRET) x B6 backcross of N. Copeland and N. Jenkins, which has been scored for RFLPs in more than 1000 genes. To date, some 250 markers have been typed in this cross. In addition, CGR entered into a collaboration with the European Collaborative Interspecific Backcross (EuCIB) to map the SSLPs at higher resolution, by typing them in a 1,000-progeny backcross.

Once the genetic map is complete, the project will turn to physical mapping. The initial goal is to construct a physical map by STS-content mapping of mouse YAC libraries. A total of 10,000 STSs will be used, consisting of the 6,000 SSLPs and some 4,000 random STSs. This collection will provide anchor points with an average spacing of 300 kb. Using these STSs, we plan to screen YAC libraries with an average size of about 700kb. CGR has already constructed a 4.3-fold coverage mouse YAC library, most of which has an average insert size of about 700 kb. Additional YAC libraries are under construction. Once the initial STS content mapping is completed, we hope to close as many gaps as possible in a directed fashion.

HUMAN PHYSICAL MAP

(continued from page 1)

Over the past year, we have been developing methods to scale up physical mapping, which includes an automated PCR setup process which generates 18,000 PCR reactions per day. The detection of these PCR products is done by a chemiluminescent hybridization assay, and data capture using CCD cameras.

YAC screening is currently carried out by a two-level scheme. At the top level, we screen superpools from eight microtiter plates. From each positive superpools, we then screen 28 subpools corresponding to the rows, columns and plates of a block. These coordinates should provide a unique address provided that the block contains only a single YAC containing the STS; the address can be ambiguous (i.e., have more than one row, column or plate) if the block has two or more YACs containing the STS.

To date, we have identified YACs for 3419 STSs which fall into the following categories: (i) 1427 Genetically mapped polymorphic STSs, which allow contigs to be anchored to the genetic map; (ii) 838 Random genome-wide STSs, of which approximately 75% are unambiguously assigned to a chromosome using the NIGMS Human/Rodent Somatic Cell Hybrid Mapping Panel #1; (iii) 857 unpublished CA-repeat-containing STSs generously provided by J. Weissenbach. These CA-repeats, which were not sufficiently polymorphic to be genotyped for the Genethon human genetic map, are an additional source of random STSs; (iv) 114 new chromosome 22 STSs generated from sequences derived from flow-sorted chromosome libraries, in collaboration with the Human Genome Center for Chromosome 22 in Philadelphia. These STSs and the corresponding YACs already provide considerable coverage of the long arm of chromosome 22; and (v) 253 STSs from public data bases.

With the current data, the average spacing between STSs is about 875 kb. This is still not sufficient to allow the assembly of a comprehensive and reliable STS content map. As additional STSs are added, contigs covering substantial portions of the genome are expected to fall together.

CGR Resources

CGR strives to ensure rapid and convenient access by the scientific community to information, reagents, and software tools developed by our mapping projects.

Databases

Human physical mapping and mouse genetic mapping data are released on a quarterly basis in January, April, July, and October. The data is available in a number of ways.

Via ftp - You will need access to an *ftp* program such as *Fetch* on the Macintosh. Set your program to log into **genome.wi.mit.edu**. Use "anonymous" for the user name, and use your e-mail address for the password. Data files are stored in **/distribution/mouse_sslp_releases** and **/distribution/human_STS_releases.**

Via e-mail - Send e-mail to the address **genome_database@genome.wi.mit.edu**, with the word "help" appearing as the first word on the subject line or body text. You will receive instructions for accessing the data by return mail. As of spring 1994, only the mouse genetic mapping information is available via this route.

Via World Wide Web - You will need a World Wide Web client, such as *Mosaic* (widely available for multiple platforms). Tell your client to connect to **http://www-genome.wi.mit.edu**, and follow the links to the data directories.

For further help with database services, call Lincoln Stein, Assistant Director of Informatics, 617-252-1916, lstein@genome.wi.mit.edu.

Software

Software is available via World Wide Web and ftp. All software is stored in the directory **/distribution/software.**

The programs currently available are as follows:

MAPMAKER-an interactive computer package for construction of genetic maps in experimental crosses and human reference families.

MAPMAKER/QTL-an interactive computer package for genetic mapping of quantitative trait loci (QTLs) in experimental crosses, using LOD scores.

PRIMER-a computer program for selection of PCR primers satisfying specified conditions.

EXCEL TCP/IP PLUGIN- a Microsoft Excel add-on that allows regions of the spreadsheet to be sent to and received from UNIX hosts on the Internet.

Distribution of Biological Reagents

CGR tries to promote broad and immediate access to biological reagents, by encouraging and assisting distribution services by the private sector.

• To ensure access to mouse and human STSs, CGR pioneered in 1990 an arrangement under which Research Genetics, Inc. retains a portion of all PCR primer pairs synthesized for our use and sells aliquots to the scientific community at discount prices, under the name "MapPairs." The arrangement has since been extended to include PCR primers from other genome centers, as well.

• To ensure access to mouse and human YACs, CGR distributed copies of its mouse YAC library (4.3-fold coverage in YACs of about 700kb; Kusumi et al. Mammalian Genome, 4:391-2 (1993)) and the CEPH mega-YAC library (plates 613-984) to both Research Genetics, Inc. and Genome Systems, Inc. These companies provide library services, including: screening service for individual STSs; purchase of YAC DNA pools for PCR screening of STSs; and purchase of a copy of the entire library. For more information contact:

Research Genetics, Inc.
2130 Memorial Parkway, SW
Huntsville AL 35801
Phone: (800)-533-4363
Fax: (205) 536-9016

Genome Systems, Inc.
7166 Manchester Road
St. Louis MO 63143
Phone: (800)-248-7609
FAX: (314)-647-4134

CGR receives no royalties from distribution of reagents by private companies.

Policy on Data Release and Patents

To promote the broadest possible use of the biological tools developed under the Human Genome Project, Whitehead Institute has adopted a policy that:

• Genetic and physical maps (including all clones, genetic markers, primers and sequences)

CGR Resources Continued

will be made promptly available to the scientific community, will be placed in the public domain, and will not be patented.

• No access to maps will be granted to any commercial entity in advance of public release.

Data releases are scheduled every 90 days. At the end of each calendar quarter, all genomic mapping data are reviewed and prepared for distribution via CGR's electronic databases. Data releases typically occur within two weeks of the close of the quarter (i.e., in mid-January, mid-April, mid-July and mid-October). Releases are announced by electronic messages posted to the following two newsgroups: "bionet.genome.chromosomes" and "bionet.announce".

The purpose of CGR's data release policy is to ensure that scientific colleagues have immediate access to information that may assist them in the search for genes. Data releases do not constitute scientific publication of CGR's work, but rather provide scientists with a regular look into our lab notebooks. For projects aimed at the analysis of particular genes or small subchromosomal regions, permission is hereby granted to use our data without the need for a formal collaboration, subject only to appropriate acknowledgement. For projects aimed at large-scale mapping of entire chromosomes or entire genomes, use of the data and markers should be on a collaborative basis. The information for the mouse genetic map should be cited as:

Whitehead Institute/MIT Center for Genome Research, Genetic Map of the Mouse, Data Release 7 (July 1994)

Dietrich, W.F. *et al.* A genetic map of the mouse with 4,006 simple sequence length polymorphisms. *Nature Genetics* 7:220-245 (1994).

The information for the human genome mapping project should be cited as:

Whitehead Institute/MIT Center for Genome Research, Human Genetic Mapping Project, Data Release 2 (July 1994).

If you would like to receive future issues of the Whitehead Institute/MIT Center for Genome Research Newsletter in hard copy, please complete and send in the following form:

*Issues are also available on the World Wide Web. Our URL is http://www-genome.wi.mit.edu/

_____

Name

_____

Institution

_____

Department

_____

Street/Building

_____

City, State, Zip (Country)

Return to: Newsletter Editor, Whitehead Institute/MIT, Center for Genome Research, One Kendall Square, Building 300, Cambridge, MA 02139-1561 or e-mail: newsletter @genome.mit.edu

References

The following references may provide useful background on CGR projects and related topics.

Mouse Mapping

Copeland, N.G. et al. A genetic linkage map of the mouse: Current applications and future prospects. Science 262: 57-66 (1993).

Copeland, N.G. et al. Genome maps IV: The mouse. Science 262: 67-82 (1993).

Dietrich, W.F. et al. A genetic map of the mouse suitable for typing intraspecific crosses. Genetics 131: 423-447 (1992).

Dietrich, W.F. et al. A genetic map of the mouse with 4,006 simple sequence length polymorphisms. Nature Genetics 7:220-245 (1994).

Dietrich, W.F. et al. in Genetic Maps 1992 (ed. O'Brien, S.) 4.100-4.142 (Cold Spring Harbor Laboratory Press, New York, 1992.)

Kusumi, K. et al. Construction of a large-insert yeast artificial chromosome (YAC) library of the mouse genome. Mammalian Genome 4:391-392 (1993).

Miller, J.C. et al. SSLP/Microsatellite genetic linkage map of the mouse. In M. F. Lyon and A.G. Searle (eds.) Genetic Variants and Strains of the Laboratory Mouse, Third Edition, Oxford University Press, Oxford (1994).

Human Mapping

Foote, S. et al. The human Y chromosome: Overlapping DNA clones spanning the euchromatic region. Science 258: 60-66 (1992).

Hudson, T.J. et al. Isolation and chromosomal assignment of 100 highly-informative human simple sequence repeat polymorphisms. Genomics 13:622-629 (1992).

Vollrath, D. et al. The human Y chromosome: A 43-interval map based on naturally occurring deletions. Science 258: 52-59 (1992).

Informatics

Goodman, N. et al. Building a laboratory information system around a C++-based object-oriented DBMS. In Proceedings of the 20th International Conference on Very Large Data Bases. Santiago (1994).

Goodman, N. et al. Requirements for a deductive query language in the MapBase genome-mapping database. In Proceedings of the Workshop on Programming with Logic Databases: 18-32 Vancouver (Oct. 1993).

Goodman, N. An object-oriented DBMS war story: Developing a genome mapping database in C++. In Modern Database Management: Object-Oriented and Multidatabase Technologies, Won Kim, ed. ACM Press, NY (1994).

Rozen, S. et al. Constructing a domain-specific using a persistent object system. Sixth International Workshop on Persistent Object Systems. Tarascon (1994).

Stein, L. et al. Splicing UNIX into a genome mapping laboratory. In USENIX Summer 1994 Technical Conference : 221-229 (June 1994).

Computer Programs

Lander, E.S. et al. MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1: 174-181 (1987).

Lander, E.S. and Botstein, D. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121: 185-199 (1989).

Lincoln, S.E. and Lander, E.S. Constructing genetic linkage maps with MAPMAKER: A tutorial and reference manual. Whitehead Institute Technical Report: 107 (1987).

Patterson, A. et al. Resolution of quantitative traits into mendelian factors by using a complete RFLP linkage map. Nature 335: 721-726 (1988).

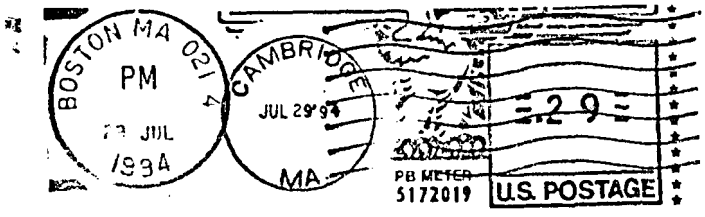
WHITEHEAD INSTITUTE/MIT
CENTER FOR GENOME RESEARCH STAFF

Director	Eric S. Lander
Associate Directors	David C. Page
	Nathan Goodman
Mouse Genome Mapping:	
Co-Directors	William F. Dietrich
	Joyce C. Miller
Assistant Director	Robert Steen
Human Genome Mapping:	
Director	Thomas J. Hudson
Assistant Director	Sebastian Gerety
Sequencing Core:	
Director	Trevor Hawkins
Assistant Director	Kim Ingalls
Informatics Core Director	Nathan Goodman
Informatics Asst. Director	Lincoln D. Stein

Whitehead Institute/MIT
Center for Genome Research
One Kendall Square, Bldg 300
Cambridge, MA 02139-1561

Phone: (617) 252-1900
FAX: (617) 252-1902

This newsletter will be released quarterly by mail and is also available on the World Wide Web URL at "http://www-genome.wi.mit.edu/".



Whitehead Institute / MIT Center for Genome Research
One Kendall Square, Bldg 300, Cambridge, MA 02139-1561

Jane Peterson
Central Genome Office
National Institutes of Health
Bldg 38A, 6th floor
9000 Rockville Pike
Bethesda, MD 20892

WHITEHEAD INSTITUTE

HUMAN GENOME

Généthon to Sequence Promoters

In its short 4-year life, the Généthon genome center, in Evry, near Paris, has carved out a position as a world leader in human genome mapping. Now, it is hoping to move into the front ranks of another area of human genomics: large-scale DNA sequencing. To do this, it is embarking on its first major collaboration with a for-profit company—a potentially controversial move for a lab funded through public donations. Last week, Généthon's paymaster, the French Muscular Dystrophy Association (AFM), announced an $11-million, 2-year joint initiative with Genset, a Paris-based genomics company, to establish an independent lab within Généthon. Dubbed the Très Grand Séquençage (TGS) laboratory, it will employ some 25 staff members at Evry and operate a battery of more than 20 automated DNA sequencers, working around the clock.

Several labs worldwide are planning similar-sized sequencing efforts, but the TGS project is unique because it is targeted at the promoter sequences that regulate gene expression. When proteins called transcription factors bind to these promoter sequences, the sequences trigger the production of messenger RNA (mRNA)—genetic messages copied from stretches of DNA that code for proteins. Several major sequencing labs are working on complementary DNA—DNA copies derived in the lab from this mRNA—to identify the coding regions of the roughly 100,000 genes contained in the human genome. Marc Vasseur, Genset's chief scientist, argues that by targeting the sequences that regulate mRNA production, TGS will complement these efforts.

Collaborating with Genset was a necessity for AFM: It could not have afforded to launch the project alone, given that it is also planning a major effort, dubbed Généthon II, to clone the genes underlying neuromuscular disease (*Science*, 18 March, p. 1554). Nevertheless, AFM general secretary Pierre Birambeau is confident that the terms of the agreement with Genset will deflect any criticism of Généthon's new link with the world of commerce: Sequence data derived from the project will be made public 6 months after collection, and any revenues that subsequently accrue to AFM will be plowed back into research on neuromuscular diseases.

The project could help in "discovering connections between genes."
—Moshe Yaniv

TGS's main strategy, says Vasseur, will be to make multiple copies of the beginning of each mRNA protein-coding message. These amplified sequences will then be used to probe the genome, binding to the DNA from which they were transcribed. By sequencing about 1000 bases of DNA upstream from that point, Vasseur claims, it should be possible to hit the majority of the corresponding promoter sequences. "I think it's a worthwhile thing to try," agrees gene control expert Robert Tjian of the University of California, Berkeley. But for a given gene, he warns, the sequences that bind to transcription factors may be spread over many thousands of bases of DNA and are not always easy to identify merely by examining the raw sequence data. Nevertheless, other researchers are more optimistic that the TGS project will yield useful information. "It certainly will be interesting in discovering connections between genes," says gene expression researcher Moshe Yaniv of the Pasteur Institute in Paris. Comparison of promoter sequences from different genes, he says, should indicate which genes are controlled by the same transcription factors.

Both Genset and AFM also hope the TGS project will lead to clinical advances. For Genset, the targets are conditions such as cancer and inflammatory disease, including arthritis. The company is developing "transcription factor decoys"—short pieces of DNA that mimic the promoter sequences of, say, cancer-causing oncogenes. These should bind to the relevant transcription factor and so "mop them up" and prevent the genes from being activated. AFM, meanwhile, has its long-term sights set on gene therapy for neuromuscular disorders and hopes to find promoter sequences that could be used to control the expression of therapeutic genes.

–Peter Aldhous

U.S. R&D POLICY

Clinton Inaugurates Science Council

Last week, U.S. science and technology policy stood at the top of the President's agenda —for about 30 minutes. That's how long Bill Clinton presided over the first meeting of the National Science and Technology Council (NSTC). The council was created by executive order last November, but it wasn't until 30 June, at 11:30 a.m., that its 24 members—Cabinet secretaries, agency heads, and senior White House officials—finally got together to discuss efforts to coordinate the government's $75-billion R&D portfolio.

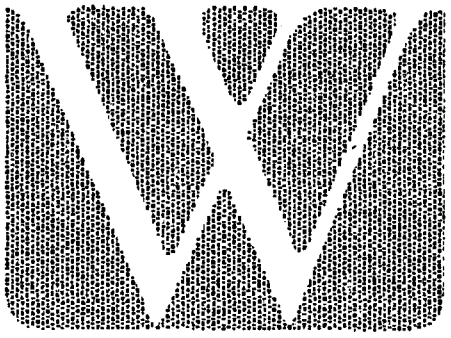
The 75-minute session in the Roosevelt Room of the White House (Vice President Al Gore took over after Clinton departed in midmeeting) didn't alter the course of U.S. policy toward research. Rather, its major purpose was to tell the President about the activities of the nine committees that serve as the operating units of the NSTC. These committees, which together form what Gore described at the meeting as "a virtual agency," are supposed to set spending priorities within nine areas that stand at the intersection of science and society—from health, transportation, education, and the environment to national security and civilian technologies. The first tangible results of their labor, which began over the winter and included two large conferences (*Science*, 4 February, p. 604, and 25 March, p. 1675), will appear in the President's 1996 budget request to Congress next February.

The President opened the meeting by stressing the importance of the federal investment in both fundamental and applied research. Then he went around the room, asking each participant to describe his or her activities on behalf of science and technology. "The discussions were strategic, not tactical," said a White House aide.

The NSTC is, in theory, on a par with the National Security Council in the White House hierarchy. But there's little chance the two councils will be equal in influence. The security council meets frequently, and its members place a high priority on attending. Last week's NSTC meeting took 7 months to arrange, in part because of the difficulty of finding time on the President's schedule. Even so, the meeting was called with less than 2 days' notice, and some of the members were forced to send stand-ins: Nine of the 24 principals dispatched their deputies or lower ranking officials. Among the no-shows was National Institutes of Health (NIH) Director Harold Varmus, who had lobbied hard for NIH membership on the council. Varmus and his family were bicycling through France as part of a long-planned vacation, and NIH's seat was occupied by deputy director Ruth Kirschstein.

There was no mention at the meeting of the President's Committee of Advisors on Science and Technology, created last fall but yet to be formed. Its 15 members are supposed to provide the president with input from the academic and private sectors on science and technology matters; the committee's charter expires in November 1995.

–Jeffrey Mervis



Pls. file

WHITEHEAD INSTITUTE/MIT
Center for Genome Research
One Kendall Square, Building 300
Cambridge, MA 02139 USA
TEL: (617) 252-1900
FAX: (617) 252-1902

TO: Jane Peterson
National Institutes of Health/NCHGR

FAX#: 301-480-2770

DATE: July 29, 1994

FROM: Eric Lander/Rachel Boucher

RE: First Issue of Newsletter

6 page(s) including this cover sheet.

July 1994 Issue 1

The Whitehead Institute/MIT Center for Genome Research (CGR) opened in July, 1993 at One Kendall Square in Cambridge, supported by a grant from the National Center for Human Genome Research. The major goals of the Center are: (1) to construct genetic and physical maps of the mouse genome; (2) to create a physical map of the human genome; and (3) to ensure that these resources are distributed to the scientific community in a timely and convenient fashion. These maps and others like them provide crucial infrastructure for the study of mammalian genetics and should be valuable in the identification of disease genes.

In order to accomplish its goals, CGR is organized in project teams and cores. At present, they are: Mouse, Human, Informatics, Sequencing and YAC.

This is the first issue of our Center newsletter, which is planned to appear quarterly. Its goal is to provide up-to-date information about progress and to describe resources available from the Center. We solicit your suggestions about ways to improve the utility of this newsletter in subsequent editions.

MOUSE GENETIC MAP REACHES 4537 SSLPs

The Mouse Genome Mapping Project aims at building genetic and physical maps covering the entire mouse genome. Such maps should make it rapid and simple for any mouse geneticist to map monogenic or polygenic traits and to obtain cloned DNA spanning the region of interest. In this way, tedious mapping and walking efforts should become unnecessary.

The first step is the construction of a high density genetic map consisting of 6,000 simple sequence length polymorphisms (SSLPs). These markers are easily typed by PCR and have a high polymorphism rate in both interspecies and intraspecies crosses. The vast majority of these markers are anonymous CA_n-repeat loci, isolated and sequenced from a whole genomic library.

(continued on page 2)

HUMAN PHYSICAL MAP TOPS 3419 STSs

The Human Genome Mapping Project aims at constructing a physical map of the human genome by the strategy of STS content mapping. The goal is to screen a total of 10,000 sequence tagged sites (STSs) by mid-1996 on 25,000 YACs from the CEPH mega-YAC library (average size 1 Mb), in order to identify the YACs containing each STS. In addition, the STSs will be screened on a panel of 'whole-genome' radiation hybrids to provide an important measure of top-down order. The STSs will consist of about 5,000 genetically mapped simple sequence length polymorphisms (SSLPs) from other groups (principally, Jean Weissenbach's group at Genethon and the CHLC consortium) and 5,000 completely random STSs developed at CGR.

(continued on page 2)

MOUSE GENETIC MAP

(continued from page 1)

The genetic markers are all genotyped in a single (OB x CAST) F_2 intercross with 46 progeny, for a total of 92 informative meioses. With this cross, genetic markers can thus be placed in "bins" with an average size of 1.1 centiMorgans (cM). The genetic markers are also tested in twelve inbred mouse strains, to determine their allele sizes.

As of the July 1994 data release, the genetic map contains 4537 SSLPs—with an average spacing of 0.30 cM or 660 kb. The SSLP map is also being closely integrated into the mouse gene map. Approximately one-quarter of the SSLP markers will be genotyped in DNAs from a (B6 x SPRET) x B6 backcross of N. Copeland and N. Jenkins, which has been scored for RFLPs in more than 1000 genes. To date, some 250 markers have been typed in this cross. In addition, CGR entered into a collaboration with the European Collaborative Interspecific Backcross (EuCIB) to map the SSLPs at higher resolution, by typing them in a 1,000-progeny backcross.

Once the genetic map is complete, the project will turn to physical mapping. The initial goal is to construct a physical map by STS-content mapping of mouse YAC libraries. A total of 10,000 STSs will be used, consisting of the 6,000 SSLPs and some 4,000 random STSs. This collection will provide anchor points with an average spacing of 300 kb. Using these STSs, we plan to screen YAC libraries with an average size of about 700kb. CGR has already constructed a 4.3-fold coverage mouse YAC library, most of which has an average insert size of about 700 kb. Additional YAC libraries are under construction. Once the initial STS content mapping is completed, we hope to close as many gaps as possible in a directed fashion.

HUMAN PHYSICAL MAP

(continued from page 1)

Over the past year, we have been developing methods to scale up physical mapping, which includes an automated PCR setup process which generates 18,000 PCR reactions per day. The detection of these PCR products is done by a chemiluminescent hybridization assay, and data capture using CCD cameras.

YAC screening is currently carried out by a two-level scheme. At the top level, we screen superpools from eight microtiter plates. From each positive superpools, we then screen 28 subpools corresponding to the rows, columns and plates of a block. These coordinates should provide a unique address provided that the block contains only a single YAC containing the STS; the address can be ambiguous (i.e., have more than one row, column or plate) if the block has two or more YACs containing the STS.

To date, we have identified YACs for 3419 STSs which fall into the following categories: (i) 1427 Genetically mapped polymorphic STSs, which allow contigs to be anchored to the genetic map; (ii) 838 Random genome-wide STSs, of which approximately 75% are unambiguously assigned to a chromosome using the NIGMS Human/Rodent Somatic Cell Hybrid Mapping Panel #1; (iii) 857 unpublished CA-repeat-containing STSs generously provided by J. Weissenbach. These CA-repeats, which were not sufficiently polymorphic to be genotyped for the Genethon human genetic map, are an additional source of random STSs; (iv) 114 new chromosome 22 STSs generated from sequences derived from flow-sorted chromosome libraries, in collaboration with the Human Genome Center for Chromosome 22 in Philadelphia. These STSs and the corresponding YACs already provide considerable coverage of the long arm of chromosome 22; and (v) 253 STSs from public data bases.

With the current data, the average spacing between STSs is about 875 kb. This is still not sufficient to allow the assembly of a comprehensive and reliable STS content map. As additional STSs are added, contigs covering substantial portions of the genome are expected to fall together.

CGR Resources

CGR strives to ensure rapid and convenient access by the scientific community to information, reagents, and software tools developed by our mapping projects.

Databases

Human physical mapping and mouse genetic mapping data are released on a quarterly basis in January, April, July, and October. The data is available in a number of ways.

Via ftp - You will need access to an *ftp* program such as *Fetch* on the Macintosh. Set your program to log into **genome.wi.mit.edu**. Use "anonymous" for the user name, and use your e-mail address for the password. Data files are stored in **/distribution/mouse_sslp_releases** and **/distribution/human_STS_releases.**

Via e-mail - Send e-mail to the address **genome_database@genome.wi.mit.edu**, with the word "help" appearing as the first word on the subject line or body text. You will receive instructions for accessing the data by return mail. As of spring 1994, only the mouse genetic mapping information is available via this route.

Via World Wide Web - You will need a World Wide Web client, such as *Mosaic* (widely available for multiple platforms). Tell your client to connect to **http://www-genome.wi.mit.edu**, and follow the links to the data directories.

For further help with database services, call Lincoln Stein, Assistant Director of Informatics, 617-252-1916, lstein@genome.wi.mit.edu.

Software

Software is available via World Wide Web and ftp. All software is stored in the directory **/distribution/software.**

The programs currently available are as follows:

MAPMAKER-an interactive computer package for construction of genetic maps in experimental crosses and human reference families.

MAPMAKER/QTL-an interactive computer package for genetic mapping of quantitative trait loci (QTLs) in experimental crosses, using LOD scores.

PRIMER-a computer program for selection of PCR primers satisfying specified conditions.

EXCEL TCP/IP PLUGIN- a Microsoft Excel add-on that allows regions of the spreadsheet to be sent to and received from UNIX hosts on the Internet.

Distribution of Biological Reagents

CGR tries to promote broad and immediate access to biological reagents, by encouraging and assisting distribution services by the private sector.

• To ensure access to mouse and human STSs, CGR pioneered in 1990 an arrangement under which Research Genetics, Inc. retains a portion of all PCR primer pairs synthesized for our use and sells aliquots to the scientific community at discount prices, under the name "MapPairs." The arrangement has since been extended to include PCR primers from other genome centers, as well.

• To ensure access to mouse and human YACs, CGR distributed copies of its mouse YAC library (4.3-fold coverage in YACs of about 700kb; Kusumi et al. Mammalian Genome, 4:391-2 (1993)) and the CEPH mega-YAC library (plates 613-984) to both Research Genetics, Inc. and Genome Systems, Inc. These companies provide library services, including: screening service for individual STSs; purchase of YAC DNA pools for PCR screening of STSs; and purchase of a copy of the entire library. For more information contact:

Research Genetics, Inc.
2130 Memorial Parkway, SW
Huntsville AL 35801
Phone: (800)-533-4363
Fax: (205) 536-9016

Genome Systems, Inc.
7166 Manchester Road
St. Louis MO 63143
Phone: (800)-248-7609
FAX: (314)-647-4134

CGR receives no royalties from distribution of reagents by private companies.

Policy on Data Release and Patents

To promote the broadest possible use of the biological tools developed under the Human Genome Project, Whitehead Institute has adopted a policy that:

• Genetic and physical maps (including all clones, genetic markers, primers and sequences)

CGR Resources Continued

will be made promptly available to the scientific community, will be placed in the public domain, and will not be patented.

• No access to maps will be granted to any commercial entity in advance of public release.

Data releases are scheduled every 90 days. At the end of each calendar quarter, all genomic mapping data are reviewed and prepared for distribution via CGR's electronic databases. Data releases typically occur within two weeks of the close of the quarter (i.e., in mid-January, mid-April, mid-July and mid-October). Releases are announced by electronic messages posted to the following two newsgroups: "bionet.genome.chromosomes" and "bionet.announce".

