The 1993–94 Généthon human genetic linkage map

Gabor Gyapay[1,2], Jean Morissette[1,3], Alain Vignal[1], Colette Dib[1], Cécile Fizames[1], Philippe Millasseau[1,2], Sophie Marc[1], Giorgio Bernardi[4], Mark Lathrop[5] & Jean Weissenbach[1,6]

In 1992, we described a second-generation genetic linkage map of the human genome. Using 1,267 new microsatellite markers, we now present a new genetic linkage map containing a total of 2,066 $(AC)_n$ short tandem repeats, 60% of which show a heterozygosity of over 0.7. Statistical linkage analysis based on the genotyping of eight large CEPH families placed these markers in the 23 linkage groups. The map includes 1,266 intervals and spans a total distance of 3690 centiMorgans (cM). A total of 1,041 markers could be ordered with odds ratios greater than 1000:1. About 56% of this map is at a distance of 1 cM or less from one of its markers.

[1]*Généthon, 1 rue de l'Internationale, 91000 Evry, France*
[2]*Centre d'Etudes du Polymorphisme Humain, 27 rue Juliette Dodu, 75010 Paris, France*
[3]*Réseau de Médecine Génétique, Centre Hospitalier de l'Université Laval, Québec, Canada*
[4]*Laboratoire de Génétique Moléculaire, Institut Jacques Monod, 2 place Jussieu, 75005 Paris, France*
[5]*INSERM U 358, 27 rue Juliette Dodu, 75010 Paris, France*
[6]*Unité de Génétique Moléculaire Humaine, CNRS URA 1445, Institut Pasteur, 75724 Paris Cédex, France*

The use of microsatellite markers[1,2] has permitted the construction of new genetic maps of the human genome as well as those of other mammals[3–5]. These maps can be used to map any Mendelian trait and in particular, monogenic human diseases. On the other hand, the average length of the intervals between adjacent markers often requires a great deal of work to isolate new markers which are even closer to mapped genes. In order to facilitate this work for the entire human genome and to refine the genetic linkage intervals that can be covered rapidly by cloned DNA fragments for gene identification, we have developed and mapped 1,267 new $(AC)_n$ microsatellite markers. These new markers have been integrated into the map which we constructed previously and described at the end of 1992 (ref. 4).

New markers

New markers were obtained using the same procedures as for our first 814 markers[4]. However, as there were some indications that subtelomeric regions were less densely covered with markers[4], 62 of the markers on this new map (markers AFMa120 to AFMa152) are from the H3 and H2 isochores which are preferentially located in terminal parts of chromosomes[6]. Although the chromosomal distribution of the current complete collection of markers is very similar to that obtained previously[4], the distribution of markers from the H2 and H3 isochores is quite different (Wunderle *et al.*, manuscript in preparation).

The distribution of heterozygosity of the 2,066 markers is shown in Fig. 1. Compared to the previous map of 814 markers, which had a mean heterozyosity of 0.75, the average heterozygosity is 0.70, and the percentage of markers with a heterozygosity of over 0.70 has decreased from 74.4% to 60.1%. This reduction is due to the use of all the markers which showed at least three alleles when tested on four individuals[4], whereas in the 814-marker map we selected essentially those markers for which the frequency of the majority allele was less than 0.5.

Genotyping and map construction

Genotyping was carried out according to the multiplex procedure previously described[7]. Although we tried to

