Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 7474–7476, September 1991
Medical Sciences

Parameters of the human genome

(physical map/genetic map/genomic size)

NEWTON E. MORTON

CRC Research Group in Genetic Epidemiology, University of Southampton, South Block, Southampton General Hospital, Southampton SO9 4XY, United Kingdom

Contributed by Newton E. Morton, April 23, 1991

ABSTRACT **Chromosome arm lengths are the critical parameters of the human genome. The physical length is required to scale radiation hybrid and other maps to megabases. The genetic lengths in males and females are required for probabilities of exclusion and synteny, choice of well-spaced loci for linkage tests, and comparison with centromeric maps based on nondisjunction. Interpolation of new data into a map is possible only when the length is known, including the distances from centromere and telomeres to the nearest markers. Current evidence on physical parameters includes reliable measurements of relative lengths from flow cytometry but only a crude estimate of genome size (3200 megabases). Evidence on genetic parameters includes chiasma counts and linkage maps corrected for failure to sample telomeres, giving an autosomal size of 2809 centimorgans in males and 4782 centimorgans in females. Estimates of the physical and sex-specific genetic lengths are presented for each chromosome arm. Any linkage analysis that yields substantially larger estimates raises a suspicion of an inappropriate mapping function or typing errors.**

Every science has parameters the accurate determination of which is essential, such as the speed of light for astronomy and Avogadro's number for chemistry. Immense effort has properly been devoted to their precision. The Human Genome Initiative, on the other hand, has not clearly recognized its parameters, which are, therefore, inexact.

PHYSICAL MAPS

The most critical parameter is genome size Γ, which old studies estimated to lie between 2500 and 3500 megabases (Mb) per haploid genome (1). A value of 3000 Mb is commonly used as a mean for both sexes (2). The significance of this parameter derives from the fact that the DNA content C_i of the ith chromosome or region is most accurately estimated as a proportion P_i of the haploid genome, and so $C_i = P_i\Gamma$. Therefore Γ determines the reliability of physical and composite maps. Only if the ith chromosome were partitioned into a complete set of nonoverlapping fragments γ_{ij} by pulsed-field gel electrophoresis or other methods would it be feasible to replace this top-down estimate by the bottom-up estimate $C_i = \Sigma_j \gamma_{ij}$. The size of a genome drawn from a female (XX) or a male (XY) at random is $\Gamma = A + 0.75X + 0.25Y$, where A is the size of a haploid set of autosomes.

Banded chromosomes are represented diagrammatically as an idiogram (this is sometimes misspelled ideogram, which has a different meaning). The relative size of chromosomes in an idiogram is usually based on linear measurement at some stage of mitosis or meiosis and may be distorted by differential contraction (3, 4). This distortion is most clearly seen for chromosome 19, which is slightly larger than chromosome 20 but contains less DNA. Relative lengths are obtained with greater precision by autoradiography, image cytometry, and flow cytometry. Representative values of P_i by these methods are given in Table 1. Variation is due not only to technical error but to length polymorphism, and so P_i should be estimated from a sample of individuals. In the last column mean values are converted to physical length, assuming tentatively that Γ = 3200 Mb. This value corresponds to 3.5 pg, which is typical of recent estimates (10).

Table 1. Physical lengths of human chromosomes

Chromosome	% of genome					Mb
	Ref. 5*	Ref. 6†	Ref. 7‡	Ref. 8§	Ref. 9¶	
1	8.29	8.21	8.17	8.25	8.12	263
2	7.87	8.04	7.97	8.06	7.94	255
3	6.72	6.69	6.67	6.67	6.63	214
4	6.28	6.35	6.35	6.38	6.33	203
5	5.97	6.10	6.07	6.11	6.02	194
6	5.66	5.70	5.71	5.77	5.68	183
7	5.30	5.29	5.37	5.29	5.36	171
8	4.75	4.83	4.87	4.87	4.82	155
9	4.48	4.57	4.59	4.53	4.47	145
10	4.60	4.51	4.51	4.45	4.47	144
11	4.56	4.49	4.51	4.55	4.47	144
12	4.43	4.43	4.51	4.49	4.47	143
13	3.44	3.61	3.62	3.55	3.68	114
14	3.37	3.41	3.46	3.44	3.36	109
15	3.25	3.32	3.30	3.23	3.36	106
16	3.11	3.09	3.08	2.96	3.09	98
17	2.92	2.82	2.81	2.85	2.89	92
18	2.71	2.69	2.67	2.67	2.67	85
19	2.15	2.09	2.07	2.06	2.21	67
20	2.27	2.21	2.21	2.29	2.26	72
21	1.62	1.57	1.48	1.57	1.64	50
22	1.80	1.73	1.70	1.64	1.79	56
X	5.27	5.06	5.15	5.16	5.07	164
Y	1.92	1.86	1.83	1.76	1.86	59

*Data are from autoradiography.
†Data are from image cytometry.
‡Data are from flow cytometry.
§Data are from image cytometry.
¶Data come from the mean of Hoechst and chromomycin in dual-beam flow cytometry.