The purpose of CGR's data release policy is to ensure that scientific colleagues have immediate access to information that may assist them in the search for genes. Data releases do not constitute scientific publication of CGR's work, but rather provide scientists with a regular look into our lab notebooks. For projects aimed at the analysis of particular genes or small subchromosomal regions, permission is hereby granted to use our data without the need for a formal collaboration, subject only to appropriate acknowledgement. For projects aimed at large-scale mapping of entire chromosomes or entire genomes, use of the data and markers should be on a collaborative basis. The information for the mouse genetic map should be cited as:

Whitehead Institute/MIT Center for Genome Research, Genetic Map of the Mouse, Data Release 7 (July 1994)

Dietrich, W.F. *et al.* A genetic map of the mouse with 4,006 simple sequence length polymorphisms. *Nature Genetics* 7:220-245 (1994).

The information for the human genome mapping project should be cited as:

Whitehead Institute/MIT Center for Genome Research, Human Genetic Mapping Project, Data Release 2 (July 1994).

If you would like to receive future issues of the Whitehead Institute/MIT Center for Genome Research Newsletter in hard copy, please complete and send in the following form:

*Issues are also available on the World Wide Web. Our URL is http://www-genome.wi.mit.edu/

Name

Institution

Department

Street/Building

City, State, Zip (Country)

Return to: Newsletter Editor, Whitehead Institute/MIT, Center for Genome Research, One Kendall Square, Building 300, Cambridge, MA 02139-1561 or e-mail: newsletter @genome.mit.edu

References

The following references may provide useful background on CGR projects and related topics.

Mouse Mapping

Copeland, N.G. et al. A genetic linkage map of the mouse: Current applications and future prospects. Science 262: 57-66 (1993).

Copeland, N.G. et al. Genome maps IV: The mouse. Science 262: 67-82 (1993).

Dietrich, W.F. et al. A genetic map of the mouse suitable for typing intraspecific crosses. Genetics 131: 423-447 (1992).

Dietrich, W.F. et al. A genetic map of the mouse with 4,006 simple sequence length polymorphisms. Nature Genetics 7:220-245 (1994).

Dietrich, W.F. et al. in Genetic Maps 1992 (ed. O'Brien, S.) 4.100-4.142 (Cold Spring Harbor Laboratory Press, New York, 1992.)

Kusumi, K. et al. Construction of a large-insert yeast artificial chromosome (YAC) library of the mouse genome. Mammalian Genome 4:391-392 (1993).

Miller, J.C. et al. SSLP/Microsatellite genetic linkage map of the mouse. In M. F. Lyon and A.G. Searle (eds.) Genetic Variants and Strains of the Laboratory Mouse, Third Edition, Oxford University Press, Oxford (1994).

Human Mapping

Foote, S. et al. The human Y chromosome: Overlapping DNA clones spanning the euchromatic region. Science 258: 60-66 (1992).

Hudson, T.J. et al. Isolation and chromosomal assignment of 100 highly informative human simple sequence repeat polymorphisms. Genomics 13:622-629 (1992).

Vollrath, D. et al. The human Y chromosome: A 43-interval map based on naturally occurring deletions. Science 258: 52-59 (1992).

Informatics

Goodman, N. et al. Building a laboratory information system around a C++-based object-oriented DBMS. In Proceedings of the 20th International Conference on Very Large Data Bases. Santiago (1994).

Goodman, N. et al. Requirements for a deductive query language in the MapBase genome-mapping database. In Proceedings of the Workshop on Programming with Logic Databases: 18-32 Vancouver (Oct. 1993).

Goodman, N. An object-oriented DBMS war story: Developing a genome mapping database in C++. In Modern Database Management: Object-Oriented and Multidatabase Technologies, Won Kim, ed. ACM Press, NY (1994).

Rozen, S. et al. Constructing a domain-specific using a persistent object system. Sixth International Workshop on Persistent Object Systems. Tarascon (1994).

Stein, L. et al. Splicing UNIX into a genome mapping laboratory. In USENIX Summer 1994 Technical Conference : 221-229 (June 1994).

Computer Programs

Lander, E.S. et al. MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1: 174-181 (1987).

Lander, E.S. and Botstein, D. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121: 185-199 (1989).

Lincoln, S.E. and Lander, E.S. Constructing genetic linkage maps with MAPMAKER: A tutorial and reference manual. Whitehead Institute Technical Report: 107 (1987).

Patterson, A. et al. Resolution of quantitative traits into mendelian factors by using a complete RFLP linkage map. Nature 335: 721-726 (1988).

WHITEHEAD INSTITUTE/MIT
CENTER FOR GENOME RESEARCH STAFF

Director	Eric S. Lander
Associate Directors	David C. Page
	Nathan Goodman
Mouse Genome Mapping:	
Co-Directors	William F. Dietrich
	Joyce C. Miller
Assistant Director	Robert Steen
Human Genome Mapping:	
Director	Thomas J. Hudson
Assistant Director	Sebastian Gerety
Sequencing Core:	
Director	Trevor Hawkins
Assistant Director	Kim Ingalls
Informatics Core Director	Nathan Goodman
Informatics Asst. Director	Lincoln D. Stein

Whitehead Institute/MIT
Center for Genome Research
One Kendall Square, Bldg 300
Cambridge, MA 02139-1561

Phone: (617) 252-1900
FAX: (617) 252-1902

This newsletter will be released quarterly by mail and is also available on the World Wide Web URL at "http://www-genome.wi.mit.edu/".

Phil Green

Fax cover sheet Fax cover sheet Fax cover sheet Fax cover sheet Fax cover sheet Fax cover sheet

The Global Weekly of Research

SCIENCE

PUBLISHED BY THE AMERICAN ASSOCIATION
FOR THE ADVANCEMENT OF SCIENCE
1333 H STREET, N.W., WASHINGTON DC 20005

Fax number: (202) 289-7562

Plsfile
Laurie

Jane: 8/8
Yours
Bettie has hers.
M.

DATE: 7/28/94

TO: Name: Dr Francis Collins

Organization:

Fax number: 301-402-0837

FROM: Name: Dr Barbara Jasny

COVER SHEET PLUS 11 PAGES SENT

COMMENTS:

I hope you will not mind, the given our propensity for telephone tag, that I am faxing this to you to ask if you could review it for Science. As you can see, it will be controversial. I'd be happy to talk to you about it WK phone 202-326-6515 home 703-533-3007. Please let me know if you can/cannot review it

Thank you!

Barbara

Analysis of a proposed 'first-generation' physical map of the human genome

Alan Kaufman[1,2] and James Orlin[1,2,3*]

[1]Whitehead Institute/MIT Center for Genome Research, and [2]Operations Research Center, and [3]Sloan School of Management Massachusetts Institute of Technology, Cambridge MA 02139..

*To whom correspondence should be addressed.

Abstract

Cohen and colleagues [1] recently described a project to characterize a human yeast artificial chromosome (YAC) library and offered a 'proposed data analysis strategy' that was said to yield a physical map covering 87% of the human genome. The authors provided no analytical evaluation to test the validity of their novel strategy for constructing 'paths' in the genome. We have now examined the proposed method in detail. Analytical studies show that most paths with at most two YACs or spanning less than 5 cM are valid, but most paths involving four or more YACs or spanning 5 cM or more are invalid. After restricting the map to paths with a high probability of being valid, we conclude that the remaining map properly covers at most 36% of the genome.

Cohen et al. [1] recently reported the results of their efforts to construct a 'first generation' physical map of the human genome, based on the analysis of a large-insert yeast artificial chromosome (YAC) library. Briefly, the physical mapping data involved screening the 33,000-clone CEPH mega-YAC library by two different methods, STS content mapping and Alu-PCR probe hybridization. In the first method, 2100 genetically-mapped sequence-tagged sites (STSs) [2] were screened against the YAC library (with half of the STSs screened completely and half screened partially to obtain 1-2 positives). In the second method, Alu-PCR products were prepared from 6900 individual YACs and were screened by hybridization against spotted Alu-PCR products from a subset of 25,000 of the YACs and from monochromosomal hybrid cell lines. (In addition, many YACs were also subjected to hybridization-based 'fingerprinting' [3], but these data played only a minor role in the analysis and do not significantly affect the coverage; we omit them in the discussion below.)

Based on these data, Cohen et al. [1] offered a 'proposed data analysis strategy' that was said to yield a physical map covering 87% of the human genome. The proposed data analysis strategy involved creating paths between STSs. A *path* of length k between two STSs, s_1 and s_2, is defined as a series of YACs, $y_1, y_2, \ldots, y_k$ such that (1) s_1 lies in y_1 and s_2 lies in y_k by STS content mapping and (2) for each step (y_i, y_{i+1}), at least one of the two YACs was used as an Alu-PCR probe and hybridized to the other YAC. Paths of length 1 correspond to traditional STS content mapping, while longer paths depend on the Alu-PCR hybridization data. A *chromosomally allowable* path is defined to be one with the property that (1) the starting and ending STSs, s_1 and s_2, lie on the same chromosome c, and (2) each y_i that was used as an Alu-PCR probe either gave no signal when hybridized to the monochromosomal hybrid panel or hybridized to a set of chromosomes that included chromosome c. (N.B. For

hybridized to a set of chromosomes that included chromosome c. (N.B. For technical reasons, chromosomal assignments were not always unique: 49% could be assigned to a single chromosome, 18% hybridized to multiple chromosomes, and 33% could not be assigned to any chromosome.)

The first generation physical map proposed in [1] was defined to be the set of all chomosomally allowable paths of length ≤ 7 connecting pairs of STSs with genetic distance ≤ 10 cM. The authors offered no experimental or analytical justification for the choice of path length nor any analysis to suggest that most such paths are correct. They simply noted that as longer paths are allowed, the coverage of the genome increased. With paths of length one, three, five, and seven, the strategy covered 11%, 30%, 70% and 87%, respectively, of the total genetic length of the genome.

There is a serious issue to be addressed: Using the proposed approach, utterly random data might also appear to cover the genome. Specifically, it is well-known in random graph theory [4] that, in certain random structures, paths of bounded length suffice to connect essentially all pairs of points. This phenomenon has recently gained popular attention through the award-winning play, "Six Degrees of Separation", in which it is asserted that any two people in the world can be connected through a path of at most six acquaintances. It is important to evaluate whether the same phenomenon accounts for the apparently complete coverage of human genome.

We set out to evaluate the proposed data analysis strategy using the data from the March 30, 1994 CEPH data release. We first constructed the minimum-length chromosomally allowable path [5] connecting every pair of STSs located on the same chromosome—regardless of the genetic distance between them. Figure 1a shows the proportion of STSs that could be connected, as a function of the path length and the genetic distance between them. We were interested to

determine what fraction of these paths resulted from spurious random connections.

A simple way to determine the proportion of false connections is to consider apparent short paths between STSs separated by ≥ 50 cM. Such paths must surely be spurious inasmuch as the average YAC length is only 1 Mb, corresponding to only about 1 cM in the human genome. The proportion of such distant STSs connected by chromosomally allowable paths of length 1, 3, 5, and 7 is 0.05%, 2%, 18% and 61%. In particular, the curve rises dramatically for path lengths exceeding four — indicating that random connections dominate at these lengths. Interestingly, the proportion of connected STSs at distances 5–10 cM, 10–20 cM, and 20–50 cM was no higher than for STSs at distances ≥ 50 cM. This suggested that most paths connecting STSs at distances ≥ 5 cM are also false.

To test whether these apparent paths were nothing more than would be expected in an equivalent random graph [4], we performed a simple randomization experiment: We left unchanged the genetic map, the STS content data, and the chromosomal assignment of Alu-PCR probes, but randomized the hybridization results of the Alu-PCR probes against the YAC library, preserving only the correct number of hits for each probe. Consider, for example, an Alu-PCR probe that hybridized to chromosomes 3 and 7 and detected four YACs in the library. In the randomized data, it was still assigned to the same chromosomes but the four YACs that it detected were selected using a random number generator. With these random data, we again constructed minimum length paths between all intra-chromosomal pairs of STSs following the strategy of Cohen et al. (Figure 1b). Paths of length 1 remain unchanged in the randomized data, since they depend only on the STS content data. However, longer paths are entirely spurious.

Interestingly, STS pairs at ≥ 5 cM show the same degree of connectivity in the random data as in the real data—confirming our suspicion that spurious connections are the principal mechanism linking such STSs. By contrast, STS pairs at < 5 cM show significantly higher connectivity for path length ≤ 3 in the real data than in the randomized data, with the difference attributable to valid short paths between nearby STSs.

Based on this analysis, it is possible to estimate the proportion of STSs connected by valid paths (Figure 2) and the probability that a path of a given length is valid (Figure 3). The results indicate that paths of length ≤ 2 connecting STS at < 5 cM are mostly valid, whereas paths having length ≥ 4 or joining STSs at ≥ 5 cM are generally spurious. Considering only paths of length ≤ 2 connecting STSs within 5 cM, the paths in the CEPH-Genethon data cover about 36% of the genetic length of the human genome. (The percentage coverage is defined as the proportion of total centiMorgans lying between connected STSs. This may somewhat overestimate the actual proportion of the physical length covered, inasmuch the covered genetic intervals of any given size would be expected to be biased to those with enhanced recombination relative to physical distance.)

In summary, the 'proposed data analysis strategy' of Cohen et al. [1] works well when restricted to short paths, but is unreliable for longer paths. Restricting the physical map to analytically valid paths, the CEPH-Genethon physical map is estimated to cover about one-third of the human genome. To obtain reliable coverage of the entire human genome using the strategy of Cohen et al. [1], one would require many more genetically mapped (or otherwise ordered) STSs. Not withstanding this revised assessment, the CEPH-Genethon data represent a large and impressive resource of great value to the human genetics community. It will clearly play an important role in the assembly of a comprehensive physical map.

References and Notes

[1] D. Cohen, I. Chumakov, and J. Weissenbach, *Nature*, **366**, 698 (1993).

[2] J. Weissenbach et al., *Nature*, **359**, 794 (1992).

[3] C. Bellanne-Chantelot et al. *Cell*, **70**, 1059 (1992).

[4] B. Bollobas, Random Graphs. Harcourt Brace Jovanovich, 1985.

[5] R. Ahuja, T. Magnanti and J. Orlin, Network Flows. Prentice-Hall, 1993.

[6] We thank Daniel Cohen, Ilya Chumakov and Jean Weissenbach for sharing this valuable data with us and the scientific community at large. We thank Eric Lander for suggesting the use of random graphs and for comments on the manuscript and David Page, Leonid Kruglyak and Lincoln Stein for helpful discussions. This work was supported in part by NIH grant HG00098.

Figure Legends

Figure 1. Cumulative proportion of connected STS pairs, by inter-STS distance and path length, for (A) real data and (B) randomized data. Minimal paths were constructed between all intra-chromosomal pairs of STSs. STSs hitting no YACs were excluded, as these could never form paths.

Figure 2. Estimated cumulative proportion of valid STS connections, by inter-STS distance and path length. The proportion of spuriously connected STS pairs for each path length was estimated based on the "≥ 50 cM" curve in Figure 1a. This proportion was subtracted from the observed proportion of connected STS pairs to yield the estimated proportion of truly connected pairs.

Figure 3. Estimated probability that a path connecting two STSs is valid, by inter-STS distance and path length. The probability a path is valid was approximated by $\max[(p_o-p_s)/p_o, 0]$, where p_o is the observed proportion of connected STSs and p_s is the proportion of spuriously connected STSs estimated from the randomized data. The results are similar if p_s is estimated from the "≥ 50 cM" curve in Figure 1a.

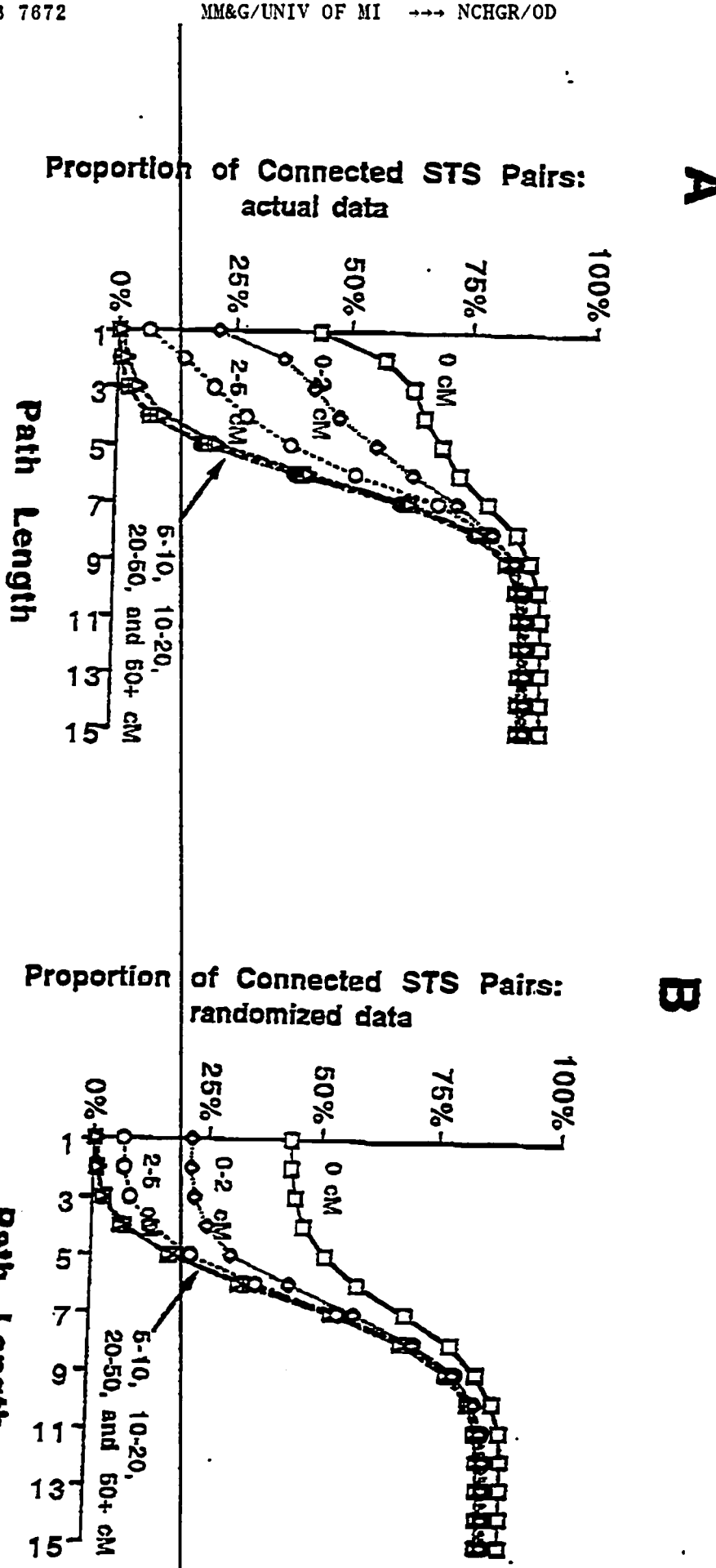

A
Proportion of Connected STS Pairs:
actual data
0%
25%
50%
75%
100%
Path Length
1
3
5
7
9
11
13
15
0 cM
0-2 cM
2-5 cM
5-10, 10-20,
20-50, and 50+ cM
B
Proportion of Connected STS Pairs:
randomized data
0%
25%
50%
75%
100%
Path Length
1
3
5
7
9
11
13
15
0 cM
0-2 cM
2-5 cM
5-10, 10-20,
20-50, and 50+ cM

Figure 1

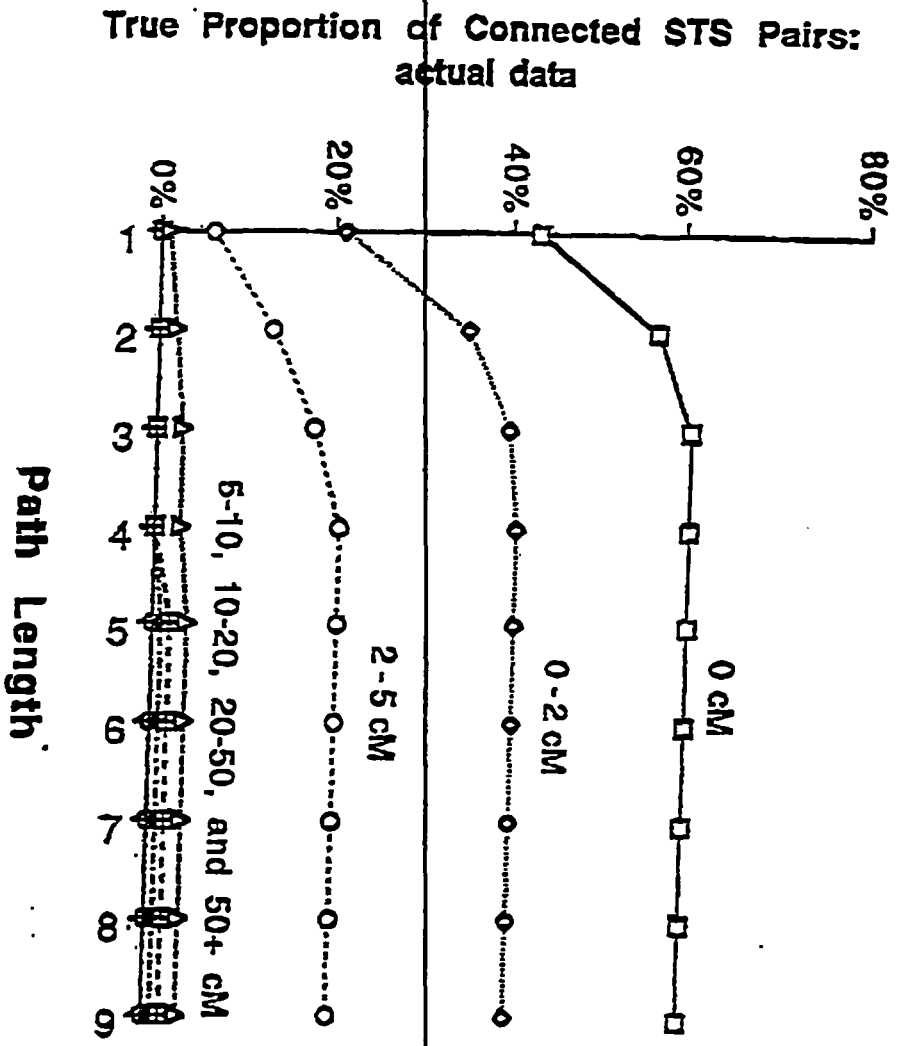

True Proportion of Connected STS Pairs:
actual data
0%
20%
40%
60%
80%
1
2
3
4
5
6
7
8
9
Path Length
5-10, 10-20, 20-50, and 50+ cM
2-5 cM
0-2 cM
0 cM

Figure 2

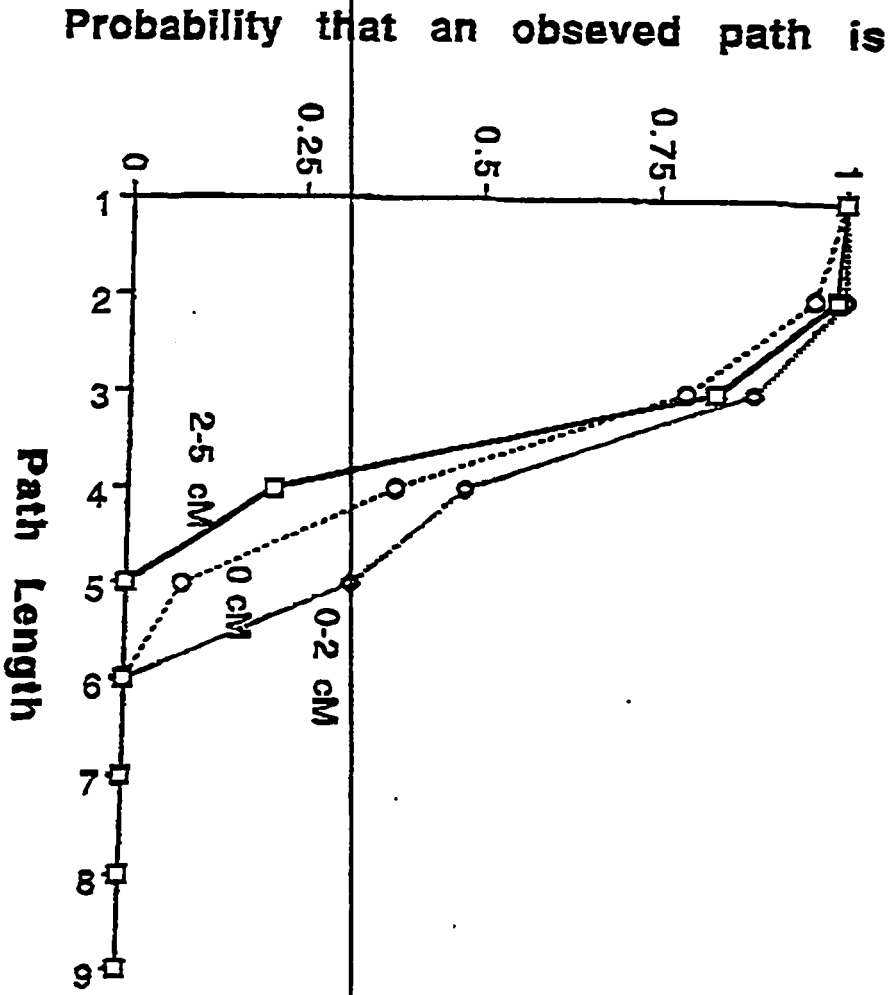

Probability that an obseved path is valid
0
0.25
0.5
0.75
1
Path Length
1
2
3
4
5
6
7
8
9
2-5 cM
0 cM
0-2 cM

Figure 3

Human Physical Mapping Project
Whitehead Institute/MIT
Center for Genome Research

Phase 1- Primary Semi-automated YAC Screening

Goal: 3000 STSs

Methods: Rosys Runs (192 well plates, 8 head pipettor)
Waffle Irons (192 well plates)
6144-Spotter
Image capture of Autoradiography

Results: 2555 markers to date
5.5 hit rate 4.7–6.5
65% definite addresses
6% apparent false positive rate
200 double-linked contigs, mostly with 2 STSs

Phase 2 - High Throughput YAC Screening

Goal: 10,000 STSs (and beyond)

Methods: Genomatron
CCD capture of chemiluminescent signal
Computer Analysis of images
Radiation hybrid mapping of 5000 genome wide STSs

Phase 3 - Map Assembly and Validation

Goal: Closure

Methods: Verification of YAC addresses using semiautomated set-up
Rapid recovery of YAC ends for STS generation

TOOLS & TECHNOLOGY

Supporting PCR, New Thermal Cyclers Find Diverse Laboratory Uses

BY CAREN D. POTTER

Thermal cyclers—or automatic temperature cyclers—have not been around very long, but, having ridden to popularity on the coattails of the polymerase chain reaction (PCR), they are fast becoming essential laboratory instruments for many biological researchers.

PCR is the DNA amplification process introduced in the 1980s that has revolutionized genetics-related research. PCR replicates a small amount of DNA in a series of heating and cooling steps and has been used in diverse research applications, including molecular biology, epidemiology, and paleontology. Reflecting the importance of the innovative process, PCR's inventor Kary Mullis was awarded this year's Nobel Prize in chemistry (see story on page 1). Thermal cyclers, for their part, have cut the time needed to run PCR by as much as two-thirds.

"In the 2½ years I've been in this industry, I've seen the uses for PCR and the market for thermal cyclers expand dramatically," says Karen Studer-Rabeler, director of new product development at Coy Corp., a thermal cycler manufacturer located in Grass Lake, Mich. "PCR is used in anything from the study of fossil ambers to genetic engineering of corn."

Thermal cyclers allow the PCR process to proceed automatically by subjecting the reagents—DNA nucleotides and a heat-tolerant polymerase, among others—to a user-specified heating and cooling sequence. In PCR, a thermal cycler heats samples to open the double helices of DNA, lets the temperature drop to bind primers, increases the temperature somewhat to build new strands, then heats up again to begin a new cycle.