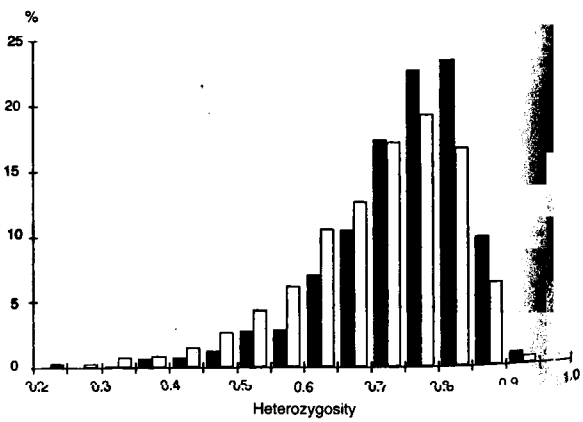

Fig. 1 Comparison of relative distribution of heterozygosity values of the $(CA)_n$ microsatellites in the 1992 and 1993–94 versions of the Généthon map. Dark columns: 814 markers; $H \geq 0.5$ for 97% and $H \geq 0.7$ for 74.4%; mean heterozygosity value = 0.75. Open columns: 2,066 markers; $H \geq 0.5$ for 93% and $H \geq 0.7$ for 60.1%; mean heterozygosity value = 0.70.

avoid genotyping microsatellites which were identical to sequences deposited in databases by others, occasionally the sequences of some of our markers were deposited in databases while our work was in progress. We have retained these markers on our maps because they are useful as common reference points among several maps (see notes, Fig. 2).

Genotyping errors were detected using the same diagnostic software programmes as described previously[4]. Genotypes of markers which appeared as double recombinants after several verifications were used for construction of the maps, although most of these double recombination events probably correspond to mutations or gene conversions.

A total of 305 mutations was observed in 278,338 genotypings. This percentage of 0.1% (0.05% of haploid genotypes) is very close to that observed previously by us[4] and by others[8]. About half of these mutations probably result from the use of DNA from continuous lymphoblastic cell lines[8].

The procedure used to construct the map was again similar to the one used for the previous version. However, our published map of 814 markers was used as a framework to position the new markers. Several markers which appeared on our previous map have been removed. In the majority of cases (AFM263zh9, AFM046xc11, AFM224xf10, AFM234tg3, AFM200ya9, AFM163xa1, AFM238yb10 and AFM182xg9), these markers had one or more alleles which could not be amplified with the primer pairs that we selected (cryptic alleles). These markers will be returned to the map when the primers have been modified so that all the alleles can be amplified. AFM136xe3, AFM165zd4, AFM217yc5 and AFM120xf6 each corresponded to two polymorphic loci and were therefore eliminated. During the initial sequence comparison, it was not noticed that marker AFM186xd2 was identical to AFM070ya9, so that the former was elimininated. Marker AFM262vg9 had a tendency to produce nonspecific bands and was also removed from the map. Finally, marker AFM123yf8 could not be repositioned unambiguously on the present map.

The new map

The new map covers a total distance of 3,690 cM, which represents an increase of 114 cM over the 814-marker map (Table 1). The extensions, which account for 135 cM, are due to the addition of more telomeric markers (Table 1). The total increase is less than these telomeric extensions because of modifications of distances in the interior regions of some chromosomes. These modifications usually involve small decreases in distances, but also increases. No new markers were added to the ends of eleven chromosomes: 3–10, 15, 18 and 19. However, chromosome 6 and especially chromosome 3 show significant increases in their total length, whereas chromosome 4 shows a marked decrease. Chromosomes 1 and X show decreases in total length despite the addition of one or more markers to at least one of their extremities, and 14 and 17 show increases which are clearly less than the extensions of their extremities. Sex-specific distances of each chromosome are indicated in Table 1.

There are a total of 1,266 intervals (Fig. 2) in the map which corresponds to an average distance of 2.9 cM between markers. We were able to position 1,041 markers

Table 1 Comparison of the main features of the 1992 (ref. 4) and 1993–1994 versions of the Généthon human genetic linkage map

