Received 3 March; accepted 5 October 1992.

1. Bogue, S. W. & Merrill, R. T. *A. Rev. Earth planet. Sci.* **20,** 181-219 (1992).
2. Hoffman, K. A. *J. geophys. Res.* **89,** 6285-6292 (1984).
3. Tric, E. *et al. Phys. Earth planet. Int.* **65,** 319-336 (1991).
4. Clement, B. M. *Earth planet. Sci. Lett.* **104,** 48-58 (1991).
5. Laj, C., Mazaud, A., Weeks, R., Fuller, M. & Herrero-Bervera, E. *Nature* **351,** 447 (1991).
6. Valet, J.-P., Tucholka, P., Courtillot, V. & Meynadier, L. *Nature* **356,** 400-407 (1992).
7. Runcorn, S. K. *Nature* **356,** 654-656 (1992).
8. Constable, C. *Nature* **358,** 230-233 (1992).
9. Hoffman, K. A. *Nature* **354,** 273-277 (1991).
10. Coe, R. S., Hse, V. & Theyer, F. *Eos* **66,** 872 (1985).
11. Chauvin, A., Roperch, P. & Duncan, R. A. *J. geophys. Res.* **95,** 2727-2752 (1990).
12. Roperch, P. & Duncan, R. A. *J. geophys. Res.* **95,** 2713-2726 (1990).
13. Langereis, C. G., van Hoof, A. A. M. & Rochette, P. P. *Nature* **358,** 226-230 (1992).
14. Duncan, R. A. & McDougall, I. *J. Volc. geotherm. Res.* **1,** 197-227 (1976).
15. Hoffman, K. A. *Nature* **320,** 228-232 (1986).
16. Davies, G. F. in *The Encyclopedia of Solid Earth Geophysics*, 806-819 (Van Nostrand Reinhold, New York, 1989).
17. Tric, E. *et al. Earth planet. Sci. Lett.* **102,** 1-13 (1991).
18. Bogue, S. W. & Coe, R. S. *Nature* **295,** 399-401 (1982).
19. Prévot, M., Mankinen, E. A., Gromme, C. S. & Coe, R. S. *Nature* **316,** 230-234 (1985).
20. Clement, B. M. & Kent, D. V. *Earth planet. Sci. Lett.* **81,** 253-264 (1987).
21. Laj, C., Guitton, S., Kissel, C. & Mazaud, A. *J. geophys. Res.* **93,** 11655-11666 (1988).
22. Valet, J. P., Tauxe, L. & Clement, B. M. *Earth planet. Sci. Lett.* **94,** 371-384 (1989).
23. Coe, R. S. & Prevot, M. *Earth planet. Sci. Lett.* **92,** 292-298 (1989).
24. Watkins, N. D. *Geophys. J. R. Astr. Soc.* **17,** 121-149 (1969).
25. Hoffman, K. A. *Nature* **294,** 67-68 (1981).
26. Clement, B. M. *Eos* **70,** 1073 (1989).
27. Gubbins, D. & Bloxham, J. *Nature* **325,** 509-511 (1987).
28. Bloxham, J. & Jackson, A. *Rev. Geophys.* **29,** 97-120 (1991).
29. Bloxham, J. & Gubbins, D., *Scient. Am.* **261,** 68-75 (1989).
30. Gubbins, D. *Nature* **326,** 167-169 (1987).
31. Barton, C. E. in *The Encyclopedia of Solid Earth Geophysics*, 560-577 (Van Nostrand Reinhold, New York, 1989).
32. Dziewonski, A. M. & Woodhouse, J. H. *Science* **236,** 37-48 (1987).
33. Clement, B. M. & Martinson, D. G. *J. geophys. Res.* **97,** 1735-1752 (1992).
34. Valet, J.-P. & Laj, C. *Nature* **311,** 552-555 (1984).
35. Shaw, J. *Geophys. J. R. Astr. Soc.* **40,** 345-350 (1975).

ACKNOWLEDGEMENTS. I thank D. Vandamme for field assistance, R. Howell for field and laboratory assistance, M. Cardella for assisting with computer simulations, S. Lund for plots of the geomagnetic field and R. Merrill for discussions. This project was supported by the NSF.

A second-generation linkage map of the human genome

Jean Weissenbach*†, Gabor Gyapay*‡, Colette Dib*, Alain Vignal*, Jean Morissette*§, Philippe Millasseau*‡, Guy Vaysseix* & Mark Lathrop‡‖

* Genethon, 1 rue de l'Internationale, 91000 Evry, France
† Unité de Génétique Moléculaire Humaine, CNRS URA 1445, Institut Pasteur, 75724 Paris cédex, France
‡ Centre d'Etudes du Polymorphisme Humain (CEPH) and § Réseau de Médecine Génétique, Centre Hospitalier de l'Université Laval, Québec, Canada
‖ INSERM U 358, 27 Rue Juliette Dodu, 75010 Paris, France

A linkage map of the human genome has been constructed based on the segregation analysis of 814 newly characterized polymorphic loci containing short tracts of (C-A)$_n$ repeats in a panel of DNAs from eight large families. Statistical linkage analysis placed 813 of the markers into 23 linkage groups corresponding to the 22 autosomes and the X chromosome; 605 show a heterozygosity above 0.7 and 553 could be ordered with odds ratios above 1,000:1. The distance spanned corresponds to ~90% of the estimated length of the human genome.