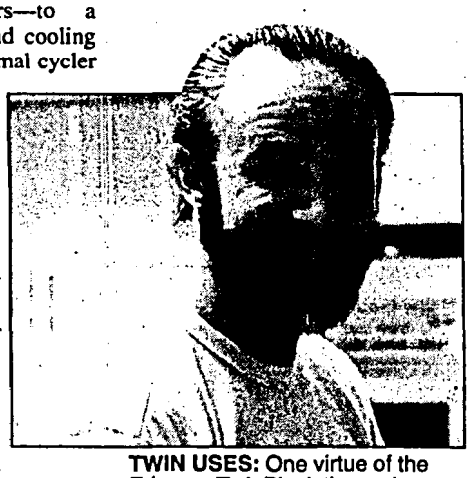
The development of thermal cyclers lagged behind that of PCR itself because the first enzymes used for PCR were thermolabile (unstable when heated, and therefore unusable after one cycle), explains Simon Foote, senior research scientist at the Whitehead Institute for Biomedical Research in Cambridge, Mass. PCR had to be done manually by placing sample tubes in water baths set at various temperatures, then adding new enzymes to the tubes after each heat cycle. "There was no way to automate the process with a device such as a thermal cycler until thermostable enzymes became available," Foote says.

Such enzymes are now available, making the use of thermal cyclers a significant improvement over the manual method. The most significant benefits of thermal cyclers are unattended operation, faster throughput (since thermal cyclers are designed to reach target temperatures as quickly as possible), and enhanced temperature control to provide uniform heating and cooling over the entire body of samples.

Capacity Range

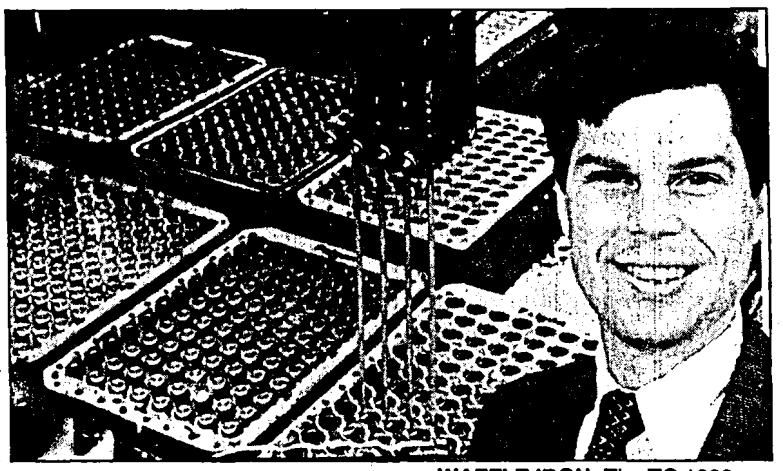
One of the most striking ways in which the thermal cyclers now available differ from each other is in the number of samples they are designed to process at once. At one end of the spectrum is a small, lightweight model called the MiniCycler, from M.J. Research in Watertown, Mass., that has a capacity of 16 0.5-ml tubes or 25 0.2-ml tubes. At the other end is what is commonly known as "the waffle iron" because the honeycomb pattern of its large well plates resembles the surface of that appliance. The official name of this instrument is the TC 1600 Thermocycler, and it is made by IAS Products Inc. of Cambridge, Mass. Depending on the configuration chosen by the researcher, it can process simultaneously up to 3,072 samples (16 microtitration plates times 192 wells).

"The waffle iron was spun out of a custom project we did for the Whitehead Institute to help them automate their work on the Human Genome Project," says Steven Gordon, president of IAS Products. This thermal cycler is the most expensive on the market at $45,000, but, as Gordon says, "It's cost-effective if you need that kind of throughput." The MiniCycler, by contrast, sells for $2,795.



TWIN USES: One virtue of the Ericomp TwinBlock thermal cycler, says David Brown of the University of Georgia, Athens, is that it can run two experiment programs at once.

Four waffle irons equipped with sixteen 96-well plates are in constant use at the Whitehead Institute, supporting the institute's work of mapping the complete human genome. "We average three runs per waffle iron per day," says Foote. The Whitehead lab is in the process of converting to 192-well plates for even greater capacity, he adds.



WAFFLE IRON: The TC 1600 Thermocycler from IAS Products—commonly called the "waffle iron" because it resembles that appliance—is capable of processing 3,072 samples at once, says company president Steven Gordon.

Some thermal cyclers, the waffle iron included, offer researchers the ability to divide the instrument's capacity into independently cycling sections. For example, the waffle iron can process four different heating and cooling profiles, one for each quadrant of the device. A smaller, more affordable model called the TwinBlock System from Ericomp Inc., San Diego, has the ability to run two different cycling programs simultaneously. David Brown, a research coordinator who works with a TwinBlock in a University of Georgia in Athens genetics lab, praises this feature.

"Aside from the confidence that the instrument reliably produces the temperatures you expect from a particular program, the ability to run two independent programs was a real selling point," he says. "Often two people in our lab run different programs on the TwinBlock. If you had another machine with the same capacity but only one cycling program, others would have to wait until the first person was finished."

Heating And Cooling

Thermal cyclers must reach appropriate temperatures quickly and provide a uniform temperature over all samples. To achieve these objectives, manufacturers of thermal cyclers have turned to different technologies for heating the samples and then cooling them down. Most, but not all, use an electrically heated element to deliver heat to a metal plate (usually aluminum) that surrounds the sample tubes.

For cooling, several approaches are used. Some models do not offer active control when it comes to cooling, they simply let excess heat escape into the ambient air. "These are the cheapest to manufacture, but they can have uniformity problems," says John Hansen, director of special projects at M.J. Research.

Another method of cooling is that used by Perkin-Elmer, the largest manufacturer of thermal cyclers. This approach relies on a vapor compression heat pumping, which is similar to a typical refrigeration unit. Other devices such as the waffle iron use water for cooling the samples. "You can get much more efficient cooling out of water because there is a physical mass that absorbs the heat and pulls it away," says Gordon.

Efficient cooling is a must for a unit that generates as much heat as the waffle iron. Because it handles such a large number of samples, this device requires a tremendous amount of power. "When you start multiplying things by 16 [the number of microtitration plates in the waffle iron], you start getting to numbers like 200 volts times 70 amps," says Gordon. "This becomes a potentially dangerous device." (Compare this with the requirements of a clothes dryer or oven, about 10 amps each.) IAS Products built five redundant safety systems into the waffle iron, Gordon adds.

Another technology used in thermal cyclers is an electronic process called the Peltier effect. Depending on the direction of the electrical current in a Peltier unit—two ceramic outer layers sandwiching an inner layer of semiconductor material—it can actively transport heat either into or out of a sample block. As current passes through the semiconductor material, electrons migrate from one surface of the sandwich to the other, dragging a small amount of heat with them. This effect can cause a temperature differential between the top and bottom of the unit of as much as 70 degrees C. Reversing the flow of the current reverses the flow of heat as well.

Discovered in 1834 by Jean Peltier of France, this electronic means of pumping heat remained a lab curiosity until the 1930s, when Maria Telkes, a solid-state physicist at Westinghouse Research Laboratories, discovered how to use a crystal instead of a bi-metallic junction in the device, according to Hansen of M.J. Research. "Telkes's findings increased the efficiency of Peltier units an order of magnitude." Today's Peltier units are efficient semiconductor heat pumps that involve no moving parts or chlorofluorocarbons.

M.J. Research and Coy Corp. introduced thermal cyclers based on the Peltier effect in 1988. Thermal cyclers from M.J. Research have bi-

(Continued on Page 19)

Thermal Cyclers

(Continued from Page 17)

directional Peltier control (that is, the Peltier effect is used for both heating and cooling); models from Coy use the Peltier effect only for cooling.

Initially, the materials used in Peltier units proved problematic for thermal cycling applications. "They were designed for steady-state conditions where the temperature doesn't vary," says Hansen. "If you put these modules into a thermal cycler they wouldn't last very long, which is why many manufacturers have shied away from them. We've devoted years of research to building better Peltier units specifically for a temperature cycling process."

Using the Peltier effect for both heating and cooling makes thermal cyclers from M.J. Research highly adaptable to field conditions. One research team took MiniCyclers to the McMurdo Sound region of Antarctica to investigate genetic diversity in moss. "Preliminary isozyme and morphological studies gave no conclusive clues, but with our little MiniCyclers we were able to conduct DNA amplification at two sites in the field," says Dieter Adam, principal investigator from the University of Waikato in New Zealand. "A little gas generator could run both a MiniCycler and a gel box simultaneously and the speed of the machine allowed us to run several amplifications a day."

In Situ Amplification

DNA amplification was, until recently, always performed in tubes.

Photo: Franklin Hoke

TOUGH IN THE FIELD: The rugged PTC-100 thermal cycler from M.J. Research was used on the battlefield in the Gulf War to perform PCR testing for biological warfare agents, says special projects director John Hansen. It accepts tubes or microassay plates for in situ PCR.

Although this method is unquestionably a powerful tool for molecular biologists and related researchers, those who deal with whole organisms often need to know the location within the cell of the DNA sequence of interest. With traditional DNA amplification procedures, they may know that there was at least one template in the tube when they started the process, but not where it came from.

With in situ DNA amplification, sections of tissue are put on glass slides and the process is carried out while the DNA is still inside the cell. "This technique has not been perfected, and there are some who doubt its ultimate validity, but others consider in situ DNA amplification to be the most significant breakthrough in molecular biology since the development of PCR," says Hansen.

Since in situ amplification still requires temperature cycling, thermal cyclers can automate the procedure in much the same way they automate the process when it takes place in tubes. Several vendors have already adapted their instruments to handle slides. With these devices, PCR can now be performed in morphologically intact cells, making the process more useful in applications such as clinical diagnostics, particularly virology, histopathology, and detection of genetic mutations.

For a detailed protocol for in situ amplification, see O. Bagasra, et al., *Journal of Immunological Methods*, 158:131-45, 1993. Also, Coy Corp. offers a technical brochure on the procedure. Even before these in situ units became available, innovative researchers were taking matters into their own hands and modifying their traditional tube thermal cyclers with aluminum foil to accommodate slides. ❑

Caren D. Potter is a freelance science writer based in McKinleyville, Calif.

THERMAL CYCLER VENDORS

The following suppliers are among those offering thermal cyclers for use in PCR-related experiments.

Applied Biosystems
Division of Perkin-Elmer Corp.
850 Lincoln Center Dr.
Foster City, Calif. 94404
(415) 570-6667
Fax: 572-2743
(800) 545-7547 (for sales information and ordering)

Coy Corp.
14500 Coy Dr.
Grass Lake, Mich. 49240
(313) 475-2200
Fax: (313) 475-1846

Ericomp Inc.
6044 Cornerstone Court West
Suite E
San Diego, Calif. 92121
(619) 457-1888
Fax: (619) 457-2937

IAS Products Inc.
142 Rogers St.
Cambridge, Mass. 02142
(617) 354-3830
Fax: (617) 547-9727

M.J. Research Inc.
149 Grove St.
Watertown, Mass. 02172
(617) 924-2266
Fax: (617) 924-2148

A Genetic Map of the Mouse Suitable for Typing Intraspecific Crosses

William Dietrich,*,† Hillary Katz,* Stephen E. Lincoln,* Hee-Sup Shin,*,† Jeffrey Friedman,‡ Nicholas C. Dracopoli† and Eric S. Lander*,†

**Whitehead Institute for Biomedical Research, Cambridge Massachusetts 02142, †Center for Genome Research and Department of Biology, Massachusetts Institute of Technology, Cambridge Massachusetts 02139, and ‡The Rockefeller University, New York, New York 10021*

Manuscript received December 3, 1991
Accepted for publication February 19, 1992

ABSTRACT

We report the construction of a genetic linkage map of the mouse, consisting entirely of genetic markers that can be rapidly typed by polymerase chain reaction and that show a high degree of polymorphism among inbred laboratory strains. Specifically, the map contains 317 simple sequence length polymorphisms at an average spacing of 4.3 cM and is detectably linked to approximately 99% of the mouse genome. In typical crosses between inbred laboratory strains, about 50% of the markers are polymorphic, making it straightforward to follow inheritance in almost any cross.

THE mouse is a powerful genetic system for the study of mammalian biology: a century of work has yielded thousands of mutants defining single gene variation and scores of inbred strains defining polygenic variation affecting physiology, development and behavior (GREEN 1989; FESTING 1979). Because most of these genes are known only by their phenotypic effect, detailed study requires cloning the genes based on their chromosomal position relative to a genetic map. The ideal genetic map for this purpose would consist of genetic markers that were (1) highly abundant and evenly distributed, so that the entire genome could be simultaneously followed in a cross; (2) highly polymorphic, so that one could study any cross between laboratory strains; (3) rapidly typed, so that scoring a cross would be short relative to generation time; and (4) easily disseminated, so that any laboratory would have ready access to them. Such a genetic map would allow initial localization of genes and then provide starting points for chromosomal walks to clone them.

The first genetic map of the mouse was based on visible mutant phenotypes. Given the difficulty of isolating large numbers of mutants and the considerable effort needed to map two mutations relative to one another, this work proceeded slowly. Although the first linkage group in the mouse was found (HALDANE, SPRUNT and HALDANE 1915) soon after the notion of linkage was first elucidated in Drosophila (STURTEVANT 1913), it took more than 60 years before linkage groups were found corresponding to all 20 mouse chromosomes in the mid-1970s (EICHER 1981; DAVISSON, RODERICK and DOOLITTLE 1989). Moreover, this map was tedious to apply in practice because at most a few visible markers could be used simultaneously in a cross.

The situation was transformed by the recognition that minor variations in DNA sequence provide a virtually inexhaustible supply of genetic markers that can be used to follow inheritance (BOTSTEIN *et al.* 1980). At the time, such variations could be most conveniently detected as restriction fragment length polymorphisms (RFLPs). In the mouse, the RFLP approach proved to be extremely powerful in interspecies crosses (ROBERT *et al.* 1985; AVNER *et al.* 1988). Comparing the laboratory mouse *Mus musculus* and the exotic species *Mus spretus*, a typical DNA probe had greater than 90% probability of detecting an RFLP with only a handful of enzymes. Using such interspecific crosses, detailed genetic maps have been constructed showing the positions of hundreds of genes (BUCHBERG *et al.* 1989; KINGSLEY, JENKINS and COPELAND 1989; CECI *et al.* 1989, 1990a,b; JUSTICE *et al.* 1990a,b; SIRACUSA *et al.* 1990; BAHARY *et al.* 1991; COPELAND and JENKINS 1991).

Notwithstanding the great utility of RFLPs, they still have several major limitations. (1) The rate of polymophism is considerably lower among inbred laboratory strains, making it difficult to type crosses between such strains. (2) Typing RFLPs is time-consuming and difficult to automate. (3) Disseminating RFLPs involves managing and distributing large numbers of DNA probes. The first limitation is especially serious. Although interspecies crosses are quite useful, there are many circumstances in which it is preferable to use crosses between two inbred laboratory strains—including mapping of many mutations whose phenotypes are affected by genetic background, mapping of modifier genes, and mapping of polygenic factors underlying physiological differences between strains. Ideally, crosses should be designed according to phe-

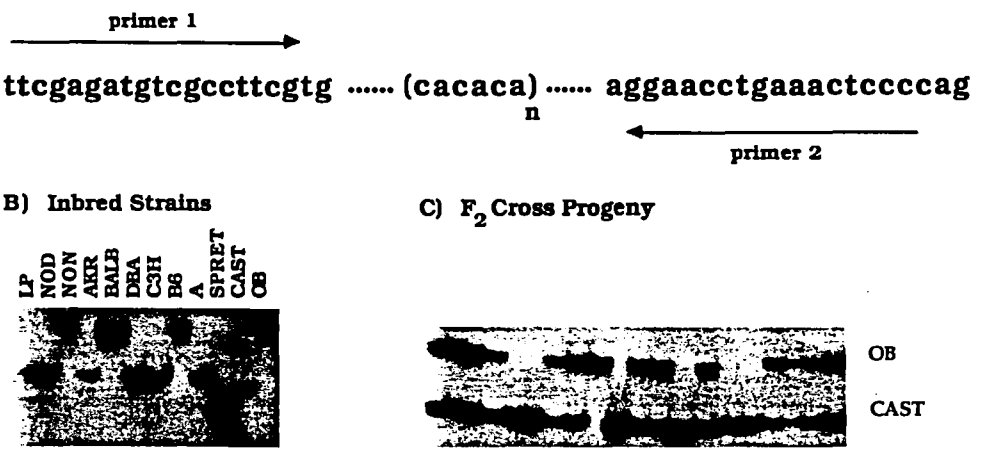


FIGURE 1.—Illustration of a simple sequence length polymorphism, D3Mit21. (A) Diagram of PCR primers flanking region containing CA-repeat; (B) characterization of SSLP alleles in 12 inbred strains (left to right: LP/J, NOD/MrkTacBr, NON/Lt, AKR/J, BALB/cJ, DBA/2J, C3H/HeJ, C57BL/6J, A/J, SPRET/Ei, CAST/Ei, and C57BL/6J-ob/ob), showing four distinct allele sizes; (C) segregation of SSLP alleles in 21 progeny from the OB × CAST intercross used for constructing the genetic map.

notypic and biological considerations, rather than to maximize polymorphism for mapping.

An alternative source of DNA polymorphism has recently been described (WEBER and MAY 1989), based on variation in the length of simple sequence repeats (SSRs) (also called microsatellites) that occur frequently in most eukaryotic genomes (HAMADA, PETRINO and TAKUNAGA 1982; STALLINGS *et al.* 1991). Such simple sequence length polymorphisms (SSLPs) can be easily typed by using the polymerase chain reaction (PCR) with primers flanking the SSR (Figure 1). Recent studies show that SSRs show extraordinarily high rates of polymorphism in both human (WEBER and MAY 1989; WEBER 1990) and mouse (LOVE *et al.* 1990; CORNALL *et al.* 1991; HEARNE *et al.* 1991). Moreover, the typing of SSRs is rapid and automatable and the genetic markers are easily disseminated simply by publishing the primer sequences.

Here, we report the construction of a complete genetic linkage map of the mouse consisting entirely of SSLPs. The map contains 317 SSLPs at an average spacing of 4.3 cM and is detectably linked to about 99% of the mouse genome. In typical crosses between inbred laboratory strains, about 50% of the markers are polymorphic, making it straightforward to follow inheritance throughout the genome in almost any cross.

MATERIALS AND METHODS

Overview: Briefly, the map was constructed as follows. (1) Random clones containing SSRs (specifically, $(CA)_n \cdot (GT)_n$ repeats) were isolated from an M13 library of mouse genomic DNA containing small inserts and their DNA sequences were determined. (2) Public computer databanks were searched to find the sequence of known genes containing SSRs. (3) From each such DNA sequence, PCR primers were selected using a computer program to generate assays designed to work under a single uniform set of experimental conditions. (4) Each PCR assay was tested to determine whether it revealed an SSLP between the two parental strains, OB and CAST (see below), used in the mapping cross and, if so, to measure the allele sizes generated in each of 12 inbred strains. (5) To construct a genetic linkage map, the SSLPs were used to genotype the progeny of an F_2 intercross between OB and CAST and linkage analysis was carried out using a computer program. (6) The newly generated map was then anchored relative to the existing genetic map by two methods: those markers that were polymorphic between the strains C57BL/6J and DBA/2J were typed in the BXD recombinant inbred lines so as to compare them with known strain distribution patterns and those markers that were chosen from the DNA sequences of genes with known chromosomal positions were assigned accordingly.

Isolation of clones containing simple sequence repeats: Random genomic libraries were constructed by digesting male C57BL/6J DNA to completion with *Mbo*I (New England Biolabs), fractionating the DNA on a 4% NuSieve GTG Agarose gel (FMC Bioproducts), and cloning the fragments in the size range 250–500 bp into the *Bam*HI site of M13 mp19 (Boehringer Mannheim). (The use of male DNA was inadvertent; we had intended to use female DNA so that the *X* chromosome would have been equimolar with the autosomes, rather than half-molar.) The libraries were plated at low density of about 500 plaques per 150-mm plate so that individual clones could be picked without the need for secondary purification. Duplicate plaque lifts (Colony/Plaque Screen, Du Pont) were prepared, simultaneously hybridized with end-labeled $(CA)_{15}$ and $(GT)_{15}$ oligonucleotides (T4 polynucleotide kinase, New England Biolabs; [γ-^{32}P]ATP, 5000 Ci/mmol, New England Nuclear) at 65° in hybridization solution as described by CHURCH and GILBERT (1984) and washed in 0.1 × SSC/0.1% SDS at 65° four times for 5 min each. We screened for $(CA)_n \cdot (GT)_n$ repeats because they are the most frequent simple sequence repeat in the mouse genome (HAMADA, PETRINO and TAKUNAGA 1982; J. SEGRE, personal communication). Strongly hybridizing plaques were picked into 1 ml Luria broth (LB).

Length screen of clones: Clones were screened prior to sequencing to determine the length of the insert. Using 5 μl of the supernatant from the plaque picked into LB, phage DNA was amplified in a 50-μl PCR reaction (Amplitaq DNA polymerase, Perkin Elmer Cetus) with the primers flanking the M13 cloning site (5′-TGTAAAACGACGCGGAGT-3′ and 5′-CAGGAAACAGCTATGACC-3′). Phage containing inserts greater than 500 bp were discarded, because they could not be sequenced in a single pass.

Sequencing: Phage DNA was prepared essentially as described (SAMBROOK, FRITSCH and MANIATIS 1989) and the DNA sequencing was carried out according to Applied Biosystem's Taq Cycle Sequencing protocol using an ABI 373A DNA sequencing apparatus. DNA sequences containing SSRs with at least 10 repeat units were used in subsequent steps.

Database searches: GenBank was searched to find DNA sequences containing SSRs, using a variety of computer programs including FASTN and BLAST (ALTSCHUL *et al.* 1990). Specifically, we searched for all occurrences of at least 10 repeats of a dimer, trimer or tetramer.

PCR primer selection: PCR primers flanking the SSRs were selected, using a computer program called PRIMER (M. J. DALY, S. E. LINCOLN and E. S. LANDER, unpublished). The primers were chosen to have a target melting temperature of 60° (BRESLAUER *et al.* 1986; RYCHLIK and RHOADS 1989) and a target length of 20 bases. In addition, primer pairs were chosen to avoid significant homology to one another or to the murine repeat elements L1, B1 and B2 (KRAYEV *et al.* 1980, 1982; LOEB *et al.* 1986). Primer pairs were tested under a single set of PCR conditions; the use of the computer program greatly increased our success in creating PCR assays that satisfied this rigorous requirement. PCR primers were obtained commercially (RESEARCH GENETICS, HUNTSVILLE, ALABAMA).

Mapping cross, recombinant inbred panel and mice: PCR assays were first tested to determine whether they revealed SSLPs between a C57BL/6J-*ob*/*ob* (OB), a congenic line carrying the recessive obese mutation, and an inbred strain of *M. musculus castaneus* (CAST/Ei). If so, allele sizes were determined in female DNA from 12 inbred strains: OB, CAST, C57BL/6J, SPRET/Ei, DBA/2J, A/J, C^3H/HeJ, BALB/cJ, AKR/J, LP/J, NOD/MrkTacBr and NON/Lt. To construct the genetic map, the assays revealing polymorphism between OB and CAST were then genotyped in 46 non-obese F_2 progeny of an OB × CAST cross; this mapping panel provides 92 informative meioses corresponding to about 1 crossover per 1.1 cM. To anchor the map using recombinant inbred (RI) strains, the BXD RI lines 2, 5, 6, 8, 9, 11, 12, 13, 14, 15, 16, 18, 19, 20, 21, 22, 23, 24, 25, 27, 28 and 29 were used. (The remaining four BXD RI lines were omitted in the interests of streamlining procedures: the 22 strains used together with the two parental controls correspond to one-quarter of a microtiter plate and thus permit four markers to be genotyped per microtiter plate.) All DNA was prepared according to standard protocols (SAMBROOK, FRITSCH and MANIATIS 1989).

Genotyping by PCR: To genotype F_2 progeny for SSR polymorphisms, PCR reactions were performed with radioactively labeled primer and products were visualized on acrylamide gels. Primers were end-labeled with [γ-^{32}P]ATP (specific activity 6000 Ci/mmol, Du Pont/NEN) using T4 kinase (NEB) according to standard protocols (SAMBROOK, FRITSCH and MANIATIS 1989). A 20-ng aliquot of genomic DNA was amplified in a 10-μl PCR reaction using AmpliTaq DNA polymerase (Perkin Elmer Cetus) according to manufacturer's specifications. The primer concentrations were: 100 nM of each of the two primers unlabeled and 20 nM of one primer end-labeled. The reactions were overlaid with 40 μl of light mineral oil. Reactions were amplified on either an MJ Research PTC96 Thermal Cycler (MJ Research) or the GeneMachine 2 (USA Scientific Products) using the following thermocycling protocol: initial denaturation at 94° for 3 min, followed by 25 cycles of 94° for 1 min, 55° for 2 min and 72° for 3 min. (Recently, we have successfully used an alternative amplification protocol which may yield cleaner results for some markers: initial denaturation at 94° for 3 min, followed by 25 cycles of 94° for 15 sec, 55° for 2 min and 72° for 2 min, and finally followed by a single cycle of 72° for 7 min.) PCR products were diluted twofold with loading buffer consisting of xylene cyanol and bromophenol blue dyes in 100% formamide, denatured for 5 min on a 100° heating block and electrophoresed on 7% denaturing polyacrylamide gels (SequaGel, National Diagnostics) for 3 hr at 20 V/cm (120 W). Gels were wrapped in Saran Wrap (Dow Chemical) and exposed directly to film for 4–16 hr at −80°. Autoradiographs were independently scored twice.

Streamlining of genotype analysis: Considerable attention was devoted to streamlining the procedures, so that a single person could process some 800 PCR samples each day. Reactions were set up in flexible 96-well plates (Becton Dickinson Labware) using a Biomek 1000 Workstation (Beckman Instruments). In some cases, we coamplified two SSLPs known to yield substantially different product sizes in the same reaction, thereby increasing efficiency. In other cases, we combined two SSLPs that gave substantially different products sizes after amplification but before gel analysis. Loading of the gels was streamlined by using an array of 12 10-μl syringes (Hamilton, Reno, Nevada) spaced to fit into 96-well microtiter plates. The gel combs were handmade sharkstooth combs designed so that the syringe array loaded every other well (G. CHURCH, personal communication).

Linkage analysis: Linkage analysis was performed using the MAPMAKER computer package, essentially as described (LANDER *et al.* 1987; LINCOLN and LANDER 1987; DONIS-KELLER *et al.* 1987; CHANG *et al.* 1988). Markers were assigned into linkage groups based on pairwise LOD scores of at least 5.0. For each linkage group, a "framework" map was constructed consisting of a subset of markers that could be ordered with a LOD score of at least 3.0. Some 66% of the markers easily fell into framework maps. The remaining markers were then mapped relative to the framework maps. Some 92% of the markers could be ordered with a LOD score of at least 2.0.

Error checking: To maximize the accuracy of our data, we developed a new mathematical method for identifying potentially erroneous genotypes. Briefly, the approach is as follows. Rather than assuming that the observed data represents the true genotype, we considered it a phenotype caused by the genotype, according to a penetrance function: phenotype reflected the true genotype with probability $1 - \epsilon$, but differed from it (*i.e.*, was erroneous) with probability ϵ. Genetic linkage analysis was then carried out under this assumption, which explicitly allows for the possibility of error throughout the data. For each typing (*i.e.*, each observation of an individual at a locus), we calculated under this model the LOD score, $\text{LOD}_{\text{error}} = \log_{10}(P_{\text{error}}/P_{\text{correct}})$, where P_{error} is the probability of the overall data set arising if the given typing is erroneous and P_{correct} is the probability of the overall data set arising if the given typing is correct. For the most part, the potential errors correspond to apparent double crossovers in a relatively small region and instances in which a single crossover apparently occurs in a small interval rather than in a much larger adjacent interval. For LOD scores ≥ 1.0, the autoradiograms were independently reread and, if there was any ambiguity, the typing was repeated. In our analysis, we used a value of $\epsilon = 0.007$ based on empirical studies of our error rate (see RESULTS). This method will be described in more detail elsewhere (E. S. LANDER and S. E. LINCOLN, in preparation).

Recombinant inbred analysis: Data from the recombinant inbred strains were analyzed with the RI Manager

computer program (MANLEY and ELLIOT 1991) using the "find" function to detect linkage.