Chromosome	Length covered (cM)				Largest gaps (cM)		Map extension since 1992 (cM)		Number of markers mapped		Number of markers positioned with odds > 1000:1	
	1992 sex-average	1993–94 sex-average	1993–94 male	1993–94 female	1992	1993–94	pter or qcen	qter	1992	1993–94	1992	1993–94
1	295	292	218	362	19	14		9	69	172	48	85
2	277	281	211	334	***24***	11	4		70	189	47	97
3	221	235	193	282	19	8			59	129	35	68
4	229	211	150	263	17	10			44	125	32	63
5	201	201	151	251	14	7			44	137	31	64
6	201	208	130	274	20	13			55	124	36	58
7	195	196	131	246	15	8			54	118	38	61
8	155	155	99	202	18	18			32	93	25	48
9	160	158	121	189	18	14			32	66	22	39
10	178	179	136	212	13	12			42	118	28	56
11	161	160	119	183	14	8	2		44	106	29	62
12	172	180	138	212	***30+37***	17	4		31	92	21	42
13q	99	123	102	146	12	19		23	30	66	17	35
14q	125	127	103	153	17	9	3	11	21	65	16	37
15q	107	107	80	135	14	12			20	53	15	27
16	119	130	101	164	***27***	13	8		24	52	16	29
17	128	133	108	155	15	13		11	28	72	18	34
18	126	129	99	161	***25***	13			21	62	17	31
19	95	99	86	119	***32***	***24***			18	42	10	20
20	101	120	83	142	18	10		20	26	57	18	31
21q	29	49	45	61	9	12	12	4	11	21	6	14
22q	32	49	40	67	8	8	18		14	27	11	13
X	170	168	–	168	***22***	18	4	2	25	80	17	27
Total	3576	3690	2644	4481			55	80	814	2066	553	1041

with an odds ratio of 1,000:1 or better. These maps, constructed with several tools based on the LINKAGE programme package, were compared to those using the same genotypes processed with the MultiMap algorithm[9], based on CRI-MAP[10]. Both sets of maps were essentially identical in order as well as distance. The main difference was observed with a few markers that were rejected by the GMS algorithm (Gene Mapping System)[11] because they could not be mapped to a region that was sufficiently precise whereas they were included in the comprehensive maps resulting from the MultiMap process. Refinement of the map and the increase in the number of markers has led to some modifications in the previous order of the markers. Four of these modifications concern markers previously ordered with odds ratios of greater than 1,000:1 and 14 concern markers positioned with lower probabilities.

Distribution of markers

Only one gap of over 20 cM remains on the map. The other gaps have been reduced in size. There are only 22 remaining gaps of over 10 cM, which represent 6 of the gaps which were over 20 cM on the previous map. A significant proportion of the markers from the H2 and H3 isochores (8 out of 62) were found to map to the distal end of the chromosomes and a number of the others are subtelomeric, as expected[12]. This indicates that markers from these GC-rich regions should permit a more dense coverage of numerous subtelomeric regions.

Correspondance between genetic and physical distance must await integration of genetic and physical maps. Genetic linkage maps integrating polymorphic markers from different sources including AFM markers from the first set of 814 have been established recently[9,13]. A more extended integration project using a different strategy and including new markers from the present map is in progress.

Conclusion

About 56% of our latest genetic linkage map of 3,690 cM is at a distance of 1 cM or less from one of its markers. In many cases, these distances can be covered by cloned DNA sequences[14]. Moreover, the isolation and mapping of 3,000 additional markers is in progress. This will increase the density of marker coverage and perhaps extend some of the chromosomal maps. This should accelerate considerably positional cloning of hereditary disease genes by facilitating the search for additional close genetic markers and candidate exons.

Methodology

Marker development. Marker development was carried out essentially as described[4]. DNA libraries were made from an *AluI* DNA digest of 46,XX human DNA (sized between 300–500 bp) and cloned in M13. The sequences of the templates from the $(CA)_n$ or $(GT)_n$ positive clones were used to define PCR primers. The synthesized primers were tested on four unrelated 46,XX individuals to obtain a first estimate of the polymorphism of the tested microsatellite markers. Markers with three or more alleles were first assigned to their chromosome and genotyped as described[4].

The H2 and H3 isochores were isolated by caesium sulphate density gradient centifugation of total human DNA in the presence of BAMD[6]. The $(AC)_n$ markers found in the fractions with the highest GC content were isolated by a single *AluI* digestion shotgun procedure as described[4].