ABOUT 3,000 human polymorphic markers have been isolated[1], but 90% of them have a heterozygosity of less than 50% and they are unevenly spaced throughout the genome. Although maps based on these markers[2-4] have contributed greatly to the primary mapping of a number of diseases, they are still insufficient for many applications such as mapping rare monofactorial diseases, refining linkage intervals to distances suited for gene identification, and mapping of loci contributing to complex traits[5,6]. Therefore there is a need for more informative markers and a higher density linkage map.

Highly informative markers have been difficult to obtain. Although efficient strategies for the isolation of hypervariable minisatellite markers are available[7,8], these loci tend to cluster in the tips of chromosome arms. Recently however, the source of polymorphic markers has increased with the observation that short interspersed tandem repeats or microsatellites are frequently highly polymorphic, and can be analysed using polymerase chain reaction (PCR) techniques[9,10]. Microsatellites appear widely distributed and can be readily isolated from genomic libraries.

The development of microsatellite markers on a large scale and the application of a multiplex genotyping procedure[11-13], allowed us to construct a genome map consisting mostly of markers with an elevated heterozygosity (>0.7) and covering a linkage distance spanning 90% of the genome as estimated from other maps[4].

Development of microsatellite markers

DNA libraries were made from *Alu*I DNA digests of 46,XX human DNA sized between 300 and 500 base pairs (bp) and cloned in M13. With the exception of $A_n{\cdot}T_n$ multimers, $(C\text{-}A)_n{\cdot}(G\text{-}T)_n$ repeats are the most frequent simple sequence repeats found in the human genome[14]. Recombinant plaques were therefore screened with a poly(dC-dA)·poly(dG-dT) probe. DNA templates from $(C\text{-}A)_n$-positive clones, ~1% of the M13 plaques were then sequenced and amplification primers on either side of the poly A-C tract were chosen using a computer program. From a total of 12,014 sequenced inserts, 2,995 were selected for amplification assays; 9,028 clones were discarded for assorted reasons (Table 1) including clones for which the C-A repeat was too short (<12 units long) or absent, or beyond the reliable part of the determined sequence. In some other sequences no oligonucleotide pairs fulfilling our requirements for primer selection could be found.

Primer pairs were synthesized for the 2,995 selected microsatellite markers and assayed by PCR amplification of four unrelated 46,XX individuals from the CEPH families. This assay was also used as a first rough estimate for the polymorphism of the tested microsatellites and to favour selection of highly informative markers. To aid large scale genotyping, a single set of PCR conditions was chosen[13]. Among 2,995 markers tested on the four individuals, 84% (2,506) were successfully amplified, 93% (2,327) of the amplified markers were polymorphic, 72% (1,809) showed three or more alleles and in 36% the highest allele frequency was below 0.5. About 4% of primer pairs promoted amplification of more than one allelic system. Only those showing unambiguously resolved systems were used further.

Markers with three or more alleles were first assigned to their chromosome using DNAs from a panel of cell lines of 18 human rodent hybrids each containing a subset of human chromosomes.

TABLE 1 Overall recovery of $(C\text{-}A)_n$ containing M13 templates

Type of sequence	Number of sequences	Per cent of total
Sequences used for PCR	2,995	25
Primers not found	624	5
8 to 12 (A-C) repeats	1,398	12
Microsatellite absent of <8 (A-C) repeats	956	8
Unreliable sequences	2,750	23
Misplaced microsatellites	2,910	24
Duplicates	381	3
Total	12,014	100

Sequencing was done using the dideoxynucleotide chain termination method with fluorescent primers[19] on Applied Biosystems 373A DNA sequencers. Sequencing reactions were done using the *Taq* polymerase linear amplification procedure (cyclic sequencing)[20] according to a protocol established by ABI. A single sequencing run was done for each template, and unreliable sequences were discarded. Oligonucleotide primers flanking the microsatellite tandem repeats were chosen with a computer program derived from the program OLIGO[21] and using additional criteria like base composition, avoidance of homologies with repeated sequences (Alu, Kpn and so on). Their T_m ranged from 55 °C to 70 °C. Special care was taken to avoid self priming, particularly at the 3′ ends. The primers were purchased from a company (Genset, Paris) and were used without purification for PCR reactions.

We found that 1,547 markers (86% of 1,806 tested) could be assigned to a chromosome (Table 2). The overall microsatellite density per chromosome fluctuates between 0.57 and 1.53 times the mean value of 0.49 marker per megabase (Mb). Some chromosomes such as 21, 9 and 19 contain significantly less $(C\text{-}A)_n$ than average whereas others such as 17 or 18 contain much more.

A better estimation of observed heterozygosity values of these 814 markers was based on scoring 28 unrelated CEPH individuals. Heterozygosities range from 0.32 to 0.96. 790 markers (97%) show a heterozygosity above 0.5 and 605 (74.4%) above 0.7 (Fig. 1).

Genotyping and data verification

The potentially most informative markers were genotyped on a subset of the CEPH panel composed of the eight largest families. This genotyping was extended to less informative markers for chromosomes with lower microsatellite densities. The eight families contained a total of 91 children, 16 parents and 26 grandparents. Genotypes were characterized following a multiplexing procedure that we have developed recently[13]. Results were obtained as images on films after photographic exposures of hybridized membranes[13]. Allelic systems with unclear or poorly resolved amplification patterns were discarded. The genotypes were interpreted in two independent readings of the images, and entered twice into a computerized database. The two entries were compared, and differences were either resolved by a third interpretation, or the data were judged to be ambiguous and removed. Further verifications of the data were made in families that were identified as containing possible genotype errors based on the pattern of pairwise recombination events between linked markers.