Mathematical analysis of distribution of interval sizes: To test whether the genetic markers were randomly distributed in the genome, we examined the observed distribution L_{obs} of distances between adjacent markers and compared it to the expected distribution L_{exp} under the assumption of random distribution of markers. We calculated the distribution L_{exp} as follows. For a map with an average spacing of d cM and a cross with n informative meioses, the probability $P_{d,n}(k)$ that two adjacent markers will recombine in exactly k meioses was calculated as:

$$p_{d,n}(k) = \int_0^\infty \left[\binom{n}{k} \theta(x)^k (1 - \theta(x))^{n-k}\right] \frac{e^{-x/d}}{d}\, dx$$

where $\theta(x)$ is an appropriate mapping function. We used Kosambi's mapping function for this calculation. (Although no simple mapping function perfectly fits the recombinational data from the mouse, the choice is adequate for the purpose inasmuch as the same mapping function was used in the construction of the map.) To understand the equation, observe that the last term in the integral is the probability density that the two adjacent markers lie at a distance of x cM apart while the preceding bracketed term is the probability that two markers at x cM will recombine in k of n meioses. Here, we have an average spacing of $d = 4.3$ cM and the 46 animal F_2 intercross provides $n = 92$ informative meioses.

Nomenclature: Loci defined by SSLPs are named according to standard convention–*e.g.*, *D1Mit7* refers to a locus on chromosome *1* isolated at the MIT Center for Genome Research, with arbitrary reference number 7. We have used this nomenclature both for SSRs in anonymous sequence and also for SSRs occurring within known gene sequences. By distinguishing between a gene and a particular SSR within the gene, the nomenclature remains unambiguous even for situations in which a single gene contains multiple SSRs, as happens in a number of cases. J. TODD has concurred in this decision and has assigned such designation to SSLPs in genes previously published by his group; these names are given in the tables. Similarly, a single SSR might be studied with various different PCR assays. To avoid ambiguity, we also distinguish between the SSR locus (locus name) and the particular PCR assay (assay name) used to study the locus. This is especially useful in the case of the six SSLPs which were independently identified twice (see below). Several SSRs previously published by J. TODD were renamed, with his permission, based on newly determined or revised chromosomal location: *D0Nds25* was renamed *D2Nds2*, *D8Nds1* was renamed *D4Nds10*, *D0Nds27* was renamed *D6Nds4*, *D0Nds22* was renamed *D10Nds3*, *D4Nds1* was renamed *D6Nds5*, *D0Nds19* was renamed *D12Nds1*, and *D1Nds3* was renamed *D15Nds2*.

Finally, we refer for simplicity to laboratory mouse as *M. musculus*, although these strains represent a combination of genomes from *M. musculus* and *Mus domesticus*.

RESULTS

Screen for polymorphism: Primer pairs flanking SSRs were first tested to determine whether they revealed polymorphism between OB and CAST, the strains used for genetic mapping. These strains were chosen because they belong to different subspecies and thus were likely to show a high rate of polymorphism, but they are sufficiently closely related that F_1 progeny of both sexes are fertile (unlike hybrids with the distinct species *M. spretus*, in which males are sterile). This allowed us to use an F_2 intercross rather than a backcross for genetic mapping–providing twice as many informative meioses per progeny.

Overall, we designed and tested 455 primer pairs, with 394 obtained from sequencing random clones containing CA- or GT-repeats and 61 obtained from searching GenBank for SSRs. Of these, 393 (86%) produced working PCR products of the expected size under the single uniform set of PCR conditions employed. This success rate increased steadily over the course of the project as the PRIMER program was refined, so that the success rate near the end of the project exceeded 90%. Of these 393 assays, 303 (77%) yielded SSLPs between OB and CAST. To this collection, we added 34 SSLPs previously described by J. TODD and colleagues (LOVE *et al.* 1990; CORNALL *et al.* 1991; HEARNE *et al.* 1991) for a total of 337 SSLPs. Of these, 18 produced patterns that we found difficult to interpret reliably. The remaining 319 produced easily scored polymorphisms (accompanied, in some cases, by background bands). These 319 SSLPs were used for genetic mapping; the primers are listed in Table 1.

To facilitate the use of these markers in other crosses, we determined the allele sizes in twelve commonly used inbred laboratory strains (Table 2). The typical rate of polymorphism between an inbred laboratory strain and either *M. musculus castaneus* or *M. spretus* was about 90% and, more remarkably, the typical rate of polymorphism between inbred laboratory strains was about 50% (Table 3). For relatively short CA-repeats, the length of the SSR is known to be correlated with its rate of polymorphism in humans (WEBER 1990); we saw no such correlation in our data, however, probably because the vast majority of the SSRs used were very long (85% had more than 15 repeats).

Genetic map construction: To construct the genetic linkage map, we typed the 319 SSLP markers in 46 progeny from an OB × CAST F_2 intercross. The primary genetic data is available by request from the authors. Based on linkage analysis, 317 of the 319 markers fell into 20 linkage groups. These markers defined a genetic map of the mouse genome, with an average spacing of about 4.3 cM (Figure 2). The remaining two loci show no significant linkage to other markers in the map; these markers were retyped several times to confirm the data but no errors were found. We estimate that the map is linked to some 99% of the mouse genome.

Error checking: Given the large size of our data set (nearly 15,000 genotypes), some errors are bound to occur. Such errors pose problems for the construction of dense genetic maps: they spuriously inflate appar-

TABLE 1

Primer sequences for simple sequence repeats

Locus name	Gene name	Assay name	Left primer	Right primer
D1Mit1		L33	GATCCTCAGATTGAAGAATC	GAGCCACCAGAGATGTAAGA
D1Mit2		A26	TTGAATTCAAACATCATCAGGC	CTATCTGTAACCCCAGCTCCC
D1Mit3		M253	TTTTTGTTTTCTTTTCTTTTCCC	CCCTCTTCTGGTTTCCACAT
D1Mit4		M46	GCTACTGCTTTGGAGTCAGT	ATGACTTGAGCTCAGTCTCTG
D1Mit5		L20	AGATAGCAGAGCCTGAGCCA	CCTGAACTCCACCATTTAGC
D1Mit7		A80	TGGTAGAGGAAGGTGCACG	GCAGGGGAGTAGTACCACCA
D1Mit8		L31	CTGAAAATCGTCCCTTGACC	CAGGAGCATGAAATGGGGAT
D1Mit9		M111	AACTGCAGGCTAGAGACCCA	ATGTGCACATACCAAAGGCA
D1Mit10		A117	AAACCATGCAGGTACTGATATGG	GAAGAAATTAACTGAGAGCAAGGC
D1Mit11		M17	GATCAGATTAAGATGTATATTATAA	GAACCCCAAAAAGAAATCTG
D1Mit12		M93	ACCATATCTCTACATGCTTGTGC	GCATTTGGTTTATTTTTCCACG
D1Mit13		L30	TGATGCTTGCACGTTGAGAT	AAAACTGGTTCCTGGTTCCC
D1Mit14		M193	GCCAGACAGGGCTACATTGT	AGACTGAACTCTGGCCTCCA
D1Mit15		M146	TCCACAGAACTGTCCCTCAA	ATACACTCACACCACCCCGT
D1Mit16		L46	AGAGTTAGCTGCCTAGCTTGAGTG	TGGAAAGATCTAGGGTTGTCAAAA
D1Mit17		M41	GTGTCTGCCTTTGCACCTTT	CTGCTGTCTTTCCATCCACA
D1Mit18		A77	TCTGGTTCCAGGCTTGATTC	TCACAAGTGAGGCTCCAGG
D1Mit19		L86	GATCCCAGCCAATAGAAGTACA	GAAAGGTTTCCTATCCTATGGC
D1Nds2		T17	ACATATATGGACTACATACATAC	AGACACATACAACATAGAATTGTT
D2Mit1		M128	CTTTTTCGTATGTGGTGGGG	AACATTGGGCCTCTATGCAC
D2Mit2		M112	AGTCCTCCTTGGACTTCCATTAG	TGGATTATATTTTCAAGACCAGA
D2Mit3		M116	GGGTATCTTCATGCCAGTGG	GGTGAGGACACGAGGCTATG
D2Mit4		M52	ACACCAACCCAAGCAATTGT	GAGCACGGAACAGGCATAAC
D2Mit5		A41	CCGGGGATCATCTTAGGACT	CCCCCTCTACACACTTGCAT
D2Mit6		L18	AACAAACAAACCCCTTGCCC	CTCTAACACAGCCCCAGGTG
D2Mit7		L44	AAGGCAAGCATTCTGCCACT	CTCCCGGCAAGAACTGTTTT
D2Mit8		M199	CTTCATTGCCAATGCTCTCA	TGAAGGTGAAAACAGAGGCA
D2Mit9		M85	GTCTGCACTCTCACCAGCAA	CAGCTTGAAATGCCTTTGAG
D2Mit10		M39	CATCAGGAAACACAGGACCA	ACCTAACCCTAATGATGGGG
D2Mit11		M134	CAAACCCCCAGCTCTCTCTT	CCATACCCAGGCTCCATCTA
D2Mit12		M179	CTACTTCCCAGGTGCTTGGA	TCCAAAGAACTGAATGGACA
D2Mit13		M130	AGGACCAGAAGGACAGACCTATC	AAATCTCACAGGGAAAAAAAAGC
D2Mit14		M163	CAACACAAATGTGAGCATGC	TGCTCACCTGCCTAGTACTT
D2Mit15		A61	ATGCCTTAGAAGAATTTGTTCCC	CTTGAAAAACACATCAAAATCTGC
D2Mit16		M186	CTTGGATAGGGATGGAGGGT	ACAGACATGTAGGCACACCA
D2Mit17		M246	AGGCAATTACAAGGCCTGG	CACCCATCTCCCTCAGTCAT
D2Mit19		A83	TGCTAAAAGTCTGGCATTTGG	CAAATGTTTGTCTTTCAAAAGCC
D2Mit21		M184	GGCTTAGGCCCAAATTTTCT	TGGAAAGCTCATCTCTTCCT
D2Mit22		M167	GCTCCCTTTCCTCTTGAACC	GGGCCCTTATTCTATCTCCC
D2Mit24		M75	ACTTGGCTTACAGGGGACCT	TACCAGTCCCTTTCCACCTG
D2Mit25		A67	TATGCCACTCAGAAGAGGTCG	ATATGTGCATTGCATGAACTCC
D2Mit26		M37	TGTTCTTTGCTCATCCACCA	AGGCTGATGGTAACAGTGGG
D2Mit27		M106	AGGCTAAGCCTTGCATCAAA	GTCGCAAAATGTGGATGATG
D2Mit28	*Snap*	D25	TGCTTCTCTCATGGTATTACCTAG	ACAAACCACACAGACATTTACAAT
D2Mit29	*Svp-4*	D115	CGGTGACGAAGCTTCTGAG	CTTTGAATATGAACTCTCACCTTCC
D2Mit30	*Trh-1*	D111	CATCCAAGCAGTAACGTAGACG	AAATGTTACACCCTCTGCGG
D2Nds1		T19	CTGCATGCACTGTATTTGTAT	ACTCATGGGTTGTGCATATGG
D2Nds2		T57	AACATTGAGGACATTTGGTGA	CGCTGTATGCATCCTTAAGAA
D2Nds3	*Il-1B*	T15	CCAAGCTTCCTTGTGCAAGTA	AAGCCCAAAGTCCATCAGTGG
D3Mit1		M28	TGTGCACAGGGGTACATACA	TCATTTTCTTCCTCCCCCTC
D3Mit3		M250	CCTTTTGAGGCAAAGCTCC	CTAAGTCCTGCACCTGCCTC
D3Mit4		L40	TGTGCCTGCAAGTTGTTCTT	CTACAGTGGGGGCAGAAGGT
D3Mit5		M123	AGCCCCTTCCAAGTGTCTCT	GGTTTCGGAATGAGATGAGC
D3Mit6		M149	AACTTCAACATGTGAGGGGC	CCTGAAACAAAGCAACAGCA
D3Mit7		M74	ATGCAACTAACTTTATTGAAAATC	TACAATTATCCGGGAGCTA
D3Mit9		A85	AACTTCATTTGCTTGGAAACTACC	TGTTTTATATTGCCCTGTATGTGC
D3Mit10		A34	CTGGCTTGGTGGAAGTCCT	CCTAAGCCAGCTACCACCAC
D3Mit11		L38	CCAACCACAGTAACACATGT	TGGAGACCAATGCGAACAAC
D3Mit12		A60	TAGACCAATCTTGGGAGTGTCC	GGAAAAGCATAAGAAACAACCG
D3Mit13		L37	TTTCTGCATTATGTGGGCTT	AACCACAGATGACAATTGAA
D3Mit13		L8	CCTTTCTGATTATGTGGGCT	CCACTGAAGGATAACCACAG
D3Mit14		M206	ATTGCGGTTAAAGTTTGCTT	TCCTGCAAATTGTCCTCTGA

TABLE 1—Continued

Locus name	Gene name	Assay name	Left primer	Right primer
D3Mit15		A55	AATTTGCATTCCAGGACCAC	AGGAAGTGACGTTGGGTTTG
D3Mit16		M159	TGCTTGTCCTGTGTTAATGA	TGAGAATGGAGGTGAACAGC
D3Mit17		M235	CATGGCTCCATGGTTCTTG	CCACGGAGAACAACTGAAGA
D3Mit18		A96	GAACAGTTCCCAGGTCCTCA	CTGCCTTTAAATTCTGTCACCC
D3Mit19		M141	CAGCCAGAGAGGAGCTGTCT	GAACATTGGGGTGTTTGCTT
D3Mit21	*Il-2*	D31	AAGCTCTACAGCGGAAGCAC	CTGGGGAGTTTCAGGTTCCT
D3Mit22	*Rp132-ps*	D122	AAGGATTGAAGAATGGTTGGG	AATCAGCGATTTCAGCACG
D3Nds2		T21	ACACATTGGAGATGCACAGCG	TCTGCATGCCAGGGTTGTGAT
D4Mit1		A73	ATGATGTACACTTAGGCATTGCA	AGAAATATGGCAAGCAAAATGG
D4Mit2		L67	GCACTCACACACTCACATGC	TGCACCAGTGACTTTACCCC
D4Mit2		L6	GGATTTCTTGGGCACTCACA	GCACCAGTGACTTTACCCCA
D4Mit4		M31	CGGAATAGGCAGCTATGCTC	TCCATAGACCCTGCATGTGA
D4Mit5		A1	CGCCTCTGTCTCTACCTCTCA	CCTAAAAAGTGTCTTCTGACCTCC
D4Mit6		M64	TGTGGGCAGTGTAAGCACTC	CTTTCCTCTGTGCTCGTGTG
D4Mit7		A71	CCGGGGATCATGTTTAGAGA	AGAGGGATAATTTTTGAATTGCC
D4Mit9		M241	GGCTTTGGAATGCTATGCAT	TGGCAGGAGGTATGACAGAA
D4Mit11		M8	GGTTCACCAAAGGACTTCGA	CCTGTGACCCCTTGGAAGTA
D4Mit12		M15	GCTTGCTTTAGGAGTGTGCC	TATTTGCTCTCCATTTCCCC
D4Mit13		M169	GCTGGTAGCTGGCTTTTCTC	CAGATGTTCCTACTGCTTGG
D4Mit14		A69	TACAATAGTTAGCTCAGGCCAGC	GGGGTGAGGAGAGTGACTCA
D4Mit15		A122	AGGAATACTGAATGTGGACTTTCC	TCCCTTGATTAACAGAAGACCTG
D4Mit16		A65	GATCACCCAAGGCTGGC	TCCCCGTGAACTTCCATC
D4Mit17	*Orm-1*	D1	TGGCCAACCTCTGTGCTTCC	ACAGTTGTCCTCTGACATCC
D4Mit205		M205	TGTGTGAACATGTCTACCCC	GGGGACCGAAGTAACAGTGA
D4Wsm1	*Ifa*	F1	TCAGTATGTACATCCATGCC	TAAAAATGATAAGTTGTTTTATGAA
D4Nds2		T24	CTTCTGTCTGCTGAGGATACC	CCATGATGAGCCAAAATGAAT
D4Nds10		T29	TGTAAGCCATTCTAATAGATC	GAGGGAATAGAACTGACTGGT
D5Mit1		A82	AATAAAGCTGTGAGGTAAACCCC	GAAACAAATGATTGTTTTGAGCC
D5Mit3		M197	AAGGGCAAGCCATTTAAGGT	GCCCCAATCTAGGAGGCTAC
D5Mit4		M189	CTAGTCATTGGCTCCAAGGG	ATGCACTGGGAGAGTGAAGG
D5Mit5		A11	TGAGTGAGGTGTGGTGATAACC	TGTGTCTTCCCCTTTCAACC
D5Mit6		L42	CTCCAAATGGAACTATGGAA	CATGATATTAAGCAGCTGTG
D5Mit7		M154	AAAGGGGGTCTTCTTTGGAA	TCTCCTGTAGTGGGTGGTTT
D5Mit9		A9	TTCCTAGCATTTCCCTGGG	ATCTGGAGAGAATTGTAGTCTGGG
D5Mit10		M207	CGAGAAGTTGGAAAGACCCA	GGCACCCATGCCTCTATG
D5Mit11		M97	GATCTTCCTACCTTCTTACCCAC	CATGATTTTATTTGGGGGG
D5Mit12	*Csnb*	D128	TTAGGCAAGTGTTAGACTAAAAGGG	GGAAATCCTCTTAGACCTTAAATGC
D5Nds2		T26	TAATCTATTGTTTGTGGAAAG	GTATCAGGCAAACTGGACC
D5Nds4	*Afp*	T61	AGCAGGGCTACACAGAGAAAC	ATTCCCATATTTGCATCTCCA
D6Mit1		A10	GGCACATTTGCCTTTGTTTT	TCTCCTATCTCTCCACCTTTTCC
D6Mit3		L59	ATGGGTACCACCCTATCATACCTA	TTATACACTGATATCTTGATAGCC
D6Mit4		M239	ACTAGGAAACACACTGATTCATATG	GAGGTGACAAAATTTTCAAAAA
D6Mit5		M161	CACGGAGAGGACCTACATGC	AGCTGCTCGTCTCCACACTT
D6Mit6		M259	TTCTCTCAGTCTTGTCTGTGTACA	GTGAGGCTCAAAGAAAGGGC
D6Mit6		M227	GAGGCTCAAAGAAAGGGCTT	TTCTCTCAGTCTTGTCTGTGTACA
D6Mit8		M240	TGCACAGCAGCTCATTCTCT	GGAAGGAAGGAGTGGGGTAG
D6Mit9		L23	GTCTGTTTTGGCATATGGCA	TCTGGGTANCCAACCATGTT
D6Mit10		M78	TCAGAGGAACAAAGCAGCAT	CCTGTGGCTAACAGGTAAAA
D6Mit11		M170	ACTGGCCTCTTTTATGTGCA	TGTGAGTGTGAGTTCAGGGG
D6Mit12		M11	CCACATCCATGTAAAAGCTG	TGGTTCAATGAAAGTTGCCA
D6Mit13	*Prp*	D34	TTTTGTTTCCTTTCAGCATG	GGGAGCCATTGTCCTATTCA
D6Mit14		M190	ATGCAGAAACATGAGTGGGG	CACAAGGCCTGATGACCTCT
D6Mit15		M148	CACTGACCCTAGCACAGCAG	TCCTGGCTTCCACAGGTACT
D6Mit16	*Ly2*	D11	AGGCTTTGATGCTGTATAGG	CACCAGGAACGTAAGTGAGC
D6Rck1	*Cpa*	F3	CAGCTGAGTCATTAGAGCACTTACC	CTCAGACCTACTAGAGAAGTGCAGAGC
D6Rck2	*Mirp*	F2	GAACACCCCTGGACCGTATTCTCA	GATCGCTGGACACTTCTCTGAGTG
D6Rck3		F103	GACAAGAGGACGCATCTTTTG	CTACGAAAAGTCAACCTCGAGG
D6Nds4		T59	ACCTCAGCGGTTCTTTATGAG	TGGTCCACCCTGAATGAGTCC
D6Nds5		T23	GGAATGTCTTATTTAAGTCAG	AGTGGAGTAATATTTGAACAA
D7Mit1		M208	GTCCCAGTGTGTATATATAATCCAG	GGATTATACACACAGATGTTGGG
D7Mit5		M187	TCGTGTCAAATTGCTTATGC	ACTGTGTGTGCCTGTGTTTG
D7Mit7		L12	ACTCAAAGGTTGTCCTGGCA	TGGTAGTGGTGGCTNCGGTG
D7Mit8		M183	TTGGCCTTTATAGGCACCTG	TAAGGCACCATGATATGGCA
D7Mit9		A89	GACAGGTGGTTCTTTAATAATCCG	GGAGCTTTAAAGGACAATTTTCA
D7Mit10		L72	GTTGTTCGGGAAGGGAAGAT	CCTTGGCACGAGATGAACTG

TABLE 1–Continued

Locus name	Gene name	Assay name	Left primer	Right primer
D7Mit10		L25	GAAGATTGGGCTGTCTGCAC	TGAAGCTGATGGAGCTGATG
D7Mit12		M23	GCTGGGTTTATTCATTGCAA	TCCAGCTCATGGGTAGAAGA
D7Mit13		A113	ATGGGGAAAGTGACTGAGGA	ATTTTTGTAGCTTGAAGGTATGGC
D7Mit14		L79	TCCCTCCTCATGTTTTCATG	GATGATTGGGAGAAGCAAGG
D7Mit15		M47	GTGTGCACCCACATGGATAC	AGGGAAAGCACTTGACCATG
D7Mit16		A13	CTGGTCTCTGTCCTTGGAGC	AAAGAAAATATTCTTGTTGCCAGC
D7Mit17		M91	CTGGCATTTATGTTGCTTCA	AACTTGCCTTCTGTCCTCCA
D7Mit18	*Gas-2*	D117	GGGAGCCCAGCTTCTACTG	TCCTAACACCCTTCCTGGTG
D7Mit19	*Tyr*	D108	GCTGCAGCTCTCTCTGGG	GATGGCTCTGATACAGCAAGC
D7Mit20	*Mb-1*	D103	GTGTAGCAATGGTGTCGGTG	AAGCCTGCCTCCAGATGTAA
D7Nds1		T27	GAGATCTTCCATACTCATATT	TAGATAGTGTTAACAGTGACC
D7Nds2		T28	CAGACTTTCATTTCTTTGGATAC	ATGCCATCATGTGTTGAAGCA
D7Nds4	*Int-2*	T63	GTGACAATACATTCCTGCTGT	CTCAGATCTTATCTCTAGCAC
D7Nds5	*Ngfg*	T62	CTCCACATGTGTATGTGTATG	ATGGAGGCCGAAGAAAGAATC
D8Mit1		M70	TTTTGCTGTCTAGGTCCTGACTC	CAGCCTCATTAGTAAGGGACCTT
D8Mit3		M195	TCCCATTCTCGCATAAGTCC	GATGGGAAGACAGGGTAGCA
D8Mit4		M71	CCAACTCATCCCCAAAGGTA	GTATGTTCAAGGCTGGGCAT
D8Mit5		M176	TCCCTTTTCCCTGTGCTATG	GCCGTTCATTTAACCCTTCA
D8Mit6		M158	CAGGCAGCTTGCTAGGACTT	TACTGCCTTTAGCCCAGTGG
D8Mit7		M138	TTGGTGAACACCAGGTTCAA	ATGATGTTAGTGGTCTGGGG
D8Mit8		M257	GAGGGGCTGGAAGAAAGAAC	AGCCCAGACTGCTTCCTTTT
D8Mit9		A62	ATTTGAATTGTGCAGACCTGG	CTGCTTGTTTTTATCTCCTGGG
D8Mit11		A105	GCAGCAGTGGTAGCAAATAGC	CTTAATCAGCAATCCTTGACACC
D8Mit12		L11	GATCTCTACATCAAAAGGGA	TTCAGTTTTGTTTCTGAAAC
D8Mit13		M77	CCTCTCTCCAGCCCTGTAAG	AACGTTTGTGCTAAGTGGCC
D8Mit14		L34	TTTTCACACTCACGTGTGCG	GTCTCTCCTTCCTGGCGCTG
D8Mit15	*Mt2*	D20	AGCTGAATTTGAGCTAGTCG	AAGCTTACGGTTTAATCCCC
D8Mit16	*Polb*	D100	GCCTGGATTTCCTCATTGAA	AGTTGGTTATCCCTGAAAATATACA
D9Mit1		M88	GAGCTGTAACACTGACAATGTGC	TATCTCAATGCACACTTTTGTGC
D9Mit2		L32	GTGGTCTGCCCTCTTCACAT	CAAAGCCAGTCCAACTCCAA
D9Mit4		M151	TGCTGAGCAAGCTATGAGGA	GACAGCCCATCACAGCTACA
D9Mit6		A78	GTACCCGGGGATCTGGTG	CTGAGAAATGGAAACGTTGTTG
D9Mit8		M211	GATGAAGACAATAAAGAACCTTAAA	AAGAGCTAACCCATTGCTGC
D9Mit9		A72	TACCCGGGGATCTTCTTTCT	AGAGCTTTCCCGCTACACAA
D9Mit10		M86	TAACCAACCCTTCAAGGCAC	AATCCTTGGCTGAAGGGAAT
D9Mit11		L60	GCCTTCATGTGTACCTGAATGCAC	GGCTCTGTAATCACTGAAGCTGGT
D9Mit12		M73	ATTCAAGGGGCAGTACACAT	TGGTCCTGGTAAAACTGCCT
D9Mit14		M236	CCAAAGGACTGCTATTTGCG	GTAATATTGCTACACTCATGCACA
D9Mit15		M160	TTCAGTCCAGTCTGGGGGTA	CCCCCAGTTTTGTTGTTTTG
D9Mit16		A5	TCTGTGCCTCTTGGAGTGTG	AGGATTGGGGCTTTGTTCTT
D9Mit17		L19	GCCAAGGCTGTCTCTTAGCC	GAGAGAAGGGTTCTGGGCAG
D9Mit18		M10	TCACTGTAGCCCAGAGCAGT	CCTGTTGTCAACACCTGATG
D9Mit19		M157	CCAAACACAACCCCTCAGAA	TCATGGCTTCAAGACTGCTT
D9Mit20		L64	CCCTTGCAGCCCATCGCCTA	TAGACACATAGCTGGAGGTTTTCT
D9Mit21	*Cyp1a2*	D15	CAGTCCCTGGTTAATAACAACAAC	TATAGTCCATTGTGGCAGAGGAGT
D9Mit22	*Ncam*	D134	ATTGCATAACACCCCCACAT	CAGTGCTTAACTGCTCAAATGC
D9Mit23	*T3d*	D4	AAGAAGTTTCCATGACATCATGAA	AGAAGAAAATTCTTGACAGCTCTG
D9Mit24	*Trf*	D26	CCTTCTAAACACAGGCTTTTTGAG	CTGATGATCACCTCATTTCCTGAG
D9Nds2		T30	TCCTTGGAGTTAAAACTTGGA	AGATAAATTCAATGAGTCCTA
D10Mit1		M153	GGAGAAAACCAACTCCTGCA	AATGTGAAAATGTGGAGTGG
D10Mit2		M24	CTGCTCACAACCCATTCCTT	GTTCATTTGAGGCACAAGCA
D10Mit3		A114	GTTGATAGTCCCACCTCACTCA	TGAGAAATTCCATCTGTGGC
D10Mit4		M139	TAGGATTACAACCTTGCCCC	CACAAGGGAAAGTCTCCAGC
D10Mit5		M67	AAGTGAAGGTGCTGGTCACC	GGGAATTTCACAAAGACAGC
D10Mit7		L62	GATCTATGTGAGTGCGAGGCTAGC	TCAAACCAGATGGCACTGAAGACT
D10Mit8		M3	AGTGTTAGTGGCTGGGGTTG	TGAACGTTTCAGTTGGTCCA
D10Mit9		A37	ATTTGGAGCACGCATCTTCT	AGGCCCACCTTGTACTTGTG
D10Mit10		M7	CCAGTCTCAAAACAACAACAAAC	TTGCACCTAGATTGCCTGA
D10Mit11		A88	GAGAAGTCACTGGGAGCTGG	TTGCCAGGTTGCTCTTCTTT
D10Mit12		M172	ATGTCCAAAACACCAGCCAG	GGAAGTGATGGAGCTCTGTT
D10Mit13		A63	GATGGAGCTTCTATGTCAACCC	TTATTTCCACTGAACTTCCTTTCC
D10Mit14		M175	AGAGGGGACAAGGAGAGACC	AAGGTTTGGGTTCAGTTCCC
D10Mit15	*Sqr3*	D30	ATGCGTACAGGCAAAACACC	GCTACATTGGTCTGTGACGC
D10Nds1		T31	TGCACACCCACAGCACACATG	AAGGTTTAAGAAGGTCAAATCATA
D10Nds2		T32	CTATTTACTTAACTCACAATT	TGGTCTTTTGCTCCATAAACT
D10Nds3		T54	TGACATTTTGCGATTTTCATTTGT	GACACATGGATCCTCACATGC