It has been suggested that location be expressed relative to arm length (11). This begs the question of how numerator and denominator are specified and gives 100 arbitrary units per arm. One unit would correspond to 0.1 Mb for the short arm of Y chromosome (Yp) and more than 1 Mb for the long arm of chromosome 1 (1q). This inherent error makes relative location inferior to absolute location expressed in Mb from an origin, conventionally taken as the end of the short arm (pter).

GENETIC MAPS

Estimates of genetic length in centimorgans (cM) are based on cytogenetic and linkage data and so are subject to different

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: Mb, megabases; cM, centimorgans.

errors (Table 2). Cytogenetic counts of chiasmata may underestimate recombination because of conservative scoring, especially of terminal chiasmata. To minimize this bias estimates have been restricted to arms with terminal chiasmata. The usable chiasma data are confined to spermatogenesis and are known to be much different for oogenesis. Fortunately there is an estimate from genetic recombination of the ratio K of map lengths in females and males for each autosome, which was applied to the male chiasma data to estimate a corresponding female value (Table 2). This value may be too large because telomeric regions (in which there is sometimes a relative male excess) were incompletely sampled or too small because phenotypic errors inflate the male map disproportionately. To some extent these errors must be compensatory. For comparison we take map lengths from recombination among n loci, obtained by multiple pairwise analysis under a supported mapping function and multiplied by $(n + 1)/(n - 1)$ to adjust for failure to sample telomeric regions. This correction assumes that loci are sampled randomly from a uniform distribution along the genetic map (16). Although map length is subject to appreciable sampling error, the cytogenetic and genetic estimates are in fair agreement. Therefore, any linkage analysis that yields substantially larger estimates raises a suspicion of an inappropriate mapping function or typing errors. If each of n loci has a probability ε of typing error, the expected bias in map length by multipoint analysis is nearly $200(n - 1)\varepsilon$ cM. Multiple pairwise analysis on which these maps are based is more robust.

Table 2. Genetic lengths of human chromosomes

		Chiasma map,† cM			Linkage map,‡ cM		
Chromosome	K*	♂	♀	Mean	♂	♀	Mean
1	1.80‖	201	362	282	218‖	392‖	305‖
2	1.82‖	185	337	261	192‖	350‖	271‖
3	1.31	163	214	188	205	269	237
4	2.28	141	321	231	150	339	244
5	1.74	144	251	198	163¶	284	224
6	1.92**	146	280	213	142	272	207
7	1.36††	151	205	178			178
8	1.49††	138	206	172			172¶
9	1.27	129	164	146			146
10	1.59	129	205	167	148§	214§	181
11	1.59	116	184	150			150
12	1.76	142	250	196	116	205	160
13	1.60‡‡	100	160	130			130§§
14	1.26	100	126	113	108	136	122
15	1.95	100	195	148	104	203	154
16	1.61	111	179	145	120	193	157
17	2.45	114	279	196	120	295	208
18	1.53	113	173	143			143
19	1.93	100	193	146	101	195	148
20	2.60	100	260	180	68	176	122
21	1.42	58	82	70	94	133	114
22	1.31	70	92	81			81
X						220¶	

*Data are from ref. 12.
†Data are from ref. 13 (arms with terminal chiasmata).
‡Data are from ref. 14.
§Data are from unpublished work by D. C. Shields, A. Collins, K. H. Buetow, and N.E.M.
¶Data are from ref. 15.
‖Data are from ref. 16.
**Data are from ref. 12, omitting a poorly determined interval.
††K was estimated as $2(W_L/W_C) - 1$, where W_L is the sex-averaged length from linkage and W_C is the male length from chiasmata.
‡‡Data are the mean of chromosomes 14 and 15.
§§Data are from a chiasma map.