Several segregation anomalies were encountered. In some families absence of amplification of one allele could be deduced from the segregation analysis. Such an allele was considered as cryptic and these genotypes were discarded. Mutations have been observed for 160 markers. In most cases they affected a single allele from one individual (120 markers) or one allele in two generally unrelated individuals (27 markers). Thirteen markers showed between three and seven mutation events. But because these latter events were more frequently found in a single family, a number of those may be ascribed to germ-line mosaicism and not to independent events. Altogether these different events correspond to a mutation rate close to 0.1%. Genotypes of individuals with obvious mutations were eliminated. Note that due to the rather important mutation rate of microsatellites one allele may sometimes be converted into another allele existing in the family, resulting in an erroneous interpretation of the origin of an allele.

Linkage and genetic maps

Genotypes in the eight families were analysed with the LINKAGE programs[15]. All 814 markers showed significant linkage (lod score >3) to at least three other markers in the CEPH database (version 5). Generally, the linkage results and the physical assignments based on somatic cell hybrids were in agreement. But, in a few instances (3%) the physical assignments were found to be erroneous. All markers that were not physically assigned (~15%) could be assigned through linkage, except for a probably pseudoautosomal marker. Pairwise lod scores were calculated between markers assigned to the same chromosome, and the recombination estimates were used to choose a subset of 5–18 loci with spacing of ~10 to 20 centimorgans (cM) as a framework map for each chromosome. In some instances some larger intervals had to be taken. The order of the framework markers was confirmed with odds >1,000:1 against possible alternatives, and the framework maps were used as a starting point to order other markers.

As expected, an overall larger genetic length was observed in female meioses (data not shown). The complete maps for all chromosomes are represented in Fig. 2 with the sex-average recombination distances between adjacent markers. Rough localizations for selected reference markers from the CEPH database[4,16] are shown to allow comparison to other genetic maps. The markers on each chromosome from a single linkage group with the exception of chromosome 12, for which the most