TABLE 1—Continued

Locus name	Gene name	Assay name	Left primer	Right primer
D11Mit1		M215	GGGTCTCTGAAGGCTTTGTG	TGAATACAGAAGCCACGGTG
D11Mit2		L14	TCCCAGAGGTCTCCAAGACA	CCACAGTGTGTGATGTCTTC
D11Mit4		A124	CAGTGGGTCATCAGTACAGCA	AAGCCAGCCCAGTCTTCATA
D11Mit5		A2	TTCTGTGAGCCTGGAGGAGT	TACAGGACTAGTTTCCATTTGGG
D11Mit7		M119	AGGGTATTCTCTAGCCTCCACAC	TTTGAGGCAAGATGTCATGTATG
D11Mit8		M212	CTTTTCATGGAGGCACAGGT	TGTGAACAGAGACACACATTCA
D11Mit10		M162	GAACCGCAAGTCATGAATCA	TGGTTTATTCCTGAAGCTGC
D11Mit11		M43	TATTCTCTCCTTCCCCCCAC	TAGAGTTGGGACACCCAAGC
D11Mit12		L3	AGGGTTATGCTCTTGGCTGC	GATTTTCCTAGGCTGGCTGG
D11Mit13	*Ace*	DACE	ATAACACCAACATTACCATAGAGGG	ATACTAAGTTCAGACTTTTCACCAATTTT
D11Mit14	*AntP91a*	D2	CCACTTAGTATATCTTGTCC	GCATGACTTGGCCTATCACC
D11Mit15	*Glut-4*	D5	TGACATTTGGCGGAGCTAAC	ACATGTACTTGCCAGGGTAC
D11Mit16	*Lif*	D133	CAGCTAGAAATGGCAATGAGG	CTTGTTCTACACCCAGCAAGC
D11Nds1		T33	TAAGAACCTTCTGTAGTTATT	ACCTTAGTTAGAGTTGGTCTC
D11Nds7	*Gfap*	T12	AACTGTTCAAAGCCATTTCG	CTATGGACTCACAGCCAGGCT
D11Nds9	*Il-5*	T14	CCTTTCTGAAAGTATTAAGAGT	ACAACCATCTGCATATCCAGC
D12Mit1		M50	TACCCGGGGATCTTTTGTTT	AAGTGGACTGCCAGAGGATG
D12Mit2		M27	ACACAGGCTAAAACATGGGC	GCATCTGTATTCCACAGGCA
D12Mit3		L41	TAAAGGGGTTTGCTTAAACA	ATGCCACTGAATGTCAAATT
D12Mit4		A64	ACATCCCCAGCTCTTGTTTG	AAACCAAACCAAAGAAGCTTAGG
D12Mit5		L58	CACATAGACCAGACAGGCATGCGT	CAAGGTCACGTTGCTAGCTAGGAA
D12Mit6		L16	ATGCTCGACATCAACCTTGG	TATCTGTGTGGCTGGAACGA
D12Mit7		M62	CCGGGGATCTAAAACTACAT	TCTAATCTCAGCCCAATGGT
D12Mit8	*Igh-C*	D7	TTGCCTAACCCACTCACACC	TGGTGACTCCTTACAGAGGC
D12Nds1		T51	AGTGATGTGATTACAGGTTTG	CACTCTATAAACCCACTGCAG
D12Nds2	*Igh-V*	T1	ACATGGTAATTTATGGGCAA	CTGGATACCTGCAATAGTAGA
D12Nds11	*Odc*	T64	CATTTGAGGACAGTCAGGATC	GGAACTTTCATGCAGTACTAG
D13Mit1		A86	TCAACTCTTCTGTAAACCAGATGC	GTCTGTTTGATTCCTGACCTCC
D13Mit3		M79	TCAGGCTCATCCCAGATACC	TTTTGCAGAGAACACACACC
D13Mit4		M231	TGTGGGACAACTGTGACAAA	CACCCAAGGCCCACTTC
D13Mit5		M38	AGAAGCCAGCAGGTGTTTTC	CCAGGAAGTAACCCCAAACA
D13Mit7		A68	CGGTACCCGGGGATCTAC	AGCCCAGCTTGTGAAGTGTT
D13Mit8		M61	GCCCCATTTCTGAAGTTTCA	AATAGACTCTTCAGCCCCCC
D13Mit9		M147	GGGTTCCAGATTGAGTGGAA	TTGCCAAAGTGTCAAAATCA
D13Mit10		L61	AGTCCTGCCATTTGTCCTCTGACC	ATGTCTTAGTCTCACATGCTGGGG
D13Mit11		A91	CATGGCTCCTTTAACCTGTTT	CAATGATTAACCCTTGAAAAAACA
D13Mit13	*Il-9*	D24	CTGTGGTAAGTCCAGATTTG	GGAAAGAGTAGGAAGATGCC
D13Mit14	*Sqr4*	D29	GGAACAGCAAGCTCTAAGGG	CTACCAGGCCTCCCAAGATA
D14Mit1		A103	GATCTATATGTCCCAACTATAAAG	ATTTTGACTAGGATTGTTTGAGGG
D14Mit2		A24	TGTCTGACCCATTGGAATTATG	TGAAGAAGACACCTAACACTGACC
D14Mit3		M32	GCAATTACACCTCCTCGGAG	CACAAGGGCATATGGTACCC
D14Mit4		M228	AGGCACCCCCTCACAGTAC	TTCATTCCTCCTGCTGACCT
D14Mit5		M214	CACATGAACAGAGGGGCAG	GTCATGAAGTGCCCACCTTT
D14Mit6		A119	GACAAACGCTTTCATCTACAAGG	TGTGCACATTCATCCACATG
D14Mit7		L27	AATGTATGGGCATGTGCGTG	GAGATAGTCAACCAAAACAA
D14Mit8		A44	TCACAGGTGCTCTCAGTCATG	GCAAATACTTCCCTTCTTGGG
D14Mit9		A93	AGGGGAAGGGAAGATGAAGA	GGTGTGACCACTGCCTAGGT
D14Nds1	*Plau*	T10	TGCTGGCTAGGAATAAACAGA	AGGGAATTCATGTTCAGGATA
D15Mit1		L29	AACATGGTCCCACAGGTGTC	AGTAGAAGCTGCAGCCCTGG
D15Mit2		L10	AGAGCATGTCCTCACCCCTT	CCTGGAAAGGTCTCAGGGAA
D15Mit3		L78	TTTCCATTTTGGAGCCAGAG	TATCCTTGTCCTGCCATCGT
D15Mit5		L1	CTTCCTAATTCCTGTCAAGCAAAT	GTTTCATTGGTCAATGGAAACTTA
D15Mit6		A59	CCTGGTCTGAAACACTTTTGC	CTTGTGAGTGCTCCATGCC
D15Mit7		M30	TTTGCAGCTGTGTTCTGCAT	GATTAGGCCACGTGAGCTTC
D15Mit8		A79	GGAAAAGGGAAAAAGATGTGC	TATATTACACTTTCCTTTGCTGCA
D15Mit9		M232	CCATGAGTCCTTCATGCCTT	TGTATATGCAGAAGCAGGCA
D15Mit10		M76	GATCTATAACCAGGGCAGGG	TTAATTCACGGAAATGTTTCAATTT
D15Mit11		M237	TGTGAGAAAAATGACAGTAAGGC	TCACAGAAAGACAAGACAAAAGG
D15Mit12		M34	ATGGACACCTGACACTGCAA	AAGGGCTTTTACCTGGGAAT
D15Mit13		A36	GGAGACAAAAATGAACTCCTGG	TTGTAAGACAAGCATAGCTCAACA
D15Mit14	*Gdc-1*	D17	GAGGAAAACCATGTCAATCACTTC	CCTCCTCTTAAACCAAGATCTCTG
D15Mit15	*Hox3.1*	D6	AGCATACACTCTCTTGTTCCTGCT	AATAAATACCAGAGAAGCACCGTG
D15Mit16	*Hoxmaa*	D131	AGACTCAGAGGGCAAAATAAAGC	TCGGCTTTTGTCTGTCTGTC

1–Continued

Locus name	Gene name	Assay name	Left primer	Right primer
D15Mit17	*Myc*	D22	GCGTCACTGATAGTAGGGAG	GTACCCCAATCCTGAACCAC
D15Nds1		T35	GAGTAGGTTGGAATTTCTCTC	ACAAATATACACTACTGGACAA
D15Nds2		T18	GCCTATTTATTTCAAAGATATGAC	TGATATCGAGGCATACATGAG
D16Mit1		A70	CGCCCTCTAAGGTGACTCAG	AGAGAGGGGTTATGGGGTTG
D16Mit2		L80	CCAATGCCCTCTTATGACCT	TTCTAGTGCGTCCTACCCAG
D16Mit3		M127	TCTAACGCCCTCTCTCTACC	CCAAATGTGATTGCACAAGG
D16Mit4		M203	AGTTCCAGGCTACTTGGGGT	GAGCCCTCATTGCAAATCAT
D16Mit5		A38	CGGGGATCATCCCTAAAAAC	TCCCCAATTCCTCTTGTGTC
D16Mit6		L7	CAGGTCCAAGAGGAGAACCA	TTTGACCTGTGAGCCTGTGG
D16Mit7		L39	CTGCCACCCCTGAACCATTA	CTACAAGATGTGGGGCATGA
D16Nds2		T37	ATTGGTGAGCTTACAGAATAC	GTGGTCATGATATTCGTAGAT
D17Mit1		M124	TGCTTGAAATCCTGGGTTCA	TGCAAAAATGTATGTGCCTG
D17Mit2		A18	ACAAACATGTTGGCCTAATTCC	TTGAGTTTAAGCCCCTAGAATCC
D17Mit3		L28	GATCTTTTCTTATTCTGGTT	GCAAAGTCATGTACTCTGAG
D17Mit4		M114	GCTGTGCTTCCACACTCAGT	TTTCTGAAAAAGCCTCTCAA
D17Mit5		M92	TGGGAACTTTCCAGACTTCC	CCCTTTCCTCCAAACTCTCA
D17Mit6		M254	GTACATGTAGAGAAATGGAGGTG	GCTTATGTTCTTTAACAAGAATGTG
D17Mit7		L4	ACTCCTTNGGGACCTGCATT	ACCGCTCAGGGAGTGCACTT
D17Mit7		A23	TCTAATCCCATGTATATGTGGTGG	TTCCTCTGGACTCCTTGGG
D17Mit9		A51	TCAGCCCTTAAAAATTACTCTTGG	CCCCACCAACTGTCCTCTAA
D17Mit10		L36	TGCACTTGCATAAGGAAAAC	GACTTTGGGGCCTACTTATG
D17Mit11		M145	TGAATTTATGAGGGGGGTCA	TGTCCCATATCTCTCTTTATACACA
D17Mit13		L57	GATCCAGACCACACCCCCTCACCA	TCCTTTGAGAGCCAAGCTTGAAGG
D17Mit16		A25	CCAGAAGACAGCATTCCACA	GTATGTCAGGGCTAGTTGACAGG
D17Mit18		M33	GCAGCTCATTCTTAGTCCCTAAT	TCATGAGTCCCCAAACTAGC
D17Mit19		M44	GAGCTGGTAAATGCTTTGGC	TTGAGTACCCCGTACTTGCC
D17Mit20	*C3*	D129	AGAACAGGACACCGGACATC	TCATAAGTAGGCACACCAATGC
D17Mit21	*Mhcab2*	D21	TAACACCAGACATTGACCTC	AGTCTAGATATGTGTCTCCC
D17Mit22	*Mhceb2*	D16	GGTAAGCATTAGATAGAGAG	TTATGATCTCCACACACGTG
D17Mit23	*Pim1*	D106	TCGAGCTGGTTGAACGAAC	CGGGAAAGCATGGAATTTAA
D17Mit24	*Thy19*	D12	ACCTCTCACCTCTCTCTGTG	TGGAGAGACGTCCTATGATG
D17Nds2	*Hsp68*	T9	GTAATTGCGTTGACTGTTAAAT	AGTGCTGCTCCCAACATTACT
D17Nds3	*Tnfb*	T68	TTCCTGTGGCGGCCTTATCAG	AGACAATGGGTAACAGAGGCA
D18Mit1		M42	TGAGCAAAATACATTGCATG	GGGATACCAGGCCAGACATA
D18Mit2		L9	TTCCCTATCCAGTTGTGTGC	CCCCTGTAGCTCAACCCACT
D18Mit3		L76	TTCCCTATCCAGTTGTGTGC	AGCAGAGAATGCACCACCTC
D18Mit4		M51	ACTGTTGCTGGGGAATGG	CCAAGTTCAAAGCTGCTGG
D18Mit5		M57	TTGTCCACTGATTGCCACAT	CGTATACCCCCACCATGTTC
D18Mit6		A104	GATGAGCTAGGAGGAGATATGAGC	CATACTTACTACAGGGTTTTGGGC
D18Mit7		M108	ACAGGAGAACGGGAACTCAG	GCCAGAGTGGACCAAGATGA
D18Mit8		L24	TTTGGAATCTGGCATGTTAC	GTCTGAAATGAAGTGCCTGC
D18Mit9		M209	AGAGGCATTGCACACACAAG	GCCCCTTGGAGAGTTGGT
D18Mit10		A100	TATCCACCCATTCCAACCTC	GGATTGAGGTTGCTCTTGGA
D18Mit12		A20	TTGTCAGTTTCTTGTGAGGGG	TGTTTAATAAGCCTTTTCCTGAGG
D18Mit14		L13	GAGGTGATGTGGACACACTC	ACACAGCCTAGAATGCACGG
D18Mit15		L87	CAGACTTCATAGCAACACCCTG	TAACATGAAAACAGAAACAGCCA
D18Mit16		A35	TTCCCTTTGGAGACTGTGCT	TGGAATTACAGGGCTTCCTG
D18Mit17	*Grl-1*	D118	TCAGGCAGATTCCAAGCAG	CTGTGGGTAGCCCAAGTCAT
D18Nds1	*Mbp*	T11	CAGTACAGCCAGGACACAGAA	ATGGCTGACCAACTCTCTAGC
D19Mit1		A17	AATCCTTGTTCACTCTATCAAGGC	CATGAAGAGTCCAGTAGAAACCTC
D19Mit2		M109	TGTTGATAGTGCAAGGTGCG	CAAGGGGCCATACCTAGTGA
D19Mit3		M13	CTTCCCCTACTGCAGTGCTC	TTGCATAGTTGGCCAAAGTG
D19Mit4		M230	CGGCTACCCGACACTCTAAA	ATTGGCTTGCCCTAACCC
D19Mit5		A75	TGTTTTGACCTATTTGTTTCATGG	GGTATCTCCTAGTTTTCCTGATTT
DXMit1		L43	CAAGCAACCGAGGAAGACAT	CAGGATGCTAATCACCCTGC
DXMit3		M131	AAAAGGTCATGGCAAAAGGA	AGGAGAAAGTGCAGGGAGGT
DXMit4		M118	TGGACAGTGCTTGAGGAATG	GCAAAACAGCTACATTTGGG
DXMit5		A19	CAACCTCTGAGCTCTCCCAC	TGTTGTCTAATTCCTTCAGGCA
DXMit6	*Zfx*	D28	ACCATTCAAATTGGCAAGGC	GTGGCTCGAGTTGTTTGCAG
DXNds1	*Hprt*	T8	TGACAACTTCTGTCCTCAACA	ATGCCGTCCTTTATCTAGAAC
DXNds2	*Plp*	T4	TAATATAACAGATAACCAACCATT	CATTTTGTAAGATGAGTTTCTA
Unmapped		*A66*	TCAGGGCTCTCTAAGGGACA	ACTATGCAGCCACCAAATCC
Unmapped		*M251*	TTCCTCAACTAAACGCTGGA	CATTTTCCTGTATACCTGAATTTT

The gene name given for SSLPs found in gene sequences from GenBank. The assay name refers to the specific assay used to genotype the locus; formal reference to the assay should be preceded by the symbol "Mit-". The primer sequences are given from 5′ to 3′.

TABLE 2

Allele sizes of simple sequence length polymorphisms in 12 inbred strains

Locus name	Gene name	Assay name	OB	CAST	B6	SPR	DBA	A	C3H	BALB	AKR	NON	NOD	LP
D1Mit1		L33	123	118	123	135	126	123	132	123	—	—	—	—
D1Mit2		A26	172	150	172	185	—	—	—	—	—	—	—	—
D1Mit3		M253	160	185	160	200	160	185	185	185	206	185	185	187
D1Mit4		M46	200	168	200	170	200	200	200	200	200	197	195	210
D1Mit5		L20	148	126	148	150	152	152	152	148	150	150	152	152
D1Mit7		A80	108	156	108	125	125	125	125	108	108	125	108	108
D1Mit8		L31	220	190	220	178	220	220	220	220	201	220	201	220
D1Mit9		M111	160	140	160	162	160	160	160	160	147	160	147	160
D1Mit10		A117	140	152	140	125	140	140	147	140	135	140	140	135
D1Mit11		M17	100	111	100	—	106	100	100	100	106	100	106	106
D1Mit12		M93	133	129	133	170	133	133	133	126	133	133	133	—
D1Mit13		L30	202	207	202	—	202	210	211	202	202	202	202	202
D1Mit14		M193	180	200	180	170	215	215	215	205	215	175	175	190
D1Mit15		M146	160	188	160	154	160	186	183	160	183	183	183	183
D1Mit16		L46	190	185	190	195	201	185	190	190	190	164	185	190
D1Mit17		M41	170	190	170	188	174	183	183	176	183	170	183	176
D1Mit18		A77	160	180	160	170	160	160	160	170	170	160	170	205
D1Mit19		L86	113	148	113	123	120	121	121	113	113	113	108	103
D1Nds2		T17	180	167	180	123	180	158	159	158	190	178	159	158
D2Mit1		M128	124	140	124	96	120	120	120	120	120	120	120	124
D2Mit2		M112	147	129	129	138	129	129	129	129	129	135	135	129
D2Mit3		M116	160	194	160	158	160	160	160	160	160	160	160	160
D2Mit4		M52	190	166	190	176	190	190	190	190	190	190	190	190
D2Mit5		A41	141	180	—	—	139	137	139	141	141	141	141	139
D2Mit6		L18	135	147	135	110	126	126	135	135	126	135	135	135
D2Mit7		L44	150	122	150	148	145	147	145	150	147	145	143	145
D2Mit8		M199	188	180	188	212	188	188	188	188	188	188	188	188
D2Mit9		M85	190	195	190	174	195	187	190	190	190	185	185	190
D2Mit10		M39	152	150	152	158	150	156	150	152	152	152	152	152
D2Mit11		M134	226	232	226	264	226	232	226	232	232	226	232	232
D2Mit12		M179	201	194	201	189	189	200	200	189	200	189	201	189
D2Mit13		M130	190	193	190	170	192	192	180	192	192	193	193	193
D2Mit14		M163	142	152	142	198	130	130	130	130	130	142	130	130
D2Mit15		A61	145	178	145	160	145	162	160	145	160	145	145	145
D2Mit16		M186	238	250	238	242	238	238	238	238	238	238	238	238
D2Mit17		M246	205	242	205	420	220	220	220	220	220	214	242	214
D2Mit19		A83	108	124	108	127	108	108	108	108	108	108	108	108
D2Mit21		M184	260	250	258	250	250	260	258	258	258	258	256	256
D2Mit22		M167	190	162	190	112	190	190	190	190	190	147	147	—
D2Mit24		M75	180	183	180	180	180	180	180	180	180	180	180	180
D2Mit25		A67	118	140	118	126	118	118	118	118	118	118	—	118
D2Mit26		M37	195	210	195	190	195	195	195	195	195	210	210	210
D2Mit27		M106	180	238	180	250	180	—	180	—	—	—	—	—
D2Mit28	*Snap*	D25	130	142	130	123	130	130	130	130	130	—	—	—
D2Mit29	*Sup-4*	D115	115	120	115	110	115	115	115	115	115	115	115	115
D2Mit30	*Trh-1*	D111	320	340	320	80	137	137	137	137	137	121	121	121
D2Nds1		T19	178	158	178	182	185	182	182	185	182	152	188	185
D2Nds2		T57	122	88	122	114	122	122	122	122	122	122	122	122
D2Nds3	*Il-1B*	T15	280	190	280	140	280	280	280	400	270	270	280	270
D3Mit1		M28	145	118	120	—	120	120	120	123	120	120	123	143
D3Mit3		M250	108	200	108	88	108	109	104	109	108	104	108	109
D3Mit4		L40	140	150	140	147	140	140	140	140	140	140	140	140
D3Mit5		M123	188	182	182	178	188	188	188	188	188	188	188	182
D3Mit6		M149	145	133	147	125	136	136	136	136	136	136	136	147
D3Mit7		M74	147	142	147	147	147	142	142	142	142	142	142	142
D3Mit9		A85	225	238	225	210	238	214	216	238	238	225	230	216
D3Mit10		A34	145	158	145	132	140	134	—	134	132	136	121	138
D3Mit11		L38	147	204	147	152	147	165	165	165	165	163	147	147
D3Mit12		A60	155	120	155	—	126	157	155	155	157	126	126	157
D3Mit13		L37	220	225	220	236	220	220	220	220	220	220	235	237
D3Mit13		L8	220	238	220	240	220	—	220	220	220	220	238	240
D3Mit14		M206	170	127	170	132	198	198	198	198	198	198	198	170

TABLE 2—Continued

Locus name	Gene name	Assay name	OB	CAST	B6	SPR	DBA	A	C3H	BALB	AKR	NON	NOD	LP
D3Mit15		A55	145	185	145	175	212	145	145	145	145	145	145	—
D3Mit16		M159	188	194	188	220	—	186	186	186	—	186	186	186
D3Mit17		M235	208	200	208	—	180	180	180	180	180	180	208	188
D3Mit18		A96	235	242	235	192	235	214	214	214	214	214	235	214
D3Mit19		M141	160	176	160	147	176	176	176	176	176	160	176	158
D3Mit21	*Il-2*	D31	236	216	236	208	218	218	218	236	218	236	218	218
D3Mit22	*Rp132-ps*	D122	240	265	240	207	238	255	240	220	238	240	245	220
D3Nds2		T21	115	147	115	133	122	115	115	115	115	115	115	115
D4Mit1		A73	120	93	120	120	120	115	120	112	112	120	112	112
D4Mit2		L67	178	172	178	178	178	172	178	172	172	172	172	172
D4Mit2		L6	195	185	195	195	195	185	195	185	185	185	185	185
D4Mit4		M31	165	169	165	158	165	163	167	163	163	165	163	163
D4Mit5		A1	138	115	138	120	148	138	138	138	138	138	138	138
D4Mit6		M64	80	60	80	58	78	62	62	62	80	80	62	83
D4Mit7		A71	151	147	151	160	151	151	151	151	151	151	151	149
D4Mit9		M241	206	212	206	238	208	210	210	210	195	200	210	200
D4Mit11		M8	144	170	144	183	144	178	144	144	144	144	149	144
D4Mit12		M15	198	190	198	185	168	168	168	168	170	169	167	170
D4Mit13		M169	92	88	92	106	97	92	108	92	111	108	108	92
D4Mit14		A69	133	130	133	145	140	140	133	140	133	133	133	140
D4Mit15		A122	280	315	280	—	280	330	330	330	280	330	330	318
D4Mit16		A65	220	245	220	226	239	239	239	239	220	239	220	239
D4Mit17	*Orm-1*	D1	147	145	147	105	141	147	147	147	138	136	136	136
D4Mit205		M205	195	197	—	190	197	204	204	204	202	200	201	202
D4Wsm1	*Ifa*	F1	160	185	—	—	—	—	—	—	—	—	—	—
D4Nds2		T24	97	95	97	98	97	91	97	97	93	89	93	97
D4Nds10		T29	90	80	90	—	90	90	90	90	90	—	90	90
D5Mit1		A82	137	145	137	149	129	137	137	137	129	137	137	129
D5Mit3		M197	165	147	165	—	165	165	165	165	165	165	165	165
D5Mit4		M189	195	238	195	250	195	195	195	195	195	195	195	195
D5Mit5		A11	145	166	145	163	145	145	145	145	145	160	145	145
D5Mit6		L42	135	125	135	108	135	135	135	135	135	135	135	131
D5Mit7		M154	160	147	160	160	147	160	147	147	147	147	147	147
D5Mit9		A9	149	180	149	138	149	149	149	149	149	149	149	149
D5Mit10		M207	196	209	196	205	203	188	194	190	201	192	200	196
D5Mit11		M97	203	195	206	210	188	188	188	199	188	188	188	203
D5Mit12	*Csnb*	D128	120	85	120	115	115	120	115	120	120	115	115	110
D5Nds2		T26	168	175	168	122	178	168	178	168	168	178	178	168
D5Nds4	*Afp*	T61	90	85	90	—	97	90	97	97	90	86	97	85
D6Mit1		A10	217	239	217	280	217	217	245	245	217	245	217	217
D6Mit3		L59	308	236	308	—	308	300	308	300	239	308	308	236
D6Mit4		M239	102	107	102	121	102	90	102	90	95	90	90	108
D6Mit5		M161	168	158	168	168	168	168	168	168	168	168	168	168
D6Mit6		M259	100	109	100	96	100	110	100	100	100	100	110	100
D6Mit6		M227	100	110	100	—	100	113	100	100	100	100	113	100
D6Mit8		M240	164	180	164	182	164	190	164	190	170	190	188	178
D6Mit9		L23	143	138	143	152	123	123	123	143	143	123	123	123
D6Mit10		M78	198	210	198	212	206	198	191	198	198	207	191	198
D6Mit11		M170	94	127	94	98	94	94	94	94	94	94	94	94
D6Mit12		M11	123	147	123	170	123	123	123	123	123	123	123	123
D6Mit13	*Prp*	D34	158	170	158	—	152	152	156	152	152	160	155	152
D6Mit14		M190	160	172	160	174	149	156	149	152	149	174	174	168
D6Mit15		M148	260	220	260	260	195	195	195	195	260	260	150	195
D6Mit16	*Ly2*	D11	155	167	155	152	147	157	147	155	130	155	130	155
D6Rck1	*Cpa*	F3	250	230	—	234	250	—	—	—	—	—	—	—
D6Rck2	*Mirp*	F2	170	155	—	147	174	—	—	—	—	—	—	—
D6Rck3		F103	110	90	—	112	—	—	—	—	—	—	—	—
D6Nds4		T59	91	114	91	112	91	91	91	91	91	91	91	91
D6Nds5		T23	98	108	98	118	—	105	—	105	—	105	98	105
D7Mit1		M208	298	309	298	298	298	298	298	298	298	298	301	298
D7Mit5		M187	215	182	215	215	215	215	215	215	215	215	215	215
D7Mit7		L12	80	90	80	—	90	77	90	77	80	80	80	90
D7Mit8		M183	150	153	150	165	148	151	148	148	146	146	146	—
D7Mit9		A89	130	145	130	—	130	130	130	130	128	128	128	130
D7Mit10		L72	180	190	180	180	180	180	180	180	180	180	—	180