Genotyping. Individuals from the eight CEPH families (102, 884, 1331, 1332, 1347, 1362, 1413 and 1416) were genotyped using standard procedures, as described[7]. PCR amplifications were performed in 50 µl reaction mixtures, containing 40 ng of genomic DNA, 50 pmol of each primer, 125 µM dNTPs, 10 mM Tris pH 9, 50 mM KCl, 1.5 mM $MgCl_2$, 0.1% Triton X-100, 0.01% gelatin and 1 U of *Taq* polymerase (Amersham). Amplifications were carried out using the "hot-start" procedure, in which the *Taq* polymerase is added to the reaction mixtures after a first denaturation step (5 min at 96 °C) after which 35 cycles of denaturation (94 °C for 40 s) and annealing (55 °C for 30 s) are performed. An elongation step (2 min at 72 °C) ends the process. For each DNA sample, 16 amplification products from different markers were ethanol-precipitated and loaded together into single lanes of 6% polyacrylamide-8M urea denaturing gels. After migration, the DNA was transferred from the gel to a Hybond N^+ nylon membrane (Amersham) by a contact blotting procedure[7]. The markers were then revealed by successive hybridizations with one of the PCR primers which was peroxidase-labelled by modification of the ECL procedure (Amersham) and exposed to autoradiographic X-ray films[7].

Map construction. Markers were assigned to chromosomes by pairwise linkage and possible genotyping errors were identified by comparisons between families of the pairwise recombination events between linked markers. After genotype corrections, markers from a chromosome-specific dataset were positioned on a framework consisting of the map of 814 markers[4] using a map construction algorithm. The order of markers in the framework and complete maps were determined with the GMS algorithm[11]. Briefly, recombination estimates for a preliminary, or trial order of the loci are used to divide the loci into subgroups of closely linked loci. Likelihoods are evaluated for different placements of subgroups, and for alternative orders of the loci within each subgroup. The best-supported order (i.e. the order with the greatest likelihood) is chosen as new trial order, and iterations are continued until convergence. Based on the best-supported order for the framework map, recombination fractions between adjacent markers were estimated with the LINKAGE programs[15]. Markers from this framework that underwent corrections since the 814-marker map were processed as new incoming markers. This led to a provisional order which was further reassessed as described[16]. Once the order remained unmodified after further computation, a search for double recombination events was undertaken. The maps were re-evaluated using the corrected genotypes until no further double recombination event could be eliminated.

Acknowledgements

We wish to acknowledge the essential technical and clerical contributions of Laurent Baron, Noëlle Becuve, Marielle Besnard-Gonnet, Isabelle Bordelais, Nathalie Cheron, Corinne Cruaud, Corinne Dumont, Evelyne Ernst, Karine Fonsat, Jacqueline Lotutala Mangua, Catherine Marquette, Elisabeta Mbimbi Bene, Delphine Muselet, Simon Nguyen, Sandra Pezard, Martine Tranchant, Nathalie Vega, Nathalie Vuillaume, Edith Wunderle and Véronique Wunderle. We would also like to thank the informatics team of Généthon, and particularly Lydie Bougueleret, Rémi Gavrel, Philippe Gesnouin, Stuart Pook, Patricia Rodriguez-Tomé, Claude Scarpelli and Guy Vaysseix. We are especially grateful to Susan Cure for her help in writing the manuscript. This work was initiated at CEPH and results from discussions with Daniel Cohen. It was supported by the Association Française contre les Myopathies, the Groupement d'Etudes et de Recherches sur les Génomes and European Union (Biomed1).