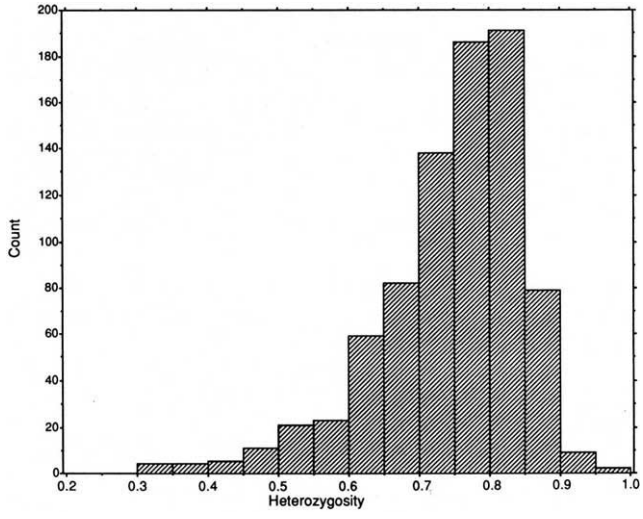
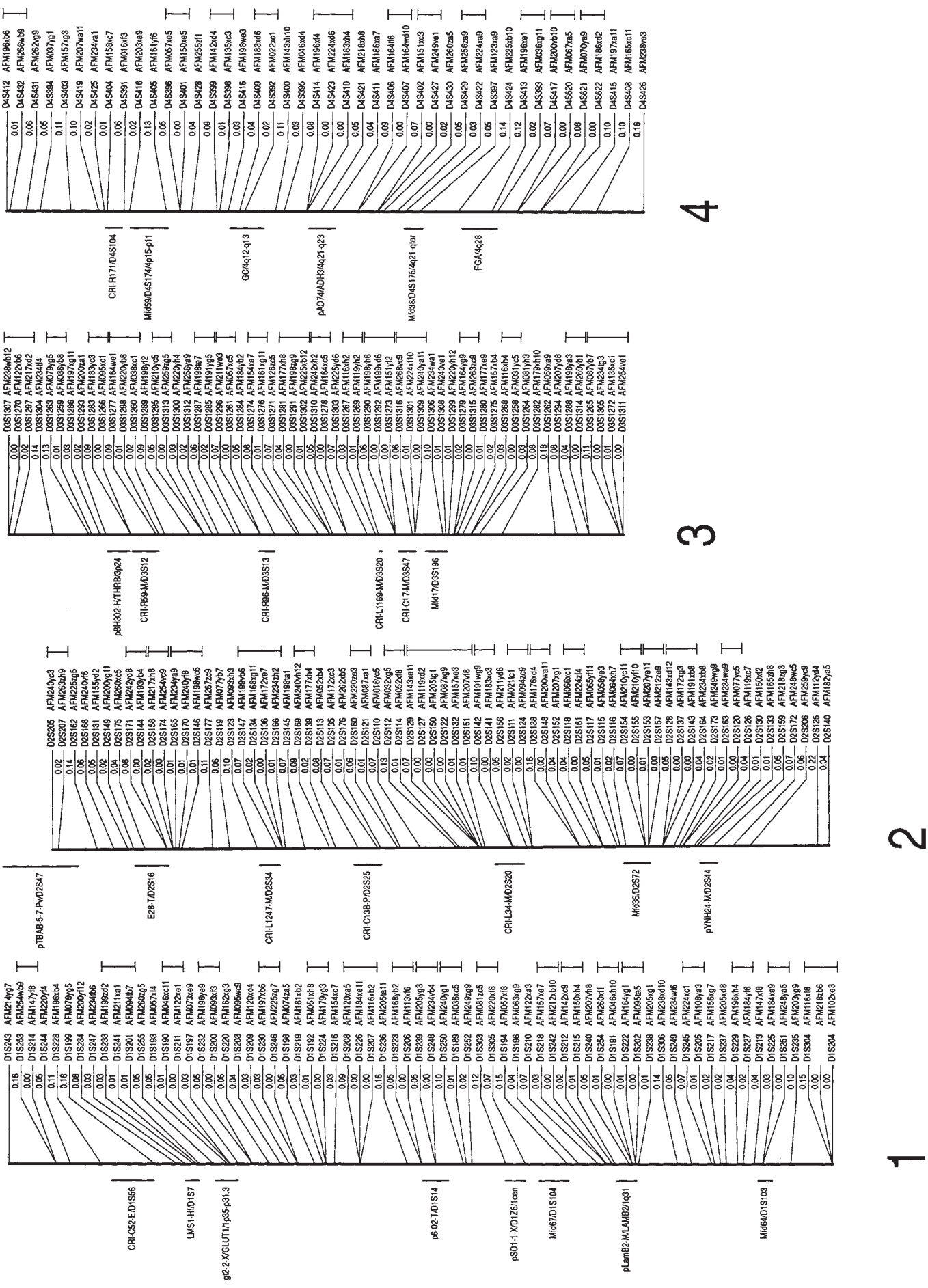