TABLE 2—Continued

Locus name	Gene name	Assay name	OB	CAST	B6	SPR	DBA	A	C3H	BALB	AKR	NON	NOD	LP
D7Mit10		L25	150	158	150	150	150	150	150	150	150	150	150	150
D7Mit12		M23	197	208	197	220	206	197	197	197	206	205	197	199
D7Mit13		A113	195	200	195	210	195	195	195	195	200	195	195	195
D7Mit14		L79	147	142	147	147	147	147	147	137	147	147	147	147
D7Mit15		M47	138	127	138	129	138	138	138	123	123	138	138	134
D7Mit16		A13	245	248	248	—	248	248	248	248	248	248	248	18
D7Mit17		M91	160	144	160	170	162	160	162	160	160	162	144	162
D7Mit18	*Gas-2*	D117	120	109	120	112	120	120	120	120	120	120	120	120
D7Mit19	*Tyr*	D108	135	131	135	127	135	135	135	135	135	135	135	135
D7Mit20	*Mb-1*	D103	107	100	107	80	107	107	107	107	107	107	95	107
D7Nds1		T27	238	301	238	270	260	260	265	270	270	247	247	270
D7Nds2		T28	118	114	118	97	112	116	112	119	119	114	114	114
D7Nds4	*Int-2*	T63	168	145	166	175	160	160	160	166	166	166	166	160
D7Nds5	*Ngfg*	T62	145	150	145	—	157	142	143	140	145	143	143	143
D8Mit1		M70	215	255	215	215	215	215	215	215	215	215	215	215
D8Mit3		M195	178	185	178	160	187	187	187	187	187	187	187	187
D8Mit4		M71	157	191	157	170	195	200	195	200	195	195	200	160
D8Mit5		M176	166	150	166	100	166	166	166	166	166	166	166	166
D8Mit6		M158	170	201	170	195	170	170	170	170	170	170	170	170
D8Mit7		M138	178	226	178	347	178	178	178	178	178	178	—	178
D8Mit8		M257	125	93	125	110	116	118	116	116	116	116	116	116
D8Mit9		A62	153	119	153	116	151	153	151	151	140	153	153	—
D8Mit11		A105	215	203	215	195	213	215	213	217	215	214	213	213
D8Mit12		L11	120	127	120	125	120	120	120	120	120	120	117	120
D8Mit13		M77	98	114	98	114	98	98	86	105	94	98	98	108
D8Mit14		L34	145	158	145	132	145	170	140	170	140	145	140	170
D8Mit15	*Mt2*	D20	180	187	180	160	180	180	180	180	178	180	185	178
D8Mit16	*Polb*	D100	310	315	310	325	300	310	300	310	300	310	310	310
D9Mit1		M88	110	132	110	110	110	110	110	110	110	110	110	110
D9Mit2		L32	177	161	177	161	177	185	185	185	176	170	185	160
D9Mit4		M151	124	131	124	120	138	138	140	138	124	138	138	136
D9Mit6		A78	144	136	142	—	140	140	140	142	140	140	140	140
D9Mit8		M211	185	180	185	210	193	195	—	193	193	193	193	178
D9Mit9		A72	126	116	126	112	126	138	138	138	126	138	126	130
D9Mit10		M86	150	178	150	156	147	150	150	150	150	147	150	150
D9Mit11		L60	76	100	76	145	108	122	122	122	115	110	112	100
D9Mit12		M73	93	100	93	—	88	82	82	82	88	91	91	93
D9Mit14		M236	78	92	78	95	78	—	—	—	80	70	—	—
D9Mit15		M160	160	166	160	138	155	155	155	155	157	155	155	155
D9Mit16		A5	180	196	180	200	180	167	167	167	176	176	180	180
D9Mit17		L19	157	130	157	145	157	161	161	161	145	143	145	140
D9Mit18		M10	180	210	180	180	204	210	210	213	204	204	180	180
D9Mit19		M157	102	92	102	108	89	108	108	108	89	89	102	102
D9Mit20		L64	114	108	114	106	106	117	117	117	114	106	106	123
D9Mit21	*Cypla2*	D15	187	210	187	168	180	187	187	187	189	187	180	180
D9Mit22	*Ncam*	D134	220	230	220	208	230	230	230	225	210	210	—	—
D9Mit23	*T3d*	D4	210	290	210	320	210	210	210	210	214	212	211	210
D9Mit24	*Trf*	D26	127	149	127	145	127	136	136	136	136	132	136	136
D9Nds2		T30	121	130	121	110	125	125	125	125	125	127	125	130
D10Mit1		M153	100	112	100	—	100	—	—	87	87	110	—	60
D10Mit2		M24	124	121	124	116	124	132	124	132	124	120	132	132
D10Mit3		A114	245	210	245	205	215	245	215	245	215	245	245	245
D10Mit4		M139	134	147	134	134	134	134	134	134	134	134	134	134
D10Mit5		M67	190	201	190	210	190	190	190	190	190	190	190	190
D10Mit7		L62	147	137	147	176	147	147	147	147	147	147	147	147
D10Mit8		M3	208	188	201	215	201	201	201	201	201	201	201	206
D10Mit9		A37	159	155	159	155	159	159	159	159	159	159	—	159
D10Mit10		M7	180	136	180	160	128	128	128	128	128	180	128	128
D10Mit11		A88	201	172	201	175	172	172	172	172	172	201	201	172
D10Mit12		M172	242	236	242	—	242	242	212	242	212	242	212	242
D10Mit13		A63	130	113	130	—	130	130	130	130	130	130	130	130
D10Mit14		M175	192	174	192	199	182	182	194	182	188	192	182	182

TABLE 2—Continued

Locus name	Gene name	Assay name	OB	CAST	B6	SPR	DBA	A	C3H	BALB	AKR	NON	NOD	LP
D10Mit15	*Sqr3*	D30	185	140	185	124	189	185	185	175	185	185	187	—
D10Nds1		T31	130	132	130	—	130	152	152	152	152	130	145	152
D10Nds2		T32	145	127	145	138	150	145	145	145	145	145	145	145
D10Nds3		T54	94	89	94	94	94	94	94	94	94	94	94	94
D11Mit1		M215	153	110	153	126	153	153	153	162	162	153	153	153
D11Mit2		L14	124	118	124	111	126	115	140	115	140	115	140	115
D11Mit4		A124	250	246	246	238	300	307	242	242	307	242	244	306
D11Mit5		A2	220	144	220	—	189	213	188	188	213	178	185	185
D11Mit7		M119	144	148	144	172	144	144	144	144	144	144	144	144
D11Mit8		M212	155	170	155	—	155	155	155	133	133	155	155	155
D11Mit10		M162	100	125	100	116	100	132	100	100	100	100	100	100
D11Mit11		M43	238	216	238	210	238	244	238	238	238	238	238	238
D11Mit12		L3	140	150	140	140	140	150	147	145	140	140	142	140
D11Mit13	*Ace*	DACE	161	165	—	—	—	—	—	—	—	—	—	—
D11Mit14	*AntP91A*	D2	158	148	158	146	161	158	167	158	139	161	158	161
D11Mit15	*Glut-4*	D5	147	143	147	143	143	147	147	147	147	151	147	151
D11Mit16	*Lif*	D133	120	135	120	113	120	120	120	120	113	113	120	—
D11Nds1		T33	102	132	102	100	108	102	108	108	108	108	108	108
D11Nds7	*Gfap*	T12	163	181	163	163	153	153	163	153	153	153	163	—
D11Nds9	*Il-5*	T14	306	309	306	—	306	306	306	306	306	302	302	306
D12Mit1		M50	255	230	255	250	244	244	244	244	244	244	244	270
D12Mit2		M27	132	178	132	132	149	132	132	132	132	149	149	132
D12Mit3		L41	123	112	123	130	127	123	123	123	127	123	123	127
D12Mit4		A64	203	270	206	214	208	208	208	196	184	184	199	208
D12Mit5		L58	180	163	180	144	163	163	163	163	163	182	180	163
D12Mit6		L16	108	125	108	110	108	108	108	108	108	108	108	121
D12Mit7		M62	108	130	108	—	121	108	108	123	123	106	123	123
D12Mit8	*Igh-C*	D7	172	180	172	148	181	148	174	174	185	174	170	170
D12Nds1		T51	93	112	93	—	93	93	93	93	93	—	93	93
D12Nds2	*Igh-V*	T1	155	159	195	195	162	193	178	165	170	183	195	165
D12Nds11	*Odc*	T64	170	178	170	158	175	178	178	178	178	—	178	—
D13Mit1		A86	149	151	149	153	149	149	149	149	140	153	149	153
D13Mit3		M79	159	196	159	178	196	188	196	188	164	188	164	163
D13Mit4		M231	185	209	185	209	185	185	185	185	185	185	185	185
D13Mit5		M38	194	190	194	—	194	194	194	194	194	194	194	194
D13Mit7		A68	140	137	140	121	140	145	140	142	140	142	142	142
D13Mit8		M61	190	200	190	250	190	190	182	190	182	190	184	182
D13Mit9		M147	126	116	126	132	145	126	145	126	126	126	145	126
D13Mit10		L61	152	144	152	105	152	160	160	160	149	160	160	160
D13Mit11		A91	147	162	147	162	158	158	158	158	158	158	160	162
D13Mit13	*Il-9*	D24	151	142	151	145	145	140	140	140	151	151	145	140
D13Mit14	*Sqr4*	D29	150	120	150	156	150	146	146	146	150	143	150	146
D14Mit1		A103	108	104	108	142	98	108	98	108	108	108	104	98
D14Mit2		A24	144	146	144	153	146	144	144	140	144	144	146	146
D14Mit3		M32	236	225	236	245	236	236	236	236	236	236	236	240
D14Mit4		M228	196	200	196	186	194	196	196	196	196	196	200	198
D14Mit5		M214	178	182	178	156	164	178	164	178	164	178	178	178
D14Mit6		A119	150	157	150	185	155	155	155	155	155	150	155	155
D14Mit7		L27	109	91	109	107	99	99	99	99	99	112	109	112
D14Mit8		A44	203	210	203	190	203	203	203	203	203	205	203	195
D14Mit9		A93	238	245	238	—	238	238	238	245	245	238	238	238
D14Nds1	*Plau*	T10	182	201	182	190	201	182	182	190	182	190	188	190
D15Mit1		L29	185	180	185	—	190	190	190	190	190	—	183	190
D15Mit2		L10	94	109	94	—	89	89	89	89	89	89	89	89
D15Mit3		L78	140	152	140	154	142	142	137	138	140	140	137	140
D15Mit5		L1	100	123	100	—	118	118	118	100	118	123	118	132
D15Mit6		A59	130	104	130	106	134	132	127	130	130	128	128	104
D15Mit7		M30	109	115	109	126	100	109	100	109	109	109	109	129
D15Mit8		A79	117	123	117	119	125	117	125	117	117	117	117	120
D15Mit9		M232	138	153	138	300	138	138	138	138	138	138	138	138
D15Mit10		M76	222	242	222	178	222	220	—	222	222	222	236	236
D15Mit11		M237	106	126	106	110	106	94	106	106	121	106	106	106
D15Mit12		M34	150	123	150	144	160	150	150	150	144	150	150	161
D15Mit13		A36	140	165	140	190	120	140	140	140	125	—	110	120
D15Mit14	*Gdc-1*	D17	190	270	190	188	190	183	183	195	190	188	190	230

TABLE 2—Continued

Locus name	Gene name	Assay name	OB	CAST	B6	SPR	DBA	A	C3H	BALB	AKR	NON	NOD	LP
D15Mit15	*Hox3.1*	D6	159	164	159	168	145	159	166	159	145	—	—	159
D15Mit16	*Hoxmaa*	D131	120	145	120	155	145	126	120	126	145	126	—	123
D15Mit17	*Myc*	D22	145	143	145	143	145	145	138	138	145	143	140	145
D15Nds1		T35	100	146	100	—	98	98	98	98	105	98	96	100
D15Nds2		T18	122	115	122	—	115	111	122	115	122	—	122	120
D16Mit1		A70	106	94	106	140	106	106	106	106	106	106	[illegible]	106
D16Mit2		L80	189	193	189	177	189	189	189	189	189	189	189	189
D16Mit3		M127	102	76	102	97	100	104	100	104	104	104	104	100
D16Mit4		M203	132	130	132	145	123	147	123	149	126	149	149	149
D16Mit5		A38	158	163	158	163	134	134	160	134	160	158	160	160
D16Mit6		L7	190	175	190	212	195	190	190	190	195	190	190	185
D16Mit7		L39	162	175	162	165	162	162	162	165	165	162	162	162
D16Nds2		T37	98	88	98	—	103	90	103	90	88	88	88	103
D17Mit1		M124	201	208	201	—	201	195	195	195	193	193	201	201
D17Mit2		A18	230	250	230	—	230	220	230	230	225	230	230	230
D17Mit3		L28	130	128	130	120	123	132	123	130	130	128	130	—
D17Mit4		M114	95	98	—	140	95	95	95	95	95	95	95	95
D17Mit5		M92	260	250	260	242	260	260	260	260	260	260	260	260
D17Mit6		M254	106	88	106	104	102	102	102	102	102	102	102	102
D17Mit7		L4	200	214	200	178	204	204	204	204	204	200	204	204
D17Mit7		A23	145	170	145	—	152	152	152	152	154	146	154	152
D17Mit9		A51	117	134	117	100	117	117	117	117	117	117	117	115
D17Mit10		L36	159	133	159	165	150	150	150	159	159	150	148	
D17Mit11		M145	176	192	176	178	150	160	176	150	176	150	178	
D17Mit13		L57	149	144	149	146	144	144	142	144	142	149	149	149
D17Mit16		A25	123	92	122	98	109	94	94	109	94	110	90	122
D17Mit18		M33	246	256	246	238	241	242	241	241	246	241	241	246
D17Mit19		M44	185	158	185	180	185	185	185	185	185	180	174	185
D17Mit20	*C3*	D129	180	198	185	212	178	178	178	178	178	185	185	185
D17Mit21	*Mhcab2*	D21	140	108	140	140	158	124	124	158	124	126	124	136
D17Mit22	*Mhceb2*	D16	160	178	160	164	185	162	162	185	160	—	160	158
D17Mit23	*Pim1*	D106	138	140	138	—	140	145	145	140	145	138	138	140
D17Mit24	*Thy19*	D12	145	140	145	120	130	145	147	130	147	145	145	130
D17Nds2	*Hsp68*	T9	110	105	110	80	105	105	—	105	105	110	125	110
D17Nds3	*Tnfb*	T68	145	120	145	90	126	126	160	126	160	132	132	145
D18Mit1		M42	154	140	154	147	154	154	154	154	154	154	154	136
D18Mit1		A104	145	130	145	143	145	145	145	145	145	145	145	126
D18Mit2		L9	130	163	130	148	130	130	130	130	130	130	130	132
D18Mit3		L76	216	158	189	213	207	207	189	218	189	—	216	218
D18Mit4		M51	212	180	210	188	195	188	195	195	195	170	180	175
D18Mit5		M57	189	200	189	208	189	189	189	189	189	189	189	200
D18Mit7		M108	93	123	93	152	123	93	123	93	93	100	93	132
D18Mit8		L24	77	90	77	88	80	74	80	74	80	—	—	80
D18Mit9		M209	170	172	170	145	160	160	160	160	170	168	160	160
D18Mit10		A100	108	117	109	109	108	108	108	108	108	108	108	108
D18Mit12		A20	122	110	122	132	122	122	132	122	122	132	132	122
D18Mit14		L13	108	130	108	103	103	103	110	103	103	110	110	108
D18Mit15		L87	162	147	162	—	164	164	173	162	160	173	173	158
D18Mit16		A35	207	201	207	199	207	207	207	207	207	207	207	207
D18Mit17	*Grl-1*	D118	212	203	214	210	190	190	190	190	190	190	190	190
D18Nds1	*Mbp*	T11	146	190	146	162	146	146	146	146	146	146	146	—
D19Mit1		A17	123	138	123	162	145	145	145	145	145	143	145	147
D19Mit2		M109	185	163	—	188	—	185	185	185	185	—	196	185
D19Mit3		M13	200	218	200	206	200	200	200	200	205	200	200	215
D19Mit4		M230	200	242	200	190	200	200	200	200	200	200	200	200
D19Mit5		A75	214	195	214	205	214	214	214	214	214	214	214	214
DXMit1		L43	100	108	100	96	86	86	86	86	86	86	86	100
DXMit3		M131	178	182	178	187	178	178	178	178	178	178	178	178
DXMit4		M118	108	100	108	102	108	108	108	108	108	108	108	108
DXMit5		A19	150	145	150	150	150	150	150	140	140	150	150	150
DXMit6	*Zfx*	D28	208	204	208	204	208	208	208	208	208	—	—	208
DXNds1	*Hprt*	T8	108	120	108	110	108	108	108	108	110	—	110	110
DXNds2	*Plp*	T4	178	181	178	—	178	178	178	178	178	—	178	—
Unmapped		M251	95	120	95	160	95	100	95	95	95	95	—	—
Unmapped		A66	242	206	242	245	202	230	206	202	230	206	206	242

The strain designations are: OB = C57 BL/6J-*ob*/*ob*, CAST = CAST/Ei, B6 = C57BL/6J, SPR = SPRET/Ei, DBA = DBA/2J, A = A/J, C3H = C3H/HeJ, BALB = BALB/cJ, AKR = AKR/J, NON = NON/Lt, NOD = NOD/MrkTacBr, LP = LP/J. All allele sizes are given in base pairs. Dashes indicate missing data. Allele sizes are determined relative to molecular weight standards run in another lane, and thus should be considered approximate.

TABLE 3

Polymorphism rates of simple sequence repeats

	OB	CAST	B6	SPR	DBA	A	C3H	BALB	AKR	NON	NOD	LP
OB	—											
CAST	100.0	—										
B6	6.5	98.6	—									
SPR	90.7	95.9	90.2	—								
DBA	52.4	92.5	51.4	90.2	—							
A	53.2	94.4	52.7	92.8	45.8	—						
C3H	52.1	95	50.5	91.2	34.8	35.1	—					
BALB	50.6	94.1	49.3	93.2	45.2	31.6	38	—				
AKR	53.8	94.4	52.4	90.5	48.3	46.2	43.9	42.9	—			
NON	50.5	95.5	49.3	88.9	50.8	51.2	46.7	47.1	53.6	—		
NOD	55.4	92.9	54.5	90.9	53.6	51.2	48.1	51.0	51.0	43.9	—	
LP	58.7	92.1	57.4	91.3	53.4	54.5	53.8	49.3	56.7	55.4	55.9	—

The polymorphism rates were determined for those SSRs that were variant between OB and CAST, thus the rate for that strain combination is necessarily 100% for the markers reported. Strain designations are: OB = C57 BL/6J-*ob*/*ob*, CAST = CAST/Ei, B6 = C57BL/6J, SPR = SPRET/Ei, DBA = DBA/2J, A = A/J, C3H = C3H/HeJ, BALB = BALB/cJ, AKR = AKR/J, NON = NON/Lt, NOD = NOD/MrkTacBr, LP = LP/J.

ent map distances and can interfere with the ability to resolve genetic order accurately (BUETOW 1991). Accordingly, we developed a novel mathematical approach (see MATERIALS AND METHODS) for identifying the potentially erroneous data, so that they could be checked with special care.

We first obtained an empirical estimate of the error rate in our data, by independently repeating the genotyping of about 10% of the loci. Comparing the duplicate typings, we found a discrepancy rate of 1.4% corresponding to an error rate of 0.7 ± 0.2%. Using this estimate, we used a computer program to identify all typings that were at least 10-fold more likely to have arisen if erroneous than if correct (*i.e.*, $LOD_{error} \geq 1.0$). Each such typing was checked by reinspecting the autoradiogram and, if there was any ambiguity, by repeating the typing from scratch. From among the typings identified as potential errors, actual errors were found in 72 cases or about 0.5% of the data. Simulation studies (not shown) showed that the expected number of actual errors that would fail to give rise to a $LOD_{error} \geq 1.0$ was about 20. About half of these errors would be expected to occur at markers that were either at the ends of linkage groups or adjacent to large intervals (since the power to detect error by virtue of double crossovers is least in these cases). Accordingly, we retyped all such markers from scratch. Overall, we estimate that approximately 10 errors may remain in the data–corresponding to a residual error rate of about 0.1%. These data should provide a firm foundation on which to build an even denser map.

Anchoring of the map: It was important to anchor our map relative to the existing mouse genetic map, in order to increase its utility for genetic studies. We used two methods. (1) Because 157 of the genetic markers are polymorphic in the BXD crosses, these markers could be mapped in the BXD recombinant inbred lines (BAILEY 1971; TAYLOR, HEINIGER and MEIER 1973). We typed a well spaced collection of 121 of these markers (Table 4), of which 100 could be unambiguously linked to known strain distribution patterns which then served as anchor points. Most anchors are indicated in Figure 2, although some are omitted when several anchors are present in the same region. (2) Because 32 of our SSLPs came from genes with previously known chromosomal positions, this provided a further collection of anchor points. [Conversely, our map provided chromosomal locations for 10 genes which were previously unmapped or incorrectly mapped (Table 5).]

Further confirmation of our anchoring came from two sources: (1) W. FRANKEL and J. COFFIN (personal communication) mapped a number of RFLPs corresponding to endogenous retroviruses segregating in our cross, six of which are shown; and (2) our map included 30 SSLPs whose positions had been previously determined in crosses by J. TODD and colleagues (LOVE *et al.* 1990; CORNALL *et al.* 1991; HEARNE *et al.* 1991).

Mutation rate: Studying the BXD recombinant inbred lines provided an excellent opportunity to measure the average mutation rate of SSLPs per generation, by looking for the occurrence of individual RI lines fixed for a nonparental allele. We observed nine such events, indicated in Table 4. Since we have typed 22 RI strains for 121 genetic markers and since the RI lines have been separated for about 75 generations (TAYLOR 1989), we estimate that there were about 200,000 opportunities for mutational

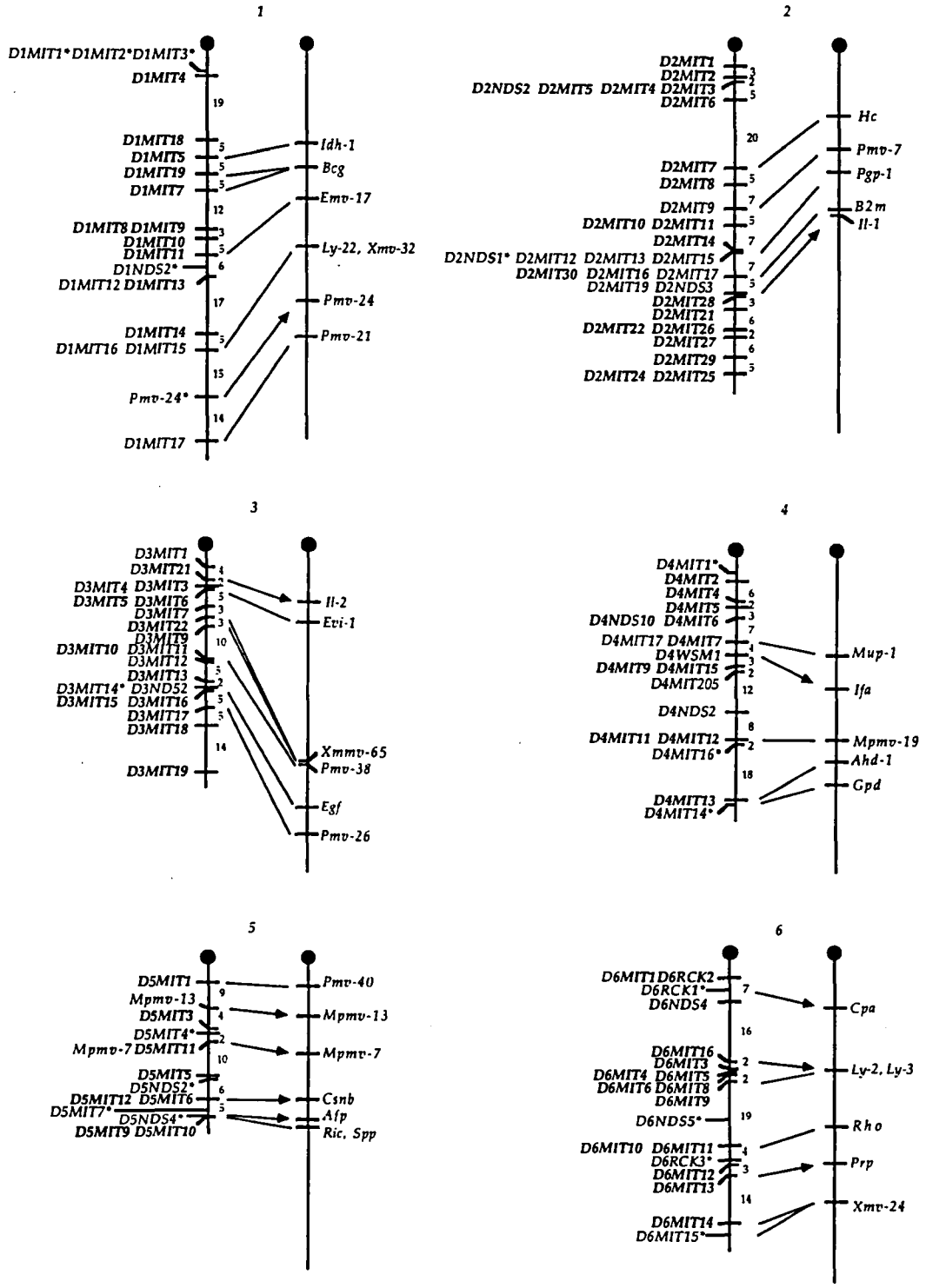
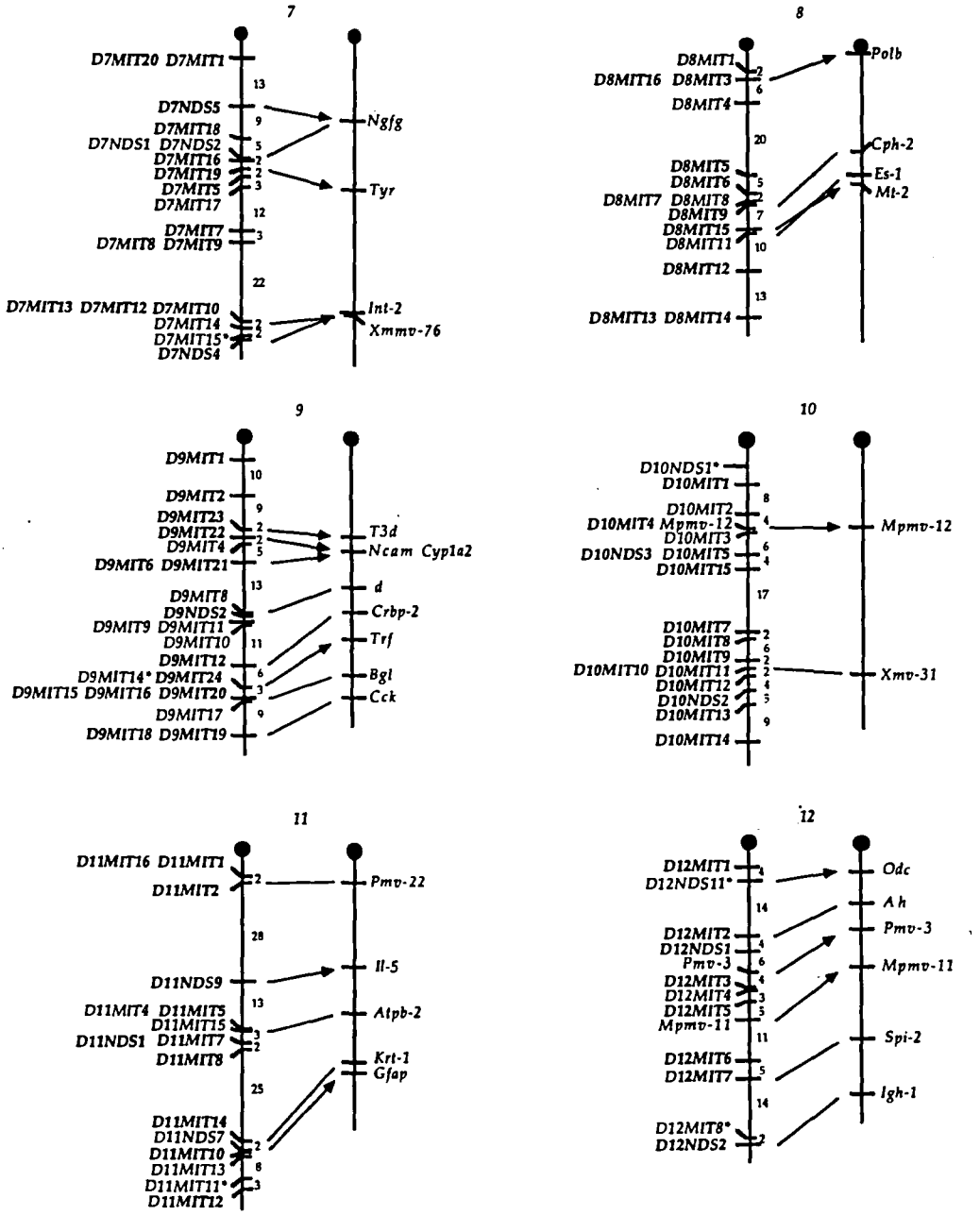


FIGURE 2.—Genetic linkage map of the mouse. Chromosomes are represented by two diagrams, the left side being the map reported in this paper, and right side being taken from the consensus map reported in the October 1990 edition of the GBASE database. For the SSLP map, a length of five cM has been arbitrarily added to each end. For the GBASE map, map lengths are equal to the fractional cytogenetic length for the chromosome multiplied by 1600 cM (the estimated genetic length of the mouse genome). Centromeres are indicated by filled circles. SSLPs are defined in Table 2. Six retroviral markers (denoted by their usual locus names) were scored in the cross and are shown on the map. Symbols indicate the degree of support for the indicated genetic order. Markers whose order relative to the map is supported by a LOD score of at least 3 are shown in bold type; by a LOD score of between 2 and 3 in plain face type; and by a LOD score of between 1 and 2 are marked with an asterisk (see MATERIALS AND METHODS). Markers listed on the same horizontal line did not recombine in the 46 animal F_2 intercross studied here. Centimorgan distances between markers are indicated, except for those less than 2 cM. Centimorgans are based on Kosambi's map function. Although the appropriate mapping function for the mouse genome is not precisely known, this function should be adequate for the present purposes. In any case, the choice of mapping function only has a significant effect on the large intervals,



whose distances should necessarily be considered to be approximate. Maximum likelihood order for LOD 1 markers relative to flanking markers is indicated, but exact distances are not. Such markers are indicated by a horizontal tick mark that does not cross the chromosome. The lines connecting the two maps indicate anchor points: Lines with arrowheads indicate that identity between markers on the two maps. Lines without arrowheads indicate that an SSLP is genetically linked to the marker shown on the GBASE map, based on analysis of BXD recombinant inbred strains. Because lines with arrowheads indicate identity and lines without arrowheads simply imply linkage, two such lines may cross one another without implying inversion of gene order (as occurs on chromosome *8*). (Figure 2 is concluded on page 440.)

events. This corresponds to an average mutation rate of about 1/22,000 per locus per generation.