ARM LENGTHS

The ratio of the long arm (q) to chromosome length (p + q), called the centromere ratio, is consistent in three studies on the physical map (Table 3). The chiasma map gives an estimate of the centromere ratio in males. Compared with the physical data it is biased toward 0.5 for metacentrics and toward 1 for autosomal acrocentrics, reflecting an obligatory chiasma on each metacentric arm and Yp but not on short arms of autosomal acrocentrics. There are no data of the same reliability for females.

Table 4 summarizes all these results. For the physical map the centromere ratios are averaged and multiplied by the estimated chromosome length. In the genetic map the centromere ratio based on chiasmata in males was multiplied by the mean chromosome length for each sex, averaging estimates from chiasmata and linkage. Because observations on chiasmata in oocytes are lacking, genetic arm lengths in females must be less reliable than in males.

Due to these errors in the female genetic map, it may well be that at the present time a better estimate of arm length is $\overline{K}W_m$, where W_m is the corresponding male length and $\overline{K}$ is the ratio of autosomal lengths in females and males, estimated as $\overline{K} = 4782/2809 = 1.7$, as was found for linkage maps many years ago (18) and used to obtain the first estimates of genetic arm length on the now proven hypothesis that terminal chiasmata are included (19). The last decade has refined these estimates and increased support, especially for the physical parameters that were originally estimated from low-resolution idiograms (20). In the near future it will be feasible for some chromosomes to supplement these data by estimates of arm length from linkage, under an appropriate level of interference and with error filtration. The quantitative analysis of Francke and Oliver (21) may be used with caution to estimate physical length of chromosome bands because the measured length of high-resolution trypsin–Giemsa bands is approximately proportional to DNA content. Physical lengths of regions defined by chromosome breakpoints in

Table 3. Centromere ratios as q/(p + q)

	Physical map			Chiasma map
Chromosome	Ref. 17*	Ref. 8*	Ref. 6*	Ref. 13*
1	0.520	0.510	0.515	0.498
2	0.615	0.610	0.613	0.557
3	0.540	0.530	0.538	0.564
4	0.730	0.720	0.728	0.645
5	0.730	0.730	0.733	0.653
6	0.645	0.650	0.645	0.603
7	0.620	0.610	0.623	0.603
8	0.685	0.670	0.685	0.638
9	0.650	0.650	0.653	0.612
10	0.700	0.690	0.697	0.612
11	0.600	0.590	0.605	0.543
12	0.730	0.720	0.726	0.648
13	0.865	0.850	0.870	1.000
14	0.855	0.850	0.854	1.000
15	0.850	0.840	0.842	1.000
16	0.595	0.600	0.599	0.486
17	0.700	0.680	0.691	0.561
18	0.770	0.750	0.766	0.558
19	0.555	0.550	0.563	0.500
20	0.575	0.550	0.582	0.500
21	0.785	0.760	0.776	1.000
22	0.780	0.770	0.776	1.000
X	0.620	0.620	0.617	0.603†
Y	0.780	0.760	0.777	0.000‡

*Data are from indicated reference.
†This value is assumed to be the same as for chromosomes 6 and 7.
‡This value assumes an obligatory chiasma in Yp and none in Yq.

Table 4. Physical and genetic arm lengths

Chromosome	Physical length, Mb	Genetic length, cM ♂	Genetic length, cM ♀
1p	128	106	189
1q	135	104	188
2p	99	83	152
2q	156	105	192
3p	99	80	106
3q	115	104	136
4p	56	52	117
4q	147	94	213
5p	52	53	93
5q	142	101	175
6p	65	57	110
6q	118	87	166
7p	65	60	81
7q	106	91	124
8p	50	50	75
8q	105	88	131
9p	51	50	64
9q	94	79	100
10p	44	54	81
10q	100	84	129
11p	58	53	84
11q	86	63	100
12p	39	45	80
12q	104	84	148
13p	16	0	0
13q	98	100	160
14p	16	0	0
14q	93	104	131
15p	17	0	0
15q	89	102	199
16p	39	60	96
16q	59	56	90
17p	28	51	126
17q	64	66	161
18p	20	50	76
18q	65	63	97
19p	30	50	97
19q	37	50	97
20p	31	42	109
20q	41	42	109
21p	11	0	0
21q	39	76	108
22p	13	0	0
22q	43	70	92
Xp	62	50	87
Xq	102	0	133
Yp	13	50	
Yq	46	0	
Autosomes	3063	2809	4782

somatic-cell hybrids may be estimated by chromosome painting with a mixture of human probes or by summing fragments defined by rare-cutting endonucleases.