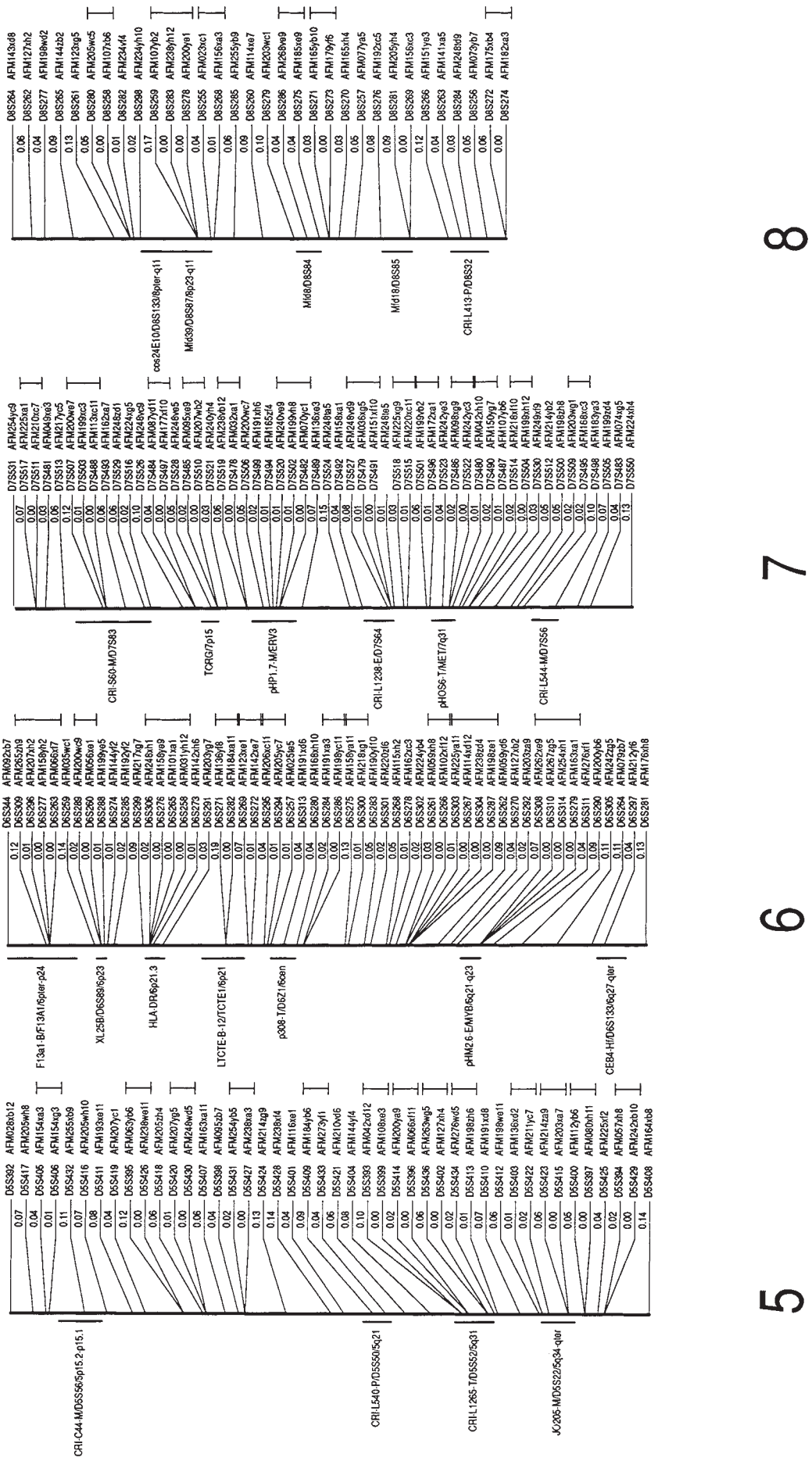
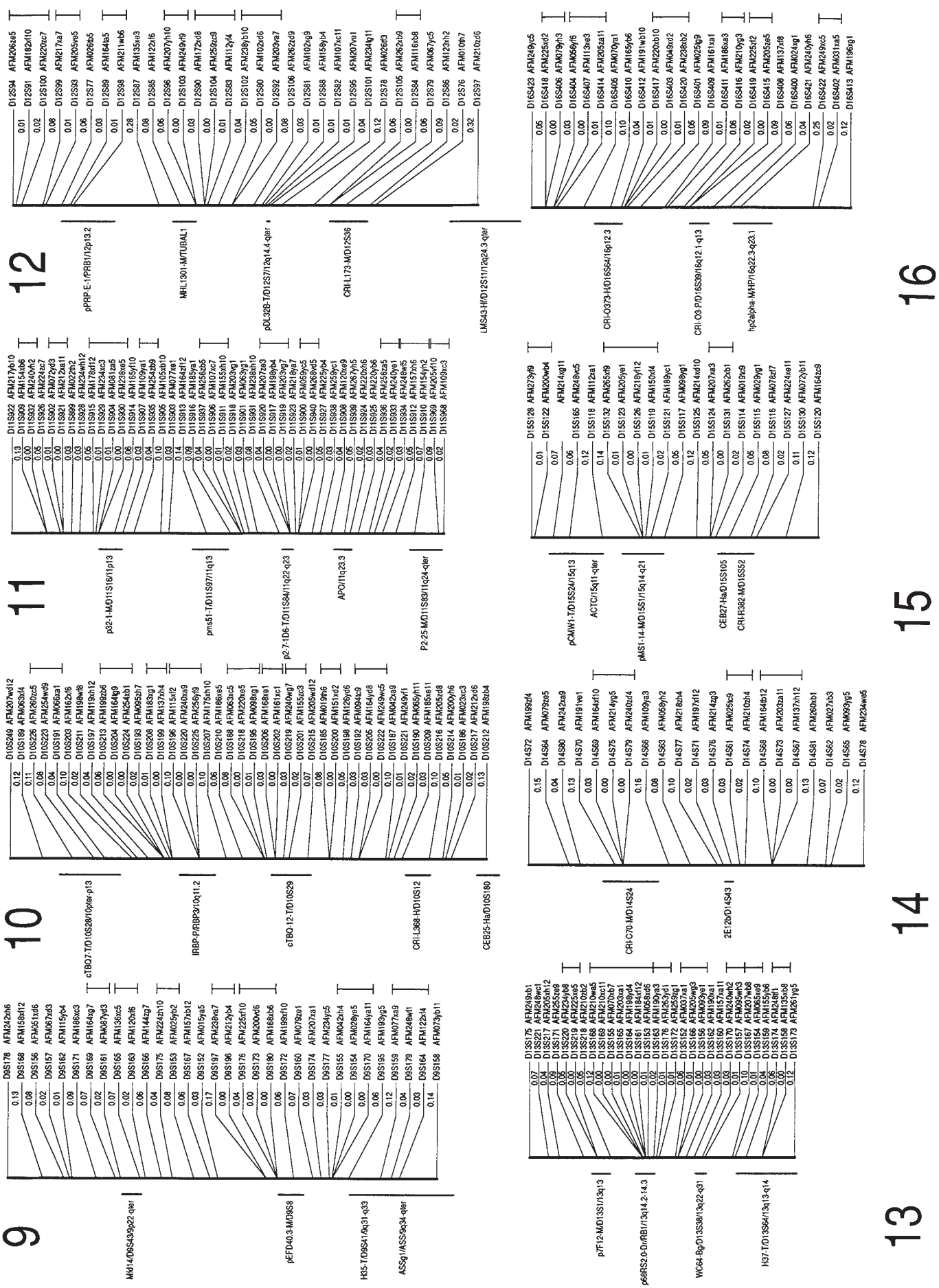