Repeat occurrence of SSRs: In selecting SSRs from the genome, it is possible that some loci will be sampled more than once. Specifically, suppose that k objects are randomly chosen with replacement from a set of size N. For $k^2 > N > k^{3/2}$, a simple Poisson approximation shows that about $k(k-1)/2N$ objects will be chosen twice and few or no objects will be chosen three times. In the present case, our 319 clones containing CA-repeats were not selected at random from the genome, but rather from the set of CA-repeats contained in *Mbo*I fragments of between 250 and 500 bp and positioned within the fragment so that we would have sufficient flanking sequence to choose PCR primers. The proportion of such CA-repeats can be estimated to be about 17% (calculations not shown). If the total number of CA-repeats in the genome is M, we would expect to see about $(319 \times 318)/(2 \times 0.17\ M) \approx 300{,}000/M$ duplicate clones

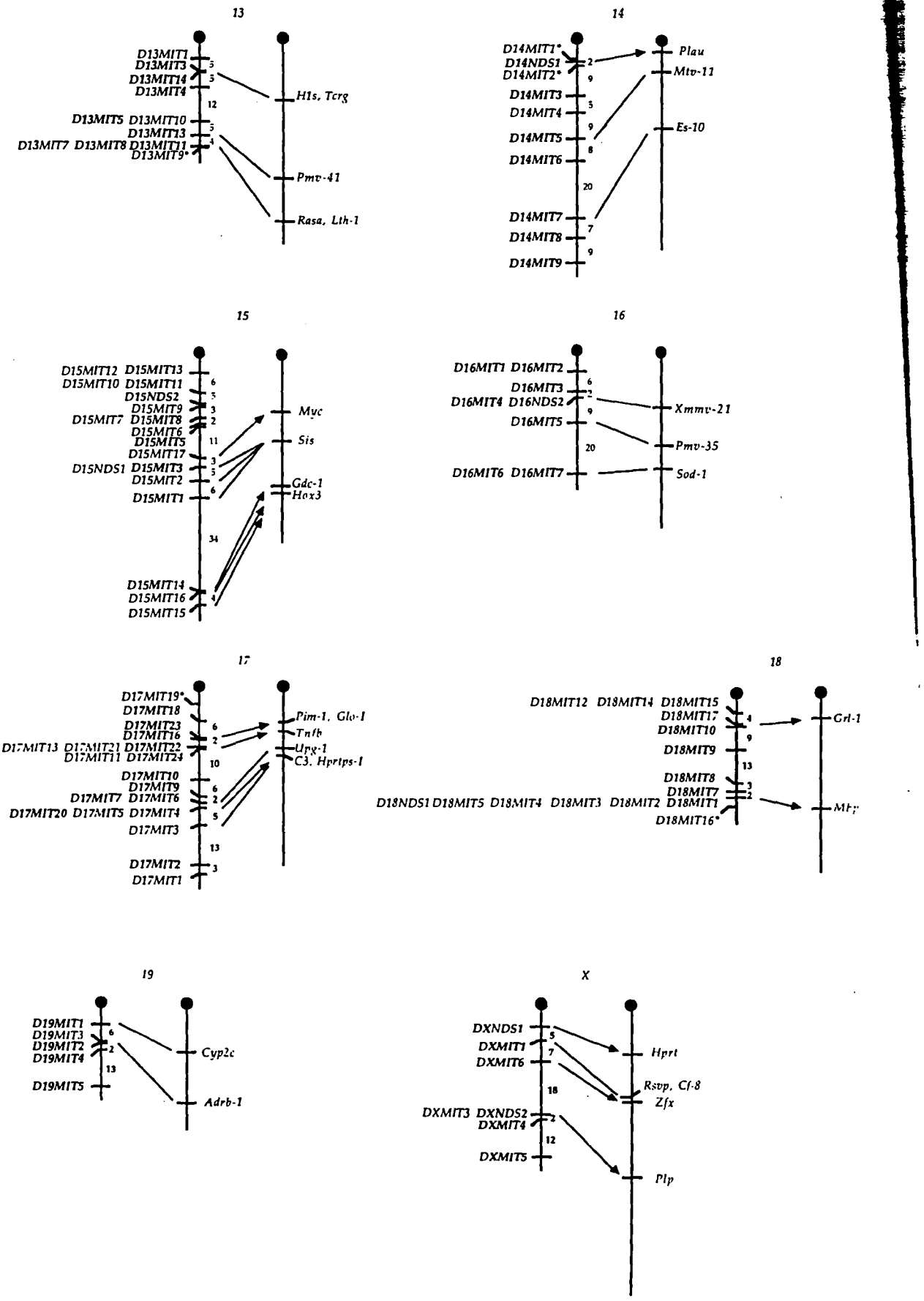


FIGURE 2.—Continued

TABLE 4

BXD strain distribution patterns

Locus name	Assay name	1	2	5	6	8	9	11	12	13	14	15	16	18	19	20	21	22	23	24	25	27	28	29	30	31	32
D1Mit1	L33	U	U	U	D	D	B	D	D	B	B	D	D	B	D	D	D	D	B	B	D	B	D	D	U	U	U
D1Mit5	L20	U	B	B	D	D	B	B	B	B	B	D	D	B	D	B	B	B	B	B	D	D	D	D	U	U	U
D1Mit7	A80	U	B	B	D	D	B	U	B	D	B	D	B	B	B	U	B	B	D	B	D	D	D	B	U	U	U
D1Mit11	M17	U	B	D	D	B	B	D	B	D	B	D	B	D	B	D	B	D	D	D	D	B	D	B	U	U	U
D1Mit16	L46	U	B	D	B	B	**	D	B	D	B	D	B	B	B	B	B	D	D	D	D	D	D	D	U	U	U
D1Mit17	M41	U	B	D	D	D	D	B	B	D	D	D	B	B	B	B	D	D	D	D	B	D	B	D	U	U	U
D1Mit19	L86	U	B	B	D	D	B	B	B	D	B	D	B	B	B	B	B	D	D	B	D	D	D	B	U	U	U
D2Mit6	L18	U	B	U	B	D	D	B	D	D	D	B	B	D	B	B	D	D	B	D	D	B	B	B	U	U	U
D2Mit7	L44	U	B	D	B	B	D	B	B	D	D	B	B	B	D	B	D	D	B	D	D	B	B	B	U	U	U
D2Mit9	M85	U	D	D	B	B	D	B	B	D	D	B	B	B	D	D	D	B	D	B	B	B	B	B	U	U	U
D2Mit12	M179	U	D	D	B	D	B	D	B	D	D	B	D	D	D	D	D	B	D	D	B	D	D	D	U	U	U
D2Mit14	M163	U	B	D	B	B	**	D	D	D	D	B	D	D	D	D	D	B	D	D	B	B	B	D	U	U	U
D2Mit17	M246	U	B	B	B	D	B	D	D	B	B	D	D	D	D	D	D	B	D	D	B	D	D	D	U	U	U
D2Mit30	D111	U	B	B	B	D	B	D	D	B	B	D	D	D	D	D	D	U	D	D	B	D	U	D	U	U	U
D2Nds1	T19	U	D	D	B	B	D	D	D	D	D	B	B	D	D	D	D	B	D	D	B	D	B	D	U	U	U
D3Mit5	M123	U	D	B	B	D	D	D	D	B	D	B	B	D	D	B	D	B	B	B	B	D	D	B	U	U	U
D3Mit9	A85	U	B	D	B	B	B	D	B	B	B	D	B	B	B	B	D	D	B	B	D	B	D	B	U	U	U
D3Mit10	A34	U	B	B	B	D	B	D	B	B	B	D	B	B	B	B	D	D	B	B	D	B	D	B	U	U	U
D3Mit12	A60	U	D	B	B	D	B	D	B	B	B	D	**	B	B	B	D	D	B	B	D	B	D	B	U	U	U
D3Mit15	A55	U	D	B	B	D	B	D	D	D	B	B	B	D	D	D	D	B	D	B	D	B	D	B	U	U	U
D3Mit17	M235	U	B	B	D	D	D	D	B	D	B	B	B	D	D	D	D	B	B	B	D	B	D	B	U	U	U
D3Mit19	M141	U	D	B	D	B	B	D	D	D	B	D	B	D	B	D	B	D	B	B	D	B	D	B	U	U	U
D3Mit21	D31	U	D	B	B	D	B	D	D	D	D	D	B	D	D	B	D	B	B	B	B	D	D	B	U	U	U
D3Mit22	D122	U	B	D	B	B	B	D	D	B	B	D	B	B	B	B	D	D	B	B	D	B	D	B	U	U	U
D3Nds2	T21	U	D	B	B	D	B	D	D	D	B	B	B	D	D	D	D	B	D	B	D	B	D	B	U	U	U
D4Mit12	M15	U	B	D	B	D	D	B	D	D	D	D	D	D	D	B	D	**	D	B	D	D	D	D	U	U	U
D4Mit13	M169	U	B	D	B	D	B	B	D	D	D	D	D	D	B	B	D	B	D	B	B	B	D	B	U	U	U
D4Mit14	A69	U	B	D	B	D	B	B	D	D	D	B	D	D	B	B	D	B	B	B	B	B	D	B	U	U	U
D4Mit16	A65	U	B	D	B	D	D	B	D	D	D	B	D	D	B	B	D	B	D	B	D	D	D	B	U	U	U
D4Mit17	D1	U	B	B	B	B	B	B	B	D	B	D	B	B	B	B	D	D	B	D	D	D	B	B	U	U	U
D5Mit1	A82	U	D	D	D	B	D	U	D	B	B	B	D	B	D	D	B	B	B	B	B	D	B	D	U	U	U
D5Mit7	M154	U	D	D	B	B	B	B	D	B	D	B	D	B	B	B	D	D	B	D	B	B	B	D	U	U	U
D5Mit10	M207	U	D	D	B	B	B	D	D	B	D	B	D	D	B	B	D	D	B	D	U	B	B	D	U	U	U
D5Mit11	M97	U	U	D	B	B	B	B	B	B	B	B	D	B	D	D	D	D	B	D	B	B	D	B	U	U	U
D6Mit9	L23	U	B	D	B	D	D	D	B	D	B	B	B	B	D	B	D	D	B	D	B	D	B	B	U	U	U
D6Mit10	M78	U	B	D	B	B	B	D	B	D	B	D	B	B	D	B	B	B	B	D	D	D	B	D	U	U	U
D6Mit13	D34	U	B	D	B	B	D	D	B	D	B	D	B	B	D	B	B	B	B	D	D	D	B	D	U	U	U
D6Mit14	M190	U	B	D	B	D	D	B	B	D	B	D	B	D	D	B	B	B	D	B	D	D	B	D	U	U	U
D6Mit15	M148	U	B	D	B	D	D	B	B	D	B	D	B	D	D	B	B	B	D	B	D	D	B	D	U	U	U
D6Mit16	D11	U	B	D	D	D	B	D	B	D	B	B	U	B	U	B	D	D	B	D	D	D	B	B	U	U	U
D7Mit7	L12	U	B	D	B	B	D	B	B	D	D	B	B	D	B	D	B	D	B	B	B	B	B	D	U	U	U
D7Mit12	M23	U	D	D	B	B	B	B	D	D	B	D	B	D	D	D	B	B	B	B	B	D	D	D	U	U	U
D7Mit17	M91	U	B	D	U	B	D	B	D	D	D	D	B	B	B	B	B	B	U	U	D	U	D	U	U	U	U
D7Nds2	T28	U	**	D	B	B	D	B	D	**	D	B	B	B	B	B	B	D	B	B	D	B	B	B	U	U	U
D8Mit4	M71	U	B	B	B	B	B	D	B	D	B	B	D	D	B	B	B	D	B	D	B	B	B	D	U	U	U
D8Mit8	M257	U	B	D	B	B	D	D	B	D	B	B	D	D	D	B	B	B	B	D	B	B	B	B	U	U	U
D8Mit9	A62	U	B	B	B	D	D	D	B	D	B	B	D	D	D	B	B	D	B	D	B	U	B	B	U	U	U
D8Mit11	A105	U	B	D	D	D	D	D	B	D	B	B	D	D	B	D	B	B	B	D	B	D	D	B	U	U	U
D9Mit4	M151	U	D	B	D	D	D	B	D	B	D	B	D	D	D	B	D	B	D	D	D	B	B	D	U	U	U
D9Mit8	M211	U	B	D	D	D	D	B	D	B	B	B	D	D	D	B	D	D	B	B	D	B	B	D	U	U	U
D9Mit11	L60	U	B	B	D	D	B	B	D	B	B	B	D	D	B	B	D	D	B	B	D	B	B	B	U	U	U
D9Mit12	M73	U	B	B	D	D	B	B	D	B	B	B	B	D	B	D	D	D	B	B	D	B	B	B	U	U	U
D9Mit15	M160	U	D	B	B	D	B	B	B	B	B	B	B	D	B	D	B	D	D	B	D	D	B	B	U	U	U
D9Mit18	M10	U	D	B	D	B	B	B	B	B	B	B	D	B	B	B	D	D	D	B	B	B	B	B	U	U	U
D9Mit19	M157	U	D	B	D	B	B	B	B	B	B	B	D	B	B	B	D	D	D	B	B	B	B	B	U	U	U
D9Mit20	L64	U	D	B	B	D	B	B	B	B	B	B	B	D	B	D	D	D	B	B	D	B	B	B	U	U	U
D9Mit21	D15	U	D	D	D	D	D	B	D	B	D	B	D	D	D	B	D	B	D	B	D	B	B	D	U	U	U
D9Nds2	T30	U	B	B	D	D	D	B	B	D	B	B	D	D	D	B	D	D	B	B	D	B	B	D	U	U	U
D10Mit3	A114	U	D	D	B	B	B	B	B	B	D	D	B	B	B	B	D	B	D	D	D	B	D	B	U	U	U
D10Mit10	M7	U	D	D	D	D	B	D	B	B	D	B	D	D	B	B	D	B	D	D	D	D	D	B	U	U	U
D10Mit11	A88	U	D	D	D	D	B	D	B	B	D	B	D	D	B	B	D	B	D	D	D	D	D	B	U	U	U
D10Mit14	M175	U	D	B	D	B	B	B	B	B	B	B	D	D	D	D	D	B	D	D	D	D	D	B	U	U	U
D10Mit15	D30	U	D	D	B	D	B	B	B	**	D	B	D	D	B	B	D	B	D	D	B	B	D	B	U	U	U
D11Mit2	L14	U	B	B	D	B	D	B	B	D	D	B	U	D	B	D	B	B	D	D	B	D	D	D	U	U	U
D11Mit4	A124	U	D	B	D	B	B	B	B	D	B	B	D	B	D	B	D	B	B	B	B	B	D	D	U	U	U
D11Mit14	D2	U	D	D	D	D	D	B	B	B	D	D	D	B	D	D	D	D	B	D	B	B	B	D	U	U	U
D11Mit15	D5	U	D	B	D	B	B	B	B	D	B	D	D	B	D	U	D	B	B	B	B	B	D	D	U	U	U
D12Mit1	M50	U	D	B	B	D	D	B	D	D	D	B	B	B	B	D	D	D	D	D	B	D	B	B	U	U	U
D12Mit2	M27	U	D	B	B	B	D	B	B	B	B	D	B	B	D	D	D	D	D	D	B	D	B	D	U	U	U
D12Mit3	L41	U	B	B	B	D	D	B	D	B	B	D	B	B	B	D	B	B	D	D	D	D	D	D	U	U	U
D12Mit5	L58	U	B	B	B	D	B	B	D	D	B	D	B	B	B	D	B	B	B	D	D	D	D	B	U	U	U
D12Mit7	M62	U	B	B	D	D	D	D	B	B	B	D	D	B	D	D	B	B	B	D	D	D	D	B	U	U	U
D12Mit8	D7	U	B	B	D	B	D	D	D	B	B	D	D	D	B	D	B	B	B	D	B	D	D	B	U	U	U
D12Nds2	T1	U	B	B	D	B	D	D	D	B	B	D	D	D	B	D	B	B	B	D	B	B	D	B	U	U	U
D13Mit3	M79	U	B	B	B	B	B	B	D	D	D	B	B	D	B	U	B	B	D	B	D	D	B	D	U	U	U
D13Mit9	M147	U	D	B	B	B	D	D	D	D	D	D	D	B	B	D	D	B	D	D	B	D	D	B	U	U	U
D13Mit11	A91	U	D	B	D	B	D	D	D	D	D	D	D	B	B	D	D	B	D	D	B	D	B	B	U	U	U
D13Mit13	D24	U	D	B	D	B	D	B	D	D	D	B	D	B	B	D	D	B	B	D	D	D	D	B	U	U	U
D14Mit1	A103	U	D	B	D	B	D	D	B	B	B	D	D	D	D	B	D	D	B	D	B	B	B	D	U	U	U
D14Mit2	A24	U	D	B	D	B	D	D	B	B	B	D	B	D	D	B	D	D	B	B	B	B	B	D	U	U	U
D14Mit5	M214	U	D	B	D	D	D	D	B	B	D	D	D	D	D	B	B	D	B	B	B	B	B	D	U	U	U
D14Mit6	A119	U	D	B	D	D	D	D	B	B	D	D	D	B	D	B	B	D	B	B	D	B	B	D	U	U	U

TABLE 4—Continued

Locus name	Assay name	1	2	5	6	8	9	11	12	13	14	15	16	18	19	20	21	22	23	24	25	27	28	29	30	31	32
D14Mit7	L27	U	D	B	D	D	D	D	B	B	D	D	U	B	D	B	B	B	D	B	D	B	B	D	U	U	U
D14Nds1	T10	U	D	B	D	B	D	D	B	B	B	D	B	D	D	B	D	D	B	B	B	B	B	D	U	U	U
D15Mit1	L29	U	B	D	B	D	D	B	B	D	D	D	B	D	B	D	D	D	B	D	B	D	D	B	U	U	U
D15Mit2	L10	U	D	D	B	D	D	B	B	D	D	D	B	D	D	D	D	D	B	D	B	D	D	B	U	U	U
D15Mit3	L78	U	B	B	B	D	D	B	B	D	D	B	B	D	D	B	D	B	B	D	B	D	D	B	U	U	U
D15Mit5	L1	U	D	B	B	B	D	B	B	B	D	B	B	D	B	B	D	B	B	D	B	D	D	B	U	U	U
D15Mit6	A59	U	D	B	B	B	D	B	B	B	D	B	B	D	B	B	D	B	B	**	B	D	D	B	U		U
D15Mit7	M30	U	D	B	B	B	D	B	B	B	D	B	B	D	B	B	D	B	B	D	B	D	D	B	U		U
D15Mit8	A79	U	D	B	B	B	D	B	B	B	D	B	B	D	B	B	D	B	B	D	B	D	D	B	U		U
D15Mit12	M34	U	D	B	B	B	D	B	B	B	D	B	B	D	B	B	D	B	B	D	D	D	**	B	U	U	U
D15Mit13	A36	U	D	B	B	B	D	B	B	B	D	B	B	D	B	B	D	B	B	D	D	D	D	B	U	U	U
D15Mit15	D6	U	B	D	B	D	D	B	D	D	B	B	B	D	B	D	D	B	B	D	B	D	B	D	U	U	U
D15Nds2	T18	U	D	D	B	B	D	B	B	B	D	B	B	D	B	B	D	B	B	D	D	D	D	B	U	U	U
D16Mit3	M127	U	B	B	B	D	B	D	B	B	B	B	D	B	D	D	B	B	U	U	D	U	B	D	U	U	U
D16Mit4	M203	U	B	B	B	D	B	D	B	B	B	B	D	B	D	D	B	B	D	D	D	D	B	D	U	U	U
D16Mit5	A38	U	B	B	B	D	B	D	B	B	B	D	D	B	B	D	D	B	B	D	D	D	B	U	U	U	U
D16Mit6	L7	U	B	B	B	B	D	D	B	B	D	D	B	B	D	D	B	B	B	D	D	D	B	D	U	U	U
D17Mit3	L28	U	U	D	B	B	B	D	B	D	D	B	B	D	B	D	B	D	D	D	B	D	D	B	U	U	U
D17Mit6	M254	U	D	D	D	B	B	D	B	D	B	B	D	D	B	B	B	D	B	D	D	D	D	B	U	U	U
D17Mit7	L4	U	D	D	D	B	B	D	B	D	B	B	D	D	B	B	B	D	D	D	B	D	D	B	U	U	U
D17Mit7	A23	U	D	D	D	B	B	D	B	D	B	B	D	D	B	B	B	D	D	D	B	D	D	B	U	U	U
D17Mit10	L36	U	D	D	D	B	B	D	B	D	B	B	D	D	B	B	B	D	B	D	D	D	D	B	U	U	U
D17Mit11	M145	U	B	D	D	B	U	D	B	B	B	B	D	D	B	B	D	D	B	D	D	D	D	B	U	U	U
D17Mit13	L57	U	B	D	D	B	D	D	D	B	B	B	D	D	B	B	D	D	B	D	D	D	D	B	U	U	U
D17Mit16	A25	U	B	D	D	B	D	D	D	B	B	B	D	D	B	B	D	D	B	D	D	D	D	B	U	U	U
D17Mit21	D21	U	B	D	D	B	D	D	D	B	B	B	D	D	B	B	D	D	B	D	D	D	D	B	U	U	U
D17Mit22	D16	U	B	D	D	B	D	D	D	B	B	B	D	D	B	B	D	D	B	D	D	D	D	B	U	U	U
D17Mit24	D12	U	B	D	D	B	D	D	D	B	B	B	D	D	B	B	D	D	B	D	D	D	D	B	U	U	U
D17Nds2	T9	U	B	D	D	B	D	D	D	B	B	B	D	D	B	B	D	D	B	D	D	D	D	B	U	U	
D18Mit4	M51	U	B	D	D	B	U	B	D	B	B	D	D	D	B	B	B	D	B	B	B	B	D	B	U	U	
D18Mit7	M108	U	B	B	D	B	D	B	D	B	B	D	D	D	B	D	B	D	B	D	B	B	B	B	U	U	
D18Mit8	L24	U	D	B	D	B	D	B	D	B	B	D	D	D	B	D	B	D	D	D	B	B	B	B	U	U	U
D18Mit9	M209	U	D	B	D	B	D	B	D	D	D	D	D	D	B	B	B	D	D	D	B	B	B	B	U	U	U
D18Mit10	A100	U	D	D	D	D	D	D	D	B	D	D	D	D	B	D	D	D	B	B	B	B	B	B	U	U	U
D18Mit14	L13	U	D	B	D	B	D	D	D	D	D	B	B	D	B	D	B	D	B	B	B	B	B	B	U	U	U
D18Mit15	L87	U	D	B	D	B	D	D	D	D	B	D	B	B	B	D	B	D	B	B	B	B	B	B	U	U	U
D18Mit17	D118	U	D	B	D	B	D	D	D	D	D	B	D	D	B	D	B	D	B	B	B	B	B	B	U	U	U
D19Mit1	A17	U	B	B	D	D	B	D	B	B	B	B	B	B	D	B	B	B	D	B	B	B	B	B	U	U	U
DXMit1	L43	U	B	B	D	B	D	B	B	B	B	B	B	B	B	D	B	B	B	B	B	B	D	B	U	U	U

The strains carrying the C57BL/6J allele are denoted by B and those carrying the DBA/2J allele are denoted by D. Strains whose allele was not determined are denoted by U. Mutant alleles, differing from both B and D, are denoted by **.

TABLE 5

Locations for previously unmapped genes

Name	Sequence	Chromosome	Reference
Trh-1 (*D2Mit30*)	His-t-RNA	*2*	MORRY and HARDING (1986)
Ace (*D11Mit13*)	Angiotensin converting enzyme	*11*	BERNSTEIN *et al.* (1989); HOWARD *et al.* (1990)
Snap (*D2Mit28*)	Synaptosomal associated protein 25	*2*	OYLER *et al.* (1991)
Rpl-32ps (*D3Mit22*)	Ribosomal protein L32′ (pseudogene)	*3*	JACKS, POWASER and HACKETT (1988)
Sqr-3 (*D10Mit15*)	Simple quadruplet repeat. pmlc3	*10*	SCHAFER *et al.* (1986)
Lif (*D11Mit16*)	Leukemia inhibitory factor	*11*	STAHL *et al.* (1990)
Antp91a (*D11Mit14*)	Tum⁻P91A antigen	*11*	LURQUIN *et al.* (1989)
Sqr-4 (*D13Mit14*)	Simple quadruplet repeat. pmlc4	*13*	SCHAFER *et al.* (1986)
Svp-4 (*D2Mit29*)	Seminal vesicle secretory protein IV	*2*	CHEN *et al.* (1991)
Mb-1 (*D7Mit20*)	Murine b-cell 1	*7*	KASHIWAMURA *et al.* (1990)

arising. (Actually, a small proportion of the clones were selected from GenBank and thus could not duplicate one another. However, this affects the estimate only slightly.)

After completing the map, we examined our data and found, in fact, six duplicate SSRs, defining the loci: *D3Mit13*, *D4Mit2*, *D6Mit6*, *D7Mit10*, *D17Mit7* and *D18Mit1*. In at least three of these cases, we can be certain that the clones were independent–either because they arose in libraries constructed at different times or because their sequences were from complementary strands. As should be the case, the independent typings of the duplicate loci showed no recombination. The number of duplicates is consistent with the genome containing about 50,000 distinct CA-repeat-containing SSLPs, which broadly agrees with previous estimates of the total number of CA-repeats in the genome (HAMADA and TAKUNAGA 1982). Although the number of duplicates is quite small, we plan to adjust our protocol in further work to check for duplicates immediately after sequencing and to use randomly sheared DNA inserts to decrease their frequency.