1. Litt, M. & Luty, J.A. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am. J. hum. Genet.* **44**, 397–401 (1989).
2. 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).
3. Dietrich, W. *et al.* A genetic map of the mouse suitable for typing intraspecific crosses. *Genetics* **131**, 423–447 (1992).
4. Weissenbach, J. *et al.* A second-generation linkage map of the human genome. *Nature* **359**, 794–801 (1992).
5. Serikawa, T. *et al.* Rat gene mapping using PCR-analyzed microsatellites. *Genetics* **131**, 701–721 (1992).
6. Bernardi, G. The isochore organization of the human genome and its evolutionary history — a review. *Gene* **135**, 57–66 (1993).
7. Vignal, A. *et al.* A non-radioactive multiplex procedure for genotyping of microsatellite markers. In *Methods in Molecular Genetics: Gene and Chromosome Analysis* (ed. Adolph, K.W.) 211–221 (Academic Press, San-Diego, 1993).
8. Weber, J.L. & Wong, C. Mutation of human short tandem repeats. *Hum. molec. Genet.* **2**, 1123–1128 (1993).
9. Matise, T.C., Perlin, M. & Chakravarti, A. Automated construction of genetic linkage maps using an expert system (MultiMap): a human genome linkage map. *Nature Genet.* **6**, 384–390 (1994).
10. Lander, E.S. & Green, P. Construction of multilocus genetic linkage maps in humans. *Proc. natn. Acad. Sci. U.S.A.* **84**, 2363–2367 (1987).
11. Lathrop, M. *et al.* A primary genetic map of markers of human chromosome 10. *Genomics* **2**, 157–164 (1988).
12. Saccone, S., De Sario, A., Della Valle, G. & Bernardi, G. The highest gene concentrations in the human genome are in T-bands of metaphase chromosomes. *Proc. natn. Acad. Sci. U.S.A.* **89**, 4913–4917 (1992).
13. Buetow, K.H. *et al.* Integrated human genome-wide maps constructed using the CEPH reference panel. *Nature Genet.* **6**, 391–393 (1994).
14. Cohen, D., Chumakov, I. & Weissenbach, J. A first-generation physical map of the human genome. *Nature* **366**, 698–701 (1993).
15. Lathrop, G.M. & Lalouel, J.M. Easy calculations of lod scores and genetic risks on small computers. *Am. J. hum. Genet.* **36**, 460–465 (1984).
16. Lathrop, G.M. & Lalouel, J.M. Statistical methods for linkage analysis. In *Handbook of Statistics* (ed. Chakraborty, C.) 81–123 (Elsevier, Amsterdam, 1991).
17. Hudson, T.J. *et al.* Isolation and chromosomal assignment of 100 highly informative human simple sequence repeat polymorphisms. *Genomics* **13**, 622–629 (1992).
18. Barber, T.D. *et al.* A highly informative dinucleotide repeat polymorphism at the D2S211 locus linked to ALPP, FN1 and TNP1. *Hum. molec. Genet.* **2**, 88 (1993).
19. Cassard, A.M. *et al.* Human uncoupling protein gene: structure, comparison with rat gene, and assignment to the long arm of chromosome 4. *J. cell. Biochem.* **43**, 255–264 (1990).
20. Goold, R.D. *et al.* The development of Sequence-Tagged sites for human chromosome 4. *Hum. molec. Genet.* **2**, 1271–1288 (1993).
21. Bosma, P.J. *et al.* Human plasminogen activator inhibitor-1 gene: Promoter and structural gene nucleotide sequences. *J. biol. Chem.* **263**, 9129–9141 (1988).
22. Wilkie, P.J., Krizman, D.B. & Weber, J.L. Linkage map of human chromosome 9 microsatellite polymorphisms. *Genomics* **12**, 607–609 (1992).
23. Litt, M., Sharma, V. & Luty, J.A. A highly polymorphic (TG)n microsatellite at the D11S35 locus. *Cytogenet. cell Genet.* **51**, 1034 (1989).
24. Reed, K.E. *et al.* Molecular cloning and functional expression of human connexin37, an endothelial cell gap junction protein. *J. clin Invest.* **91**, 997–1004 (1993).
25. Bowcock, A. *et al.* Microsatellite polymorphism linkage map of human chromosome 13q. *Genomics* **15**, 376–386 (1993).
26. Polymeropoulos, M.H., Xiao, H., Ide, S.E. & Merril, C.R. Dinucleotide repeat polymorphism at the D14S99E locus. *Hum. molec. Genet.* **2**, 490 (1993).
27. Shen, Y. *et al.* Four dinucleotide repeat polymorphisms on human chromosome 16 at D16S289, D16S318, D16S319 and D16S320. *Hum. molec. Genet.* **1**, 773 (1992).
28. Sharma, V., Guo, Z. & Litt, M. Dinucleotide repeat polymorphism at the D18S37 locus. *Hum. molec. Genet.* **1**, 289 (1992).
29. Brophy, B.K. *et al.* cDNA sequence of the pregnancy-specific β1-glycoprotein-11s (PSG-11s). *Biochim. Biophys. Acta* **1131**, 119–121 (1992).
30. Thompson, L.H. *et al.* Molecular cloning of thee human XRCC1 gene, which corrects defective DNA strand break repair and sister chromatid exchange. *Molec. Cell Biol.* **10**, 6160–6171 (1990).
31. Martin-Gallardo, A. *et al.* Automated DNA sequencing and analysis of 106 kilobases from human chromosome 19q13.3. *Nature Genet.* **1**, 34–39 (1992).
32. Lindsay, S.J. *et al.* Isolation and characterisation of highly polymorphic markers and retinal cDNAs in the vicinities of X-linked inherited eye disease loci. *Cytogenet. cell Genet.* **58**, 2075 (1991).