FIG. 1 Heterozygosity of 814 $(C\text{-}A)_n$ microsatellites mapped in this study. Heterozygosity was estimated from a total of 28 unrelated individuals of the eight largest CEPH families (CEPH pedigree numbers 102, 884, 1331, 1332, 1347, 1362, 1413 and 1416). DNA samples were amplified by PCR independently for each marker and were amplified as described[13]. The reactions were performed using a 'hot-start' procedure: *Taq* polymerase was added only after a first denaturation step of 5 min at 96 °C. Eight to sixteen different PCR products amplified on DNA samples from the same individual, were coprecipitated and comigrated in a single gel lane. About 80 coprecipitates (four families) were processed on a single gel. Coprecipitated PCR products of the M13 templates cognate with the typed markers were used as sizing standards along with ladders of poly(A). Products separated on denaturing acrylamlide gels were transferred onto Hybond N^+ nylon membranes. Membranes were successively probed with nonradioactively labelled PCR primers, as reported[13]. Primer sequences and allele frequencies of these markers have been deposited in the GDB.







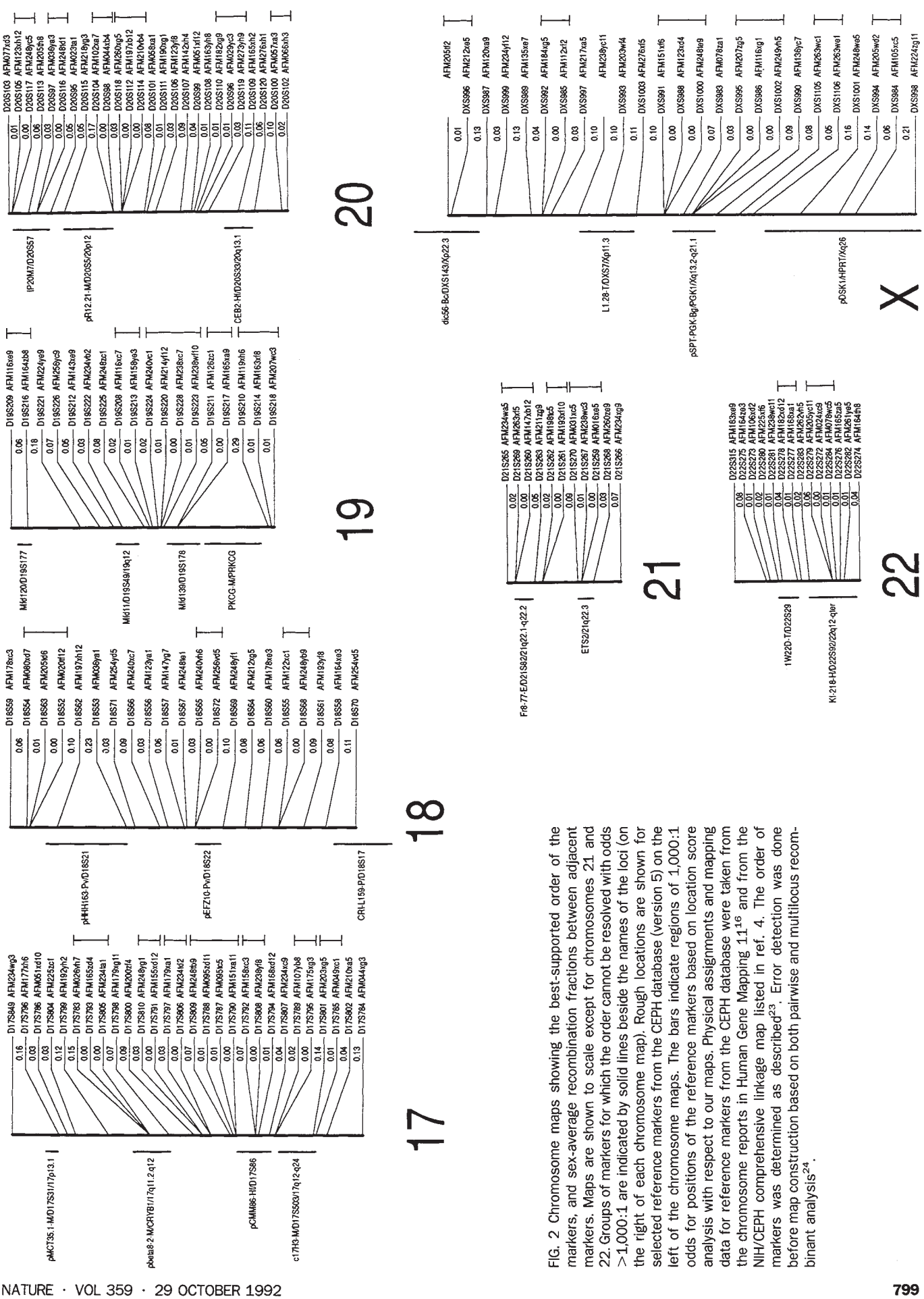


FIG. 2 Chromosome maps showing the best-supported order of the markers, and sex-average recombination fractions between adjacent markers. Maps are shown to scale except for chromosomes 21 and 22. Groups of markers for which the order cannot be resolved with odds >1,000:1 are indicated by solid lines beside the names of the loci (on the right of each chromosome map). Rough locations are shown for selected reference markers from the CEPH database (version 5) on the left of the chromosome maps. The bars indicate regions of 1,000:1 odds for positions of the reference markers based on location score analysis with respect to our maps. Physical assignments and mapping data for reference markers from the CEPH database were taken from the chromosome reports in Human Gene Mapping 11[16] and from the NIH/CEPH comprehensive linkage map listed in ref. 4. The order of markers was determined as described[23]. Error detection was done before map construction based on both pairwise and multilocus recombinant analysis[24].

TABLE 2 Summary of quantitative characteristics of the chromosome markers and maps and comparisons to reference data

Chromosome	Physical length (Mb)	Number of markers assigned	Markers per Mb assigned	Estimated genetic length* (cM)	NIH/CEPH genetic length† (cM)	Length covered (cM)	NIH/CEPH extra length pter (qcen)‡ (cM)	NIH/CEPH extra length qter (cM)	Number of markers mapped	Number of markers positioned with odds >1,000:1	Mean interval (cM)§
1	263	122	0.46	305	390	295	19	22	69	48	4.28
2	255	126	0.49	271	354	277	2	0	70	47	3.96
3	214	105	0.49	237	334	221	10	10	59	35	3.75
4	203	85	0.42	244	275	229	3	10	44	32	5.20
5	194	97	0.50	234	297	201	3	4	44	31	4.57
6	183	104	0.57	207	249	201	10	10	55	36	3.65
7	171	80	0.47	178	267	195	7	10	54	38	3.61
8	155	86	0.55	172	237	155	0	25	32	25	4.84
9	145	45	0.31	146	170	160	2	6	32	22	5.00
10	144	83	0.58	181	250	178	5	3	42	28	4.24
11	144	72	0.50	150	245	161	−4	20	44	29	3.66
12‖	143	71	0.50	160	201	172	−5	0	31	21	5.55
13q	98	46	0.47	130	219	99	25	45	30	17	3.30
14q	93	54	0.58	122	171	125	4	38	21	16	5.95
15q	89	46	0.52	154	178	107	7	2	20	15	5.35
16	98	42	0.43	157	162	119	10	8	24	16	4.96
17	92	69	0.75	208	168	128	2	12	28	18	4.57
18	85	53	0.62	143	109	126	−30	0	21	17	6.00
19	67	23	0.34	148	114	95	15	10	18	10	5.28
20	72	40	0.56	122	147	101	0	37	26	18	3.88
21q	39	11	0.28	114	67	29	15	10	11	6	2.64
22q	43	20	0.47	81	98	32	37	25	14	11	2.29
X	164	67	0.41	220	208	170	≈0	10	25	17	6.80
Total	3,154	1,547	0.49	4,084	4,910	3,576	176	317	814	553	4.39