DISCUSSION

Physical arm lengths are required to scale radiation hybrid maps to megabases from their initial centiray units, which are dose-dependent. Physical maps constructed from fragments defined by chromosome breaks or restriction enzymes are often in arbitrary units that must be scaled to megabases. Genetic arm lengths are required for probabilities of exclusion and synteny, choice of well-spaced loci for linkage tests, and comparison of standard maps with centromere maps from nondisjunction in trisomies, triploids, or ovarian teratomas. Interpolation of physical or genetic data into a map is contingent on the relevant arm length especially for pericentric and subtelomeric intervals. A composite map that reconciles physical, genetic, and cytogenetic information should be scaled in megabases and, therefore, depends on arm lengths, which are the basic parameters of the human genome.

The short-term goal of the Human Genome Initiative is a map of several thousand loci the locations of which are accurately specified in both physical (Mb) and genetic (cM) units from either telomere. This mapping is feasible only when an appropriate level of statistical support is adopted and when map lengths are known. This note will have served its purpose if it directs attention to this problem and leads to more accurate estimates, on which the truth of the map depends.

I am grateful to Darryl Green and Adrian Sumner for helpful discussion of the physical parameters.

1. Shapiro, H. S. (1976) in *Handbook of Biochemistry and Molecular Biology*, ed. Fasman, G. D. (CRC, Boca Raton, FL), Vol. 2, pp. 284–306.
2. Bodmer, W. F. (1981) *Am. J. Hum. Genet.* **33**, 664–682.
3. Harnden, D. B. & Klinger, H. P. (1985) *An International System for Human Cytogenetic Nomenclature* (Karger, Basel).
4. Van Dyke, D. L., Worsham, M. J., Fisher, L. J. & Weiss, L. (1986) *Hum. Genet.* **73**, 130–132.
5. Korenberg, J. R. & Engels, W. R. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 3382–3386.
6. Mayall, B. H., Carrano, A. V., Moore, D. H., Ashworth, L., Bennett, D. E. & Mendelsohn, M. L. (1984) *Cytometry* **5**, 376–385.
7. Harris, P., Boyd, E., Young, B. D. & Ferguson-Smith, M. A. (1986) *Cytogenet. Cell Genet.* **41**, 14–21.
8. Mendelsohn, M. L., Mayall, B. H., Bogart, E., Moore, D. H. & Perry, B. H. (1973) *Science* **179**, 1126–1129.
9. Langlois, R. G., Yu, L.-C., Gray, J. W. A. & Carrano, A. V. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 7876–7880.
10. Tiersch, T. R., Chandler, R. W., Wachtel, S. S. & Elias, S. (1989) *Cytometry* **10**, 706–710.
11. Kidd, K. K., Bowcock, A. M., Schmidtke, J., Track, R. K., Ricciuti, F., Hutchings, G., Bale, A., Pearson, P. & Willard, H. F. (1989) *Cytogenet. Cell Genet.* **51**, 622–947.
12. Keats, B. J. B., Sherman, S. L. & Ott, J. (1990) *Cytogenet. Cell Genet.* **55**, 387–394.
13. Morton, N. E., Lindsten, J., Iselius, L. & Yee, S. (1982) *Hum. Genet.* **62**, 266–270.
14. Keats, B., Ott, J. & Conneally, M. (1989) *Cytogenet. Cell Genet.* **51**, 459–502.
15. Morton, N. E. (1988) *Ann. Hum. Genet.* **52**, 309–318.
16. Morton, N. E. & Collins, A. (1990) *Ann. Hum. Genet.* **54**, 235–251.
17. Lucas, J. N. & Gray, J. W. (1987) *Cytometry* **8**, 273–279.
18. Weitkamp, L. R. (1972) in *Human Genetics*, eds. De Grouchy, J., Ebling, F. J. G. & Henderson, I. W. (Excerpta Med., Amsterdam), pp. 445–460.
19. Keats, B. J. B., Morton, N. E., Rao, D. C. & Williams, W. R. (1979) *A Source Book for Linkage in Man* (Johns Hopkins Press, Baltimore, MD).
20. Bergsma, D., Lindsten, J. E., Klinger, H. P. & Hamerton, J. L. (1978) *An International System for Human Cytogenetic Nomenclature* (Karger, Basel).
21. Francke, U. & Oliver, N. (1978) *Hum. Genet.* **45**, 137–165.