▶ Fig. 2 This map contains data for 2,066 $(AC)_n$ microsatellite markers, spanning the human genome. The chromosome maps showing the best-supported order of the markers and sex-average recombination fractions between adjacent markers are shown to scale. Most chromosomes are represented on several pages and data on the markers, such as the primer sequences, allele sizes and EMBL accession numbers, is given opposite the maps. 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 6) on the left of the chromosome maps. The bars indicate regions of 1000:1 odds for positions of the reference markers based on location score analysis with respect to our maps. All Généthon markers described here have been submitted to GDB.

The number of alleles, the heterozygosity and the maximum and minimum allele sizes (size range) have been determined by observation of 28 unrelated individuals, namely the grandparents or the parents of the CEPH families 1331, 1362, 102, 1347, 1332, 1416, 1413 and 884.

The allele sizes indicated in the reference allele column are those observed in the mother of the CEPH family 1347 (individual 134702). Therefore, 134702 can be used as a reference for allele size determination. An asterisk indicates markers for which data on 134702 allele size was not available; in such cases, the size was derived from the cloned sequence.

Notes: (aa), L22371 (*D1S333*), Weber, J.L. (1993), unpublished; (ab), M87678 (*D11S866*)[17]; (ac), M98989[2]; (ad), X71445 (*NTRK1*), Greco, A. (1993), unpublished; (ba), X67747, (*D2S211*)[18]; (ca); L02085[2]; (da), L22427, Weber, J.L. (1993), unpublished; (db), X51952 (*UCP*)[19]; (dc), L00809 (*D4S192*)[20]; (dd), L09826 (*D4S826*)[20]; (de), L00804 (*D4S610*)[20]; (ea), L22411 (*D5S811*), Weber, J.L. (1993), unpublished; (fa), Z19340 (*D0S6908E*), Genexpress, (1992), unpublished; (ga), L22426 (*D7S803*), Weber, J.L. (1993), unpublished; (gb), J03764 (*PLANH1*)[21]; (ha), M94655[2]; (ia), M83639[22]; (ib), L10620 (*D9S125*), Kwiatkowski (1993), unpublished; (ka), X52579 (*D11S35*)[23]; (kb), L20022, Weber, J.L. (1993), unpublished; (la), M96789 (*GJA4*)[24]; (ma), M99151 (*D13S144*)[25]; (mb), M99142 (*D13S121*)[25]; (na), L04461, (*D14S99E*)[26]; (pa), L02208 (*D16S318*)[27]; (ra), M88273 (*D18S37*)[28]; (sa), Z11689 (*PSG-11*)[29]; (sb), M36089 (*XRCC1*)[30]; (sc), M89651 (ref. 31); (wa), X60693 (*DXS571*)[32].