The extra length of the NIH/CEPH maps has been evaluated using, as a reference point for each telomeric region, a marker from the CEPH datase very closely linked to the most distal marker from our maps. To aid linkage analysis, markers were first physically assigned to chromosomes when possible. DNA from cell lines from a panel of 18 human rodent hybrids (NIGMS Human Genetic Mutant Cell Repository panel 1, Coriell Institute, Camden), each containing a subset of human chromosomes, were tested by PCR for all tri- or multiallelic markers and scored for the presence or absence of an amplification product. PCR amplifications were done as described[13]. PCR products were visualized on agarose gels.

* Values from ref. 22.

† Values from ref. 4.

‡ For the acrocentric chromosomes (13, 14, 15, 21, 22), this extra length applies to qcen and not pter.

§ Mean interval is the ratio of length covered per number of markers mapped.

‖ Position of the distal marker framing the largest gap of chromosome 12 (see Fig. 2) is not supported by odds of 1,000:1 and has not been included in the distance covered by this linkage group.

distal marker (AFM210zd6) is not significantly linked to the other markers. But AFM210zd6 has lod scores >12 with two other markers from the CEPH database that had previously been mapped to chromosome 12q. A subset of 553 of the 814 markers formed an anchor map in which the order of all markers were resolved with odds >1,000:1, demonstrating the power to obtain precise linkage maps from highly informative markers in the eight families we selected for genotype characterization.

Table 2 summarizes several other features of the linkage maps. The average distance between adjacent markers is ~5%, and the total distance spanned by the markers was estimated to be 36 Morgans (Kosambi's mapping function). A total of seven internal intervals larger than 20 cM between adjacent loci can be seen on Fig. 2. In addition several subtelomeric regions appear to have a low density of markers (for example 1q, 2q, 12q, 20q). Conversely, clusters of microsatellites are observed in several regions. In particular, 17 of 30 markers on chromosome 13 are localized within a 24 cM interval. The maps for several of the smaller chromosomes, such as 21 and 22, cover less than 40% of their estimated length, which also suggests clustering of microsatellite markers on these chromosomes.

Discussion

We isolated and mapped a large number of $(C\text{-}A)_n$ microsatellites to construct a linkage map of the human genome with an average resolution of 5 cM. The quality of such a map relies on the informativeness of the markers and their distribution.

The mean level of heterozygosity of our markers is close to 0.75 indicating that the selection procedure for highly polymorphic markers has been successful. This procedure was based on (1) selection of markers containing a minimum of 12 C-A doublets in the sequenced allele, (2) test of polymorphism on a sample of four individuals (eight alleles), (3) choice of markers for which the highest allele frequency was <0.5 among these eight alleles. Use of these last criteria has been evaluated recently[17].

Although the markers obtained are widely dispersed, two sets of arguments indicate that their distribution is not random. First, the distribution observed by chromosome assignment of 1,547 markers selected only on the basis of potential informativeness has an almost threefold variation between different chromosomes adjusted to their physical size. Because the source library of the markers was constructed using size fractionated *Alu*I fragments, it could be argued that it is actually the sequence of *Alu*I restriction sites (A-G-C-T) which is non-uniformly distributed among chromosomes. But in both cases this indicates that some sequence motifs do not have the same frequency on all chromosomes. Tentative correlations of this non-uniform distribution with the ratios of G to R bands of each chromosome were not conclusive. Otherwise a bias in distribution could result from the various selection steps such as exclusion of less informative or poorly amplified sequences or sequences for which no primers could be selected. It should be noted that the first set of markers (~760) was selected on the basis of high informativeness. Some less informative markers mapping to poorly covered chromosomes (for example 13, 21, 22 and so on) were included afterwards. Several other markers were added to gaps

larger than 20 cM or to extend the lengths covered for some chromosomes.

Second, the non-random distribution within a single chromosome is sometimes even more dramatic, as best exemplified by chromosomes 13 and 22. Because the cytogenetic location of our markers is unknown, correlations between physical and linkage distances is speculative. But an apparent depletion of microsatellites in terminal parts of chromosomes was found for chromosome 20 (ref. 12) and might also be related to the map inflations observed in terminal regions.

The NIH/CEPH chromosome maps include 1,416 markers and most frequently extend slightly beyond our maps. Because these map extensions amount approximately to 5 Morgans (Table 2), our maps cover about 90% of the NIH/CEPH maps. A discrepancy of 8 Morgans still remains between the two sets of data. This discrepancy varies from chromosome to chromosome as can be seen from Table 2 but might be in part attributable to the heterogeneous set of data used to construct the NIH/CEPH map and to the fact that the present maps are based on a subset of eight families from the CEPH panel.

Despite several gaps above 20 cM, the present maps provide a general tool which should allow primary linkage mapping of most single gene disorders. This should be possible even when only a limited number of affected families is available because of the high informativeness of the markers. This collection of markers should also serve as a resource to help fill the remaining gaps in the present effort to construct index marker maps with heterozygosities above 0.7 and no gaps greater than 15 cM. Note that many of the gaps correspond to subtelomeric regions in which most of the known variable number of tandem repeats loci reside. Thus the two alternative sources of micro- and minisatellites appear to be complementary, and the existing markers from these sources together may constitute at present an almost continuous set of highly informative markers covering the human genome. A mouse genetic map of similar resolution has been reported recently[18]. □

Received 25 September; accepted 8 October 1992.