DISCUSSION

Utility of maps based on SSLPs: Simple sequence length polymorphisms are rapidly becoming a method

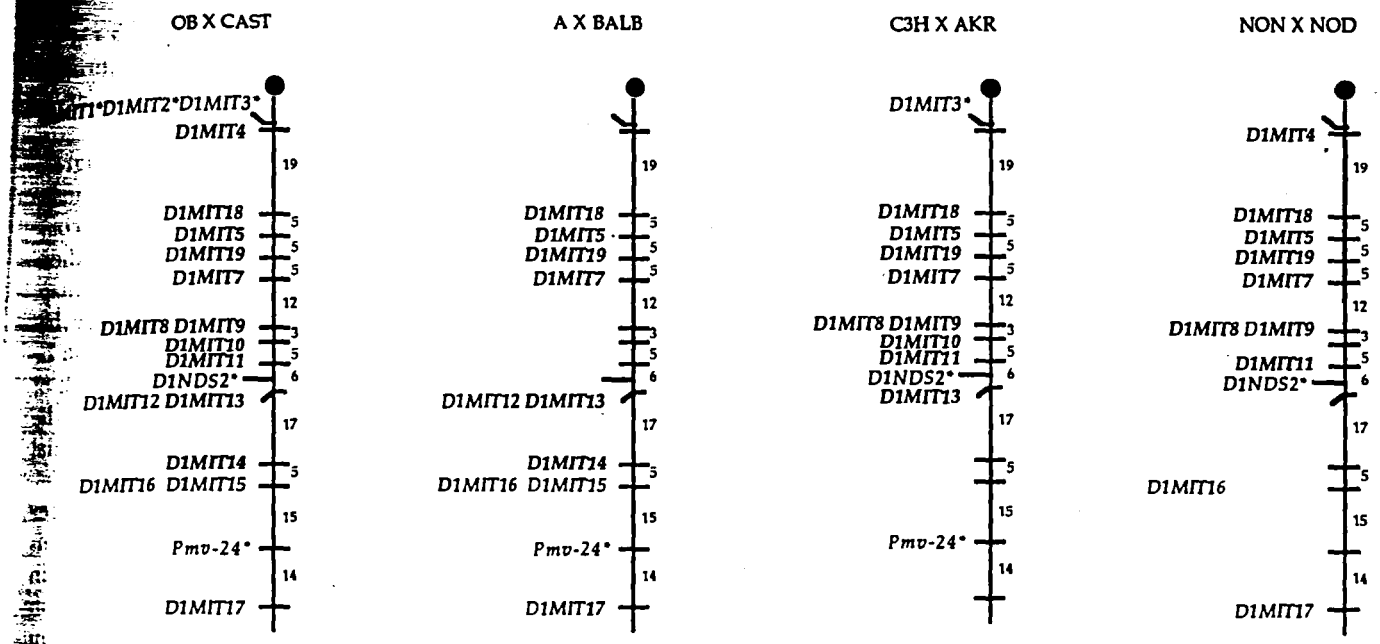


FIGURE 3.—Illustration of high polymorphism rate in different crosses. For chromosome *1*, the diagrams shows those SSLPs that are polymorphic in four typical crosses.

of choice for genetic mapping in human, mouse and rat, due to their exceptionally high rate of polymorphism and their relative ease of use. In humans, the high degree of polymorphism helps overcome the difficulties inherent in studying families in randomly breeding populations. In mouse and rat, the markers make it feasible to map the entire genome in any cross between laboratory strains; this has begun to allow genetic dissection of polygenic traits such as type I diabetes (TODD *et al.* 1991) and hypertension (JACOB *et al.* 1991; HILBERT *et al.* 1991).

We have developed a genetic map of the mouse consisting of 317 SSLP markers, with an average spacing of about 4.3 cM. Although the map was constructed in a cross between two divergent subspecies of *M. musculus* (OB × CAST), it can now be applied to map genes in most intraspecific crosses. Some 50% of the markers are polymorphic in a typical cross between two inbred laboratory strains, providing a genetic map with an average spacing of less than 9 cM. This is illustrated in Figure 3, showing the coverage of chromosome *1* in various crosses.

We hope that the map will prove useful to mouse geneticists. Because our map is anchored relative to the existing mouse map, it should be straightforward to identify the SSLPs in specific regions of interest. Additional anchor points will be added over time, by our laboratory and others. Because we have developed a dense collection of highly polymorphic SSLPs that work under a single set of PCR conditions, it should be possible to choose a relatively small subset of markers that are informative in any cross of interest and span the genome. In this fashion, it should be feasible for mouse geneticists rapidly to map any monogenic trait, as well as to undertake genetic dissection of polygenic traits. Indeed, all the laboratory work involved in constructing the map reported here was accomplished by two of us (W. DIETRICH and H. KATZ) in less than 18 months, and we have been able to apply it to genotype new crosses for the entire genome in a few weeks per cross (W. DIETRICH, unpublished results).

In addition to their utility in genetic mapping, the SSLPs should be valuable for studies of loss of heterozygosity (LOH) in murine tumors. Apart from the fact that DNA polymorphisms are generally useful in recognizing LOH, SSLPs offer the advantage that only a small tissue sample is required for PCR typing. This may be especially valuable in the case of tumors that must be dissected carefully from surrounding tissue.

Also, SSLPs may be useful in population genetic and evolutionary studies. For example, we note that the rate of polymorphism ranges from a low of about 32% for closely related strains such as DBA/2J and C3H/HeJ, A/J and C3H/HeJ, or A/J and Balb/cJ, to more than 50% for more distantly related strains such as C57BL/6J and AKR/J, or LP/J and A/J, to about 90% for intersubspecific and interspecific comparisons. This suggests that SSLPs may offer considerable power in tracing gene flow in closely related populations and may also offer advantages over simple nucleotide substitutions in reconstructing phylogenies (ATCHLEY and FITCH 1991) because they mutate more rapidly.

Coverage of the genome: The map appears to cover the vast majority of the mouse genome. One way to assess the coverage of the map is simply to observe

TABLE 6

Genetic length of mouse chromosomes

Chromosome	Consensus genetic length (cM)	Map reported in this paper (cM)
1	98	111
2	101	90
3	100	61
4	81	67
5	84	38
6	74	71
7	89	77
8	81	67
9	70	73
10	78	71
11	78	89
12	73	73
13	72	65
14	49	69
15	56	82
16	58	37
17	36	50
18	57	32
19	36	22
X	88	33
Total	1459	1267

The lengths represent genetic distance between most proximal and most distal markers. The consensus genetic length is from consensus map in Encyclopedia of the Mouse Genome, 1990. See text for description.

that only 2 of 319 markers failed to show linkage to our map. Another way is to compare our map to the consensus map reported in the GBASE database (Table 6). Of course, the two maps would not be expected to agree perfectly because genetic distance is known to be affected by strain background: our map is constructed in a single cross between two subspecies, while the GBASE consensus map represents a complex weighted average of a variety of different crosses. Nonetheless, the maps are colinear and the correspondence between them is good: our map shows a genetic length of 1267 cM contained between the most terminal markers, compared to a length of 1459 cM between the most terminal markers in the GBASE map. The difference amounts to an average of 5 cM lying beyond the most terminal marker at each end of the 20 mouse chromosomes.

A few specific features deserve mention.

The map has a few large intervals. Chromosome *15* contains the largest interval, of about 34 cM. Interestingly, the genetic length of this interval in the GBASE map appears to be only about 17 cM, suggesting enhanced recombination in this interval in our cross. The next largest interval is about 28 cM on chromosome *11*. Mathematically, an interval of this size would be expected by chance assuming a random distribution of markers.

Comparison with the GBASE map suggests that the terminal regions of most chromosomes are well covered, with the exception of the distal 20–25 cM on chromosome 5 and the distal 15–20 cM on chromosome 13. These intervals are not significantly larger than would be expected by chance.

Although the total length of chromosome 3 agrees well between our map and the GBASE consensus map, the region from Il-2 to Xmmv-65 seems to be compressed. Our map shows about 15 cM between these markers, compared to 40 cM on the GBASE map. This might be due to structural heterogeneity between OB and CAST chromosomes such as one or more inversions, although there is no large block of recombinationally inseparable markers as might be expected from a single large inversion. Additional anchors will be needed to resolve this.

Chromosome 18 shows an unusually large cluster of recombinationally unseparated markers. This might be due to an inversion or to a heterogeneity in the distribution of SSRs. The anchoring information suggests that the entire chromosome is represented in the map.

Random distribution of markers: Broadly speaking, the genetic markers appear to be randomly distributed throughout the genome. One way to assess this is to compare to the number of markers that would be expected to fall on each chromosome based on its physical size (estimated by cytogenetic length) to the number actually observed. (In this calculation, we must account for the fact that the genomic library used to isolated SSRs was made from a male mouse. We thus expect a twofold underrepresentation of the *X* chromosome.) The agreement is excellent (Table 7). Only chromosome *17* shows a significant deviation from expectation. In fact, the deviation is explained by the disproportionate number of SSLPs derived from cloned genes in GenBank on chromosome *17* (specifically, 7 of the 54 SSLPs derived from GenBank sequences in our map) owing to the extensive study of this chromosome, which is the site of the *major histocompatibility complex* and the *t complex*.

Another way to assess whether the markers are randomly distributed is to compare the observed distribution of distances between adjacent markers to that expected under the assumption that SSRs are randomly distributed across the genetic map (see MATERIALS AND METHODS). The distributions agree quite well (Figure 4). There appears to be a slight excess of zero distances–the proportion of pairs of adjacent loci that showed no recombination in our cross was 25.1% compared to an expectation of 20.3%±2.4%–but the deviation is just at the edge of statistical significance. This might hint at slight clustering of SSRs with respect to genetic distance, which could be due to uneven spacing of either SSRs or recombination along the physical map.

In short, the assumption of random distribution of

TABLE 7

Number of markers on each chromosome

Chromosome	Percent of genome based on physical map[a]	No. markers expected ±1 SD[b]	Markers in this paper	Z-score[c]
1	7.20	23.6 ± 4.7	19	−0.98
2	6.95	22.8 ± 4.6	30	1.58
3	5.99	19.6 ± 4.3	21	0.32
4	5.89	19.3 ± 4.3	19	−0.07
5	5.68	18.6 ± 4.2	12	−1.58
6	5.53	18.1 ± 4.1	20	0.46
7	5.19	17.0 ± 4.0	20	0.75
8	4.97	16.3 ± 3.9	14	−0.58
9	4.79	15.7 ± 3.9	21	1.38
10	4.74	15.5 ± 3.8	17	0.39
11	4.72	15.5 ± 3.8	16	0.14
12	4.88	16.0 ± 3.9	11	−1.28
13	4.38	14.3 ± 3.7	11	−0.90
14	4.46	14.6 ± 3.7	10	−1.23
15	4.05	13.3 ± 3.6	18	1.33
16	3.81	12.5 ± 3.5	8	−1.29
17	3.86	12.6 ± 3.5	22	2.69
18	3.88	12.7 ± 3.5	16	0.94
19	2.73	8.9 ± 2.9	5	−1.34
X	6.23	10.2 ± 3.1	7	−1.02

[a] Based on cytogenetic length EVANS (1989).

[b] Based on proportional size of each chromosome, adjusted for the *X* chromosome being at half-molar representation (since the vast majority of markers were isolated from a genomic library from male DNA).

[c] Z-score = (observed-expected)/standard deviation.

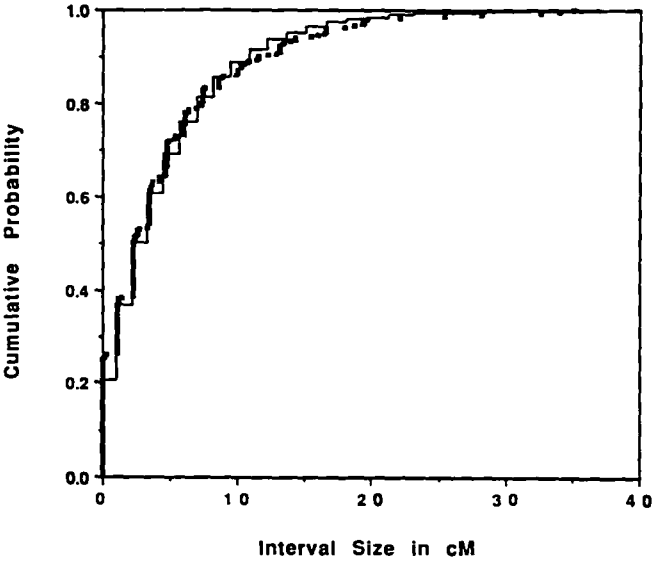


FIGURE 4.—Cumulative probability distribution of interval sizes in the genetic map. Points show observed cumulative distribution for intervals in our map. The solid line represents the expected distribution, assuming that SSLPs are randomly distributed with respect to centimorgans (see MATERIAL AND METHODS for formula). Note that the distributions of interval sizes is expected to show discrete jumps, because only a finite number *N* of meioses are studied and thus recombination fractions will be approximately integral multiples of $1/N$.

SSRs fits the data reasonably well at this level of resolution, although there may hints of clustering. Of course, significant inhomogeneity may become apparent at higher resolution. These findings bode well for the general usefulness of SSRs in the construction of genetic maps in other organisms, including the human.

Toward a dense genetic map of the mouse: The approach described here should allow the construction of much denser maps consisting of thousands of SSLPs. Indeed, SSLPs appear to be in abundant supply and to be randomly distributed throughout the genome–at least at the level of resolution examined here. With a genetic linkage map of 3000 SSLPs, one would have genetic landmarks at an average spacing of 1 million basepairs. Coupled with high quality yeast artificial chromosome libraries, such a dense collection of landmarks would permit rapid and straightforward cloning of the region containing any gene of interest and should greatly advance the genetic understanding of mammalian biology.

We thank GEORGE CHURCH, MARK DALY, NAT GOODMAN, HOWARD JACOB, YI-PEI MAO, MARY PAT REEVE and JULIA SEGRE for valuable advice and assistance at various stages of the project; JOHN COFFIN, WAYNE FRANKEL, DANIKA METALLINOS, JOE NADEAU, JOHN TODD, MICHAEL SELDIN, BEN TAYLOR and ROGER WISEMAN for sharing information and resources and for helpful discussions; and JIM HUDSON and the staff of Research Genetics, Inc. for consistent and reliable supply of a large number of oligonucleotides. This work was supported in part by grants to from the National Institutes of Health (P50HG00098 and R01HG00126 to E.S.L. and R01HG00316 to J.F. and E.S.L.), the National Science Foundation (D1R8611317 to E.S.L.), and the Markey Foundation (to E.S.L.).

Note added in proof: The locus *D18Mit6* was omitted in Figure 2. It did not recombine with *D18Mit1*.

LITERATURE CITED

ALTSCHUL, S. F., W. GISH, W. MILLER, E. MYERS and D. J. LIPMAN, 1990 A basic local alignment search tool. J. Mol. Biol. **215:** 403–410.

ATCHLEY, W. R., and W. M. FITCH, 1991 Gene trees and the origins of inbred strains of mice. Science **254:** 554–558.

AVNER, P., L. AMAR, L. DANDOLO and J. L. GUENET, 1988 Genetic analysis of the mouse using interspecific crosses. Trend Genet. **4:** 18–23.

BAHARY, N., G. ZORICH, J. E. PACHTER, R. L. LEIBEL and J. M. FRIEDMAN, 1991 Molecular Genetic linkage maps of mouse chromosomes *4* and *6*. Genomics **11:** 33–47.

BAILEY, D. W., 1971 Recombinant inbred strains. An aid to finding identity, linkage, and function of histocompatibility and other genes. Transplantation **11:** 325–327.

BERNSTEIN, K. E., B. M. MARTIN, A. S. EDWARDS and E. A. BERNSTEIN, 1989 Mouse angiotensin-converting enzyme is a protein composed of two homologous domains. J. Cell Biol. **264:** 11949–11951.

BOTSTEIN, D., R. L. WHITE, M. SKOLNICK and R. W. DAVIS, 1980 Construction of a genetic linkage map in man using restriction fragment length polymorphism. Am. J. Hum. Genet. **32:** 314–331.

BRESLAUER, K. J., R. FRANK, H. BLOCKER and L. A. MARKY, 1986 Predicting DNA duplex stability from the base se-

quence. Proc. Natl. Acad. Sci. USA **83:** 3746–3750.

BUCHBERG, A. M., E. BROWNELL, S. NAGATA, N. A. JENKINS and N. G. COPELAND, 1989 A comprehensive genetic map of murine chromosome *11* reveals extensive linkage conservation between mouse and human. Genetics **122:** 153–161.

BUETOW, K., 1991 Influence of aberrant outcomes on high-resolution linkage outcomes. Am. J. Hum. Genet. **49:** 985–994.

CECI, J. D., L. D. SIRACUSA, N. A JENKINS and N. G. COPELAND, 1989 A molecular genetic linkage map of mouse chromosome *4* including the localization of several proto-oncogenes. Genomics **5:** 699–709.

CECI, J. D., M. J. JUSTICE, L. F. LOCK, N. A. JENKINS and N. G. COPELAND, 1990a An interspecific backcross linkage map of mouse chromosome *8*. Genomics **6:** 72–79.

CECI, J. D., D. M. KINGSLEY, C. M. SILAN, N. G. COPELAND and N. A. JENKINS, 1990b An interspecific backcross linkage map of the proximal half of mouse chromosome *14*. Genomics **6:** 673–678.

CHANG, C., J. L. BOWMAN, A. W. DEJOHN, E. S. LANDER and E. MEYEROWITZ, 1988 An RFLP linkage map that facilitates gene cloning in *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA **85:** 6856–6860.

CHEN, Y. H., B. T. PENTECOST, J. A. MCLACHLAN and C. T. TENG, 1991 The androgen-dependent mouse seminal vesicle secretory protein IV: Characterization and complementary deoxyribonucleic acid cloning. Mol. Endocrinol. **1:** 707–716.

CHURCH, G., and W. GILBERT, 1984 Genomic sequencing. Proc. Natl. Acad. Sci. USA **81:** 1991–1995.

COPELAND, N. G., and N. A. JENKINS, 1991 Development and applications of a molecular genetic linkage map of the mouse genome. Trends Genet. **7:** 113–118.

CORNALL, R. J., T. J. AITMAN, C. M. HEARNE and J. A. TODD, 1991 The generation of a library of PCR-analyzed microsatellite variants for genetic mapping of the mouse genome. Genomics **10:** 874–881.

DAVISSON, M. T., T. H. RODERICK and D. P. DOOLITTLE, 1989 Recombination percentages and chromosomal assignments, pp. 432–505 in *Genetic Variants and Strains of the Laboratory Mouse*, Ed. 2, edited by M. F. LYON and A. SEARLE. Oxford University Press, New York.

DONIS-KELLER, H., P. GREEN, C. HELMS, S. CARTINHOUR, B. WEIFFENBACH, K. STEPHENS, T. P. KEITH, D. W. BOWDEN, D. R. SMITH, E. S. LANDER, D. BOTSTEIN, G. AKOTS, K. S. REDIKER, T. GRAVIUS, V. A. BROWN, M. B. RISING, C. PARKER, J. A. POWERS, D. E. WATT, E. R. KAUFFMAN, A. BRICKER, P. PHIPPS, H. MULLER-KAHLE, T. R. FULTON, S. NG, J. W. SCHUMM, J. C. BRAMAN, R. G. KNOWLTON, D. F. BARKER, S. M. CROOKS, S. E. LINCOLN, M. J. DALY and J. ABRAHAMSON, 1987 A genetic linkage map of the human genome. Cell **51:** 319–337.

EICHER, E., 1981 Foundation for the future: Formal genetics of the mouse, pp. 7–49 in *Mammalian Genetics and Cancer: The Jackson Laboratory Fiftieth Anniversary Symposium*. Alan R. Liss, New York.

EVANS, E., 1989 Standard normal chromosomes: standard idiogram, pp. 576–578 in *Genetic Variants and Strains of the Laboratory Mouse*, Ed. 2, edited by M. F. LYON and A. SEARLE. Oxford University Press, New York.

FESTING, M. F. W., 1979 *Inbred Strains in Biomedical Research*. Oxford University Press, New York.

FRANKEL, W. N., J. P. STOYE, B. A. TAYLOR and J. M. COFFIN, 1990 A genetic linkage map of endogenous murine leukemia proviruses. Genetics **124:** 221–236.

GREEN, M. C., 1989 Catalog of mutant genes and polymorphic loci, pp. 12–403 in *Genetic Variants and Strains of the Laboratory Mouse*, Ed. 2, edited by M. F. LYON and A. SEARLE. Oxford University Press, New York.

HALDANE, J. B. S., A. D. SPRUNT and N. M. HALDANE, 1915 Reduplication in mice. Science **5:** 133–135.

HAMADA, H., and T. TAKUNAGA, 1982 Potential z-DNA forming sequences are highly dispersed in the human genome. Nature **298:** 396–398.

HAMADA, H., M. G. PETRINO and T. TAKUNAGA, 1982 A novel repeated element with z-DNA-forming potential is widely found in evolutionary diverse eukaryotic genomes. Proc. Natl. Acad. Sci. USA **79:** 6465–6469.

HEARNE, C. M., M. A. MCALEER, J. M. LOVE, T. J. AITMAN, R. J. CORNALL, S. GHOSH, A. M. KNIGHT, J.-B. PRINS and J. A. TODD, 1991 Additional microsatellite markers for mouse genome mapping. Mamm. Genome **1:** 273–282.

HILBERT, P., K. LINDPAINTNER, J. S. BECKMANN, T. SERIKAWA, F. SOUBRIER, C. DUBAY, P. CARTWRIGHT, B. DE GOUYON, C. JULIER, S. TAKAHASI, M. VINCENT, D. GANTEN, M. GEORGES and G. M. LATHROP, 1991 Chromosomal mapping of two genetic loci associated with blood-pressure regulation in hereditary hypertensive rats. Nature **353:** 521–529.

HOWARD, T. E., S. SHAI, K. G. LANGFORD, B. M. MARTIN and K. E. BERNSTEIN, 1990 Transcription of testicular angiotensin-converting enzyme (ACE) is initiated within the 12th intron of the somatic ACE gene. Mol. Cell. Biol. **10:** 4294–4302.

JACKS, C. M., C. B. POWASER and P. B. HACKETT, 1988 Sequence analysis of a processed gene coding for mouse ribosomal protein L32. Gene **74:** 565–570.

JACOB, H. J., K. LINDPAINTNER, S. E. LINCOLN, K. KUSUMI, R. K. BUNKER, Y.-P. MAO, D. GANTEN, V. J. DZAU and E. S. LANDER, 1991 Genetic mapping of a gene causing hypertension in the stroke-prone spontaneously hypertensive rat. Cell **67:** 213–224.

JUSTICE, M. J., C. M. SILAN, J. D. CECI, A. M. BUCHBERG, N. G. COPLAND and N. A. JENKINS, 1990a A molecular genetic linkage map of mouse chromosome *13* anchored by the *beige* (*bg*) and *satin* (*sa*) loci. Genomics **6:** 341–351.

JUSTICE M. J., L. D. SIRACUSA, D. J. GILBERT, N. HEISTERKAMP, J. GROFFEN, K. CHADA, C. M. SILAN, N. G. COPELAND and N. A. JENKINS, 1990b A genetic linkage map of mouse chromosome *10*: localization of eighteen molecular markers using a single interspecific backcross. Genetics **125:** 855–866.

KASHIWAMURA, S.-I., T. KOYAMA, T. MATSUO, M. STEINMETZ, M. KIMOTO and N. SAKAGUCHI, 1990 Structure of the murine *mb-1* gene encoding a putative sIgM-associated molecule. J. Immunol. **145:** 337–343.

KINGSLEY, D. M., N. A. JENKINS and N. G. COPELAND, 1989 A molecular genetic linkage map of mouse chromosome *9* with regional localizations for the *Gsta*, *T3g*, *Ets-1*: and *Ldlr* loci. Genetics **123:** 165–172.

KRAYEV, A. S., D. A. KRAMEROV, K, G. SKRYABIN, A. P. RYSKOV, A. A. BAYEV and G. P. GEORGIEV, 1980 The nucleotide sequence of the ubiquitous repetitive DNA sequence B1 complementary to the most abundant class of mouse fold-back RNA. Nucleic Acids. Res. **8:** 1201–1215.

KRAYEV, A. S., T. V. MARKUSHEVA, D. A. KRAMEROV, A. P. RYSKOV, K. G. SKRYABIN, A. A. BAYEV and G. P. GEORGIEV, 1982 Ubiquitous transposon-like repeats B1 and B2 of the mouse genome: B2 sequencing. Nucleic Acids Res. **10:** 7461–7475.

LANDER, E. S., P. GREEN, J. ABRAHAMSON, A. BARLOW, M. DALY, S. LINCOLN and L. NEWBURG, 1987 MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics **1:** 174–181.

LINCOLN, S. E., and E. S. LANDER, 1987 Constructing genetic linkage maps with MAPMAKER: a tutorial and reference manual. Technical Report, Whitehead Institute, Cambridge, Mass.

LOEB, D. D., R. W. PADGETT, S. C. HARDIES, W. R. SHEHEE, M. B. COMER, M. H. EDGELL and C. A. HUTCHISON III, 1986 The sequence of a large L1Md element reveals a tandemly repeated

5′ end and several features found in retrotransposons. Mol. Cell. Biol. 6: 168–182.

LOVE, J. M., A. M. KNIGHT, M. A. MCALEER and J. A. TODD, 1990 Towards construction of a high resolution map of the mouse genome using PCR analysed microsatellites. Nucleic Acids Res. 18: 4123–4130.

LURQUIN, C., A. VAN PEL, B. MARIAME, E. DE PLAEN, J.-P. SZIKORA, C. JANSSENS, M. J. REDDEHASE, J. LEJEUNE and T. BOON, 1989 Structure of the gene of Tum⁻ transplantation antigen P91A: the mutated exon encodes a peptide recognized with L^d by cytolytic T cells. Cell 58: 293–303.

MANLEY, K. F., and R. W. ELLIOTT, 1991 R1 Manager, a computer program for analysis of data from recombinant inbred strains. Mamm. Genome 1: 123–126.

MORRY, M., and J. D. HARDING, 1986 Modulation of transcriptional activity and stable complex formation by 5′-flanking regions of mouse $tRNA^{His}$ genes. Mol. Cell. Biol. 6: 105–115.

OYLER, G. A., J. W. POLLI, M. C. WILSON and M. L. BILLINGSLEY, 1991 Developmental expression of the 25-kDa synaptosomal-associated protein (SNAP-25) in rat brain. Proc. Natl. Acad. Sci. USA 88: 5247–5251.

ROBERT, B., P. BARTON, A. MINTY, P. DAUBAS, A. WEYDERT, F. BONHOMME, J. CATALAN, D. CHAZOTTES, J. L. GUENET and M. BUCKINGHAM, 1985 Investigation of genetic linkage between myosin and actin genes using an interspecific mouse back-cross. Nature 314: 181–183.

RYCHLIK, W., and R. E. RHOADS, 1989 A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing, and *in vitro* amplification of DNA. Nucleic Acids Res. 17: 8543–8551.

SAMBROOK, J., E. F. FRITSCH and T. MANIATIS, 1989 *Molecular Cloning: A Laboratory Manual*, Ed. 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

SCHAEFER, R., E. BOLTZ, A. BECKER, F. BARTELS and J. T. EPPLEN, 1986 The expression of the evolutionarily conserved GATA/GACA repeats in mouse tissues. Chromosoma 93: 496–501.

SIRACUSA, L. D., C. M. SILAN, M. J. JUSTICE, J. A. MERCER, A. R. BAUSKIN, Y. BEN-NARIAH, D. DUBOULE, N. D. HASTIE, N. G. COPELAND and N. A. JENKINS, 1990 A molecular genetic linkage map of mouse chromosome 2. Genomics 6: 491–504.

STAHL, J., D. P. GEARING, T. A. WILLSON, M. A. BROWN, J. A. KING and N. M. GOUGH, 1990 Structural organization of the genes for murine and human leukemia inhibitory factor. J. Biol. Chem. 265: 8833–8841.

STALLINGS, R. L., A. F. FORD, D. NELSON, D. C. TORNEY, C. E. HILDEBRAND and R. K. MOYZIS, 1991 Evolution and distribution of $(GT)_n$ repetitive sequences in mammalian genomes. Genomics 10: 807–815.

STURTEVANT, A. H., 1913 The linear arrangement of six sex-linked factors in Drosophila, as shown by their mode of association. J. Exp. Zool. 14: 43–59.

TAYLOR, B. A., 1989 Recombinant inbred strains, pp. 773–796 in *Genetic Variants and Strains of the Laboratory Mouse*, Ed. 2, edited by M. F. LYON and A. SEARLE. Oxford University Press. New York.

TAYLOR, B. A., H. J. HEINIGER and H. MEIER, 1973 Genetic analysis of resistance to cadmium-induced testicular damage in mice. Proc. Soc. Exp. Biol. Med. 143: 629–633.

TODD, J. A., T. J. AITMAN, R. J. CORNALL, S. GHOSH, J. R. S. HALL, C. M. HEARNE, A. M. KNIGHT, J. M. LOVE, M. A. MCALEER, J. PRINS, N. RODRIGUES, M. LATHROP, A. PRESSEY, N. H. DELARATO, L. B. PETERSON AND L. S. WICKER, 1991 Genetic analysis of autoimmune type 1 diabetes mellitus in mice. Nature 351: 542–547.

WEBER, J. L., 1990 Informativeness of human $(dC\text{-}dA)_n$ $(dG\text{-}dT)_n$ polymorphisms. Genomics 7: 524–530.

WEBER, J. L., and P. E. MAY, 1989 Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am. J. Hum. Genet. 44: 388–396.

Communicating editor: D. BOTSTEIN