1. Williamson, R. *et al. Cytogenet. Cell Genet.* **58,** 1190-1832 (1991).
2. Donis-Keller, H. *et al. Cell* **51,** 319-337 (1987).
3. White, R. *et al.* in *Genetics Maps. Locus Maps of Complex Genomes* 5th edn (O'Brian S. J.) 5.134-5.157 (Cold Spring Harbor Laboratory Press, New York, 1990).
4. NIH/CEPH Collaborative mapping group *Science* **258,** 67-86 (1992).
5. Bishop, D. T. & Williamson, J. A. *Am. J. hum. Genet.* **46,** 254-265 (1990).
6. Risch, N. *Am. J. hum. Genet.* **46,** 242-253 (1990).
7. Nakamura, Y. *et al. Science* **235,** 1616-1622 (1987).
8. Vergnaud, G., Maria, D., Zoroastro, M. & Lauthier, V. *Electrophoresis* **12,** 134-140 (1991).
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. Church, G. M. & Kieffer-Higgins, S. *Science* **240,** 185-188 (1988).
12. Hazan, J. *et al. Genomics* **12,** 183-189 (1992).
13. Viginal, A. *et al.* in *Methods in Molecular Genetics* Vol. 1 (ed. Adolph, K. W.) (Academic, Orlando, in the press).
14. Beckmann, J. S. & Weber, J. L. *Genomics* **12,** 627-631 (1992).
15. Lathrop, G. M. & Lalouel, J. M. *Am. J. hum. Genet.* **36,** 460-465 (1985).
16. *Cytogenet. Cell Genet.* **58,** 1-2200 (1991).
17. Ott, J. *Am. J. hum. Genet.* **51,** 283-290 (1992).
18. Dietrich, W. *et al. Genetics* **131,** 423-447 (1992).
19. Smith, L. M. *et al. Nature* **321,** 674-679 (1986).
20. Saluz, H. & Jost, J. P. *Proc. natn. Acad. Sci. U.S.A.* **86,** 2602-2606 (1989).
21. Rychlik, W. & Rhoads, R. E. *Nucleic Acids Res.* **17,** 8543-8551 (1989).
22. Morton, N. E. *Proc. natn. Acad. Sci. U.S.A.* **88,** 7474-7476 (1991).
23. Lathrop, G. M. *et al. Genomics* **2,** 157-164 (1988).
24. Lathrop, G. M. & Lalouel, J. M. *Stat. Meth. biol. Med. Sci.* **8,** 81-123 (1991).

ACKNOWLEDGEMENTS. We thank L. Baron, N. Becuwe, M. Besnard, I. Bordelais, N. Cheron, C. Cruaud, D. Delporte, C. Discala, C. Dumont, C. Dupraz-Ramel, S. Duverneuil, E. Ernst, K. Fonsat, M. François, D. Gondouin, J. L. Mangua, C. Marquette, E. M. Bene, S. Nguyen, S. Pezard, M. Tranchant, N. Vega, N. Vuillaume, E. Wunderle and V. Wunderle for technical or clerical assistance and X. Benigni, L. Bougueleret, J.-M. Froussard, R. Gavrel, P. Gesnouin, S. Pook, P. Rodriguez-Tomé and C. Scarpelli for data processing. This work was supported by the Association Française contre les Myopathies.

Crystal structure of staphylococcal enterotoxin B, a superantigen

S. Swaminathan, William Furey, James Pletcher & Martin Sax

Biocrystallography Laboratory, VA Medical Center, PO Box 12055, University Drive C, Pittsburgh, Pennsylvania 15240, USA

The three-dimensional structure of staphylococcal enterotoxin B, which is both a toxin and a superantigen, has been determined to a resolution of 2.5 Å. The unusual main-chain fold containing two domains may represent a general motif adopted by all staphylococcal enterotoxins. The T-cell receptor binding site encompasses a shallow cavity formed by both domains. The MHCII molecule binds to an adjacent site. Another cavity with possible biological activity was also identified.

THE staphylococcal enterotoxins comprise a family of five serologically distinct toxins (labelled A-E), all of which are secreted by various strains of *Staphylococcus aureus*, and share significant sequence homology[1-3]. These toxins are primary causes of food poisoning, with all members inducing the same biological effects: emesis, diarrhoea and mitogenesis[1-5]. More importantly, they can induce profound immune system responses by stimulating T cells with particular Vβ elements[6-10]. The fraction of responding T cells can be orders of magnitude greater than that evoked by conventional antigens. Accordingly, staphylococcal enterotoxins have been termed 'superantigens'[11].

Superantigens have similarities to conventional antigens. For example, both types of exogenous antigens must be presented to a T-cell antigen receptor (TCR) by a class II molecule of the major histocompatability complex (MHCII)[12,13], and both types form ternary complexes (antigen-MHCII-TCR) that trigger cytokine production and T-cell proliferation. But superantigens and antigens differ in some main respects. The main differences lie in how they interact with MHCII and TCR molecules. Ordinary antigens are preprocessed into short peptide fragments, 13 to 17 residues in length[14], that bind to a cleft on the surface of the MHCII molecule for presentation to the TCR[15,16]. In contrast, with superantigens the intact protein molecule binds to MHCII[12]. There is also evidence that superantigens bind to a different place on the MHCII molecule than do processed peptides[17-19].

The difference in the mode of binding gives rise to different T-cell interactions. When a peptide-MHCII complex combines with a TCR, the antigenic fragment interacts with variable components (V) and others on both TCR α and β chains[20]. The