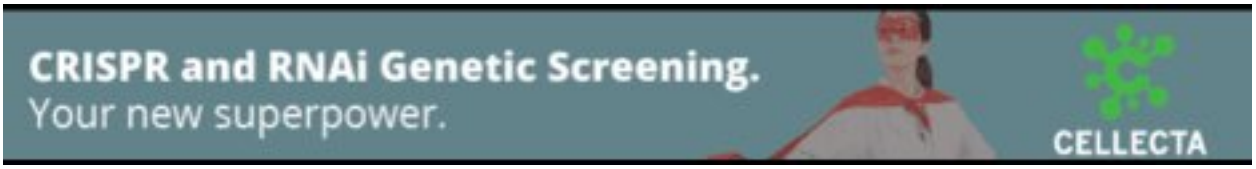
A High-Resolution Microsatellite Map of the Mouse Genome

Michael Rhodes, Richard Straw, Supem Fernando, et al.

Genome Res. 1998 8: 531-542
Access the most recent version at doi:10.1101/gr.8.5.531



RESEARCH

A High-Resolution Microsatellite Map of the Mouse Genome

Michael Rhodes,[1] Richard Straw,[1] Supem Fernando,[1] Andrew Evans,[1] Tregaye Lacey,[1] Andrew Dearlove,[1] John Greystrong,[1] Joanne Walker,[1] Paula Watson,[1] Paul Weston,[1] Maria Kelly,[1] Dilip Taylor,[1] Keith Gibson,[1] Chris Mundy,[1] Franck Bourgade,[2] Christophe Poirier,[2] Dominique Simon,[2] Ana Lucia Bueno Brunialti,[2] Xavier Montagutelli,[2] Jean-Louis Guénet,[2] Andy Haynes,[3] and Steve D.M. Brown[3,4,5]

[1]United Kingdom Human Genome Mapping Project (HGMP) Resource Centre, Hinxton CB10 1RQ, UK; [2]Institut Pasteur, Paris, France; [3]Medical Research Council Mouse Genome Centre and [4]Medical Research Council Mammalian Genetics Unit, Harwell OX11 ORD, UK

The European Collaborative Interspecific Backcross (EUCIB) resource was constructed for the purposes of high-resolution genetic mapping of the mouse genome (Breen et al. 1994). The large *Mus spretus*/C57BL/6 backcross of 982 progeny has a genetic resolution of 0.3 cM at the 95% confidence level (~500 kb in the mouse genome). We have used the EUCIB mapping resource to develop a genome-wide high-resolution genetic map incorporating 3368 microsatellites. The microsatellites are distributed among 2302 genetically separated bins with 1.46 markers per bin on average. Average bin separation is 0.61 cM. This high-resolution genetic map will aid the construction of a robust physical map of the mouse genome.

The mouse is a pivotal model organism for the genome program and with its battery of mutagenic, transgenic, and developmental biology approaches is set to play a key role as mammalian genetics moves from genomics to studies of gene function (Copeland et al. 1993; Dietrich et al. 1995; Brown and Peters 1996). The development of genetic and physical maps in the mouse is an important step toward providing the genome resources for future gene function studies (Dietrich et al. 1995). The construction of a high density map of mouse simple-sequence polymorphisms at intermediate resolution is complete (Dietrich et al. 1996). The future development of genome-wide physical maps (Hudson et al. 1996) will assist gene mapping as well as providing the clone resources for gene identification. In addition, the map will provide the substrates for preparing sequence-ready maps for comparative sequencing, which will itself speed the process of gene identification. Physical maps will also underpin the development of comprehensive, high-resolution gene maps (Schuler et al. 1996) that can be used for the characterization of mouse mutations and provide a new cache of gene sequences that can be related to loci on the human genome by the conserved linkage groups identified between the two species (Copeland et al. 1993; Andersson et al. 1996). Nevertheless, there are no complete physical maps yet available for any mouse chromosome. The development of a high-resolution genetic map can enhance the production of a robust physical map on any mouse chromosome.

The ability to undertake large genetic crosses between defined mouse strains means the construction of high-resolution genetic maps can be readily achieved. Most notably, large interspecific or intersubspecific backcrosses between laboratory strains of mice and wild species such as *Mus spretus* or *Mus castaneus* (Avner et al. 1988) has transformed mouse genetic mapping (for review, see Copeland et al. 1993). Large numbers of backcross progeny can be readily derived from such crosses providing the requisite numbers of meioses to achieve high genetic resolution. Additionally, the use of crosses between relatively diverged species contributes to the large numbers of markers that are variant between the parental strains. Approximately 90% of microsatellites show size variation between laboratory strains and the wild species, *M. spretus* and *M. castaneus*

[5]Corresponding author.
E-MAIL s.brown@har.mrc.ac.uk; FAX 01235 824542.

(Dietrich et al. 1992). Small intersubspecific crosses have been used for the construction of genome-wide gene or microsatellite maps at intermediate resolution (Dietrich et al. 1996; see above). Large interspecific backcrosses of a 1000 progeny or more carrying a specific mutation of interest have been used widely for high-resolution genetic mapping of the mutant locus as a route to positional cloning of the gene (Brown 1994, 1996). However, to date there has been no systematic attempt to use the high resolution afforded by large interspecific backcrosses to construct genome-wide high-resolution maps.

Recently, we reported the construction of a high-resolution mouse mapping resource consisting of an interspecific backcross of nearly 1000 progeny (Breen et al. 1994). A backcross of this size has a genetic resolution of 0.3 cM at the 95% confidence level that equates to ~600 kb in the mouse genome. We have now used this backcross to construct a high-resolution and high-density microsatellite map of the mouse genome. This high-resolution genetic map will be the anchor for the construction of a high integrity physical map of the mouse genome.

RESULTS

Identification of Panels of Recombinants from EUCIB for High-Resolution Mapping

We have described the construction of a large interspecific backcross between C57BL/6 and *Mus spretus*—the European Collaborative Interspecific Backcross (EUCIB)—comprising 982 backcross progeny. Backcross progeny were initially typed for 78 primary anchor loci spanning the entire genome with 3–6 anchors per chromosome (Breen et al. 1994). Subsequently, a number of additional anchor markers were added. Where it became apparent from further mapping studies that the proximal and distal anchors available on a particular chromosome did not represent either the most centromeric or telomeric markers, additional anchors were added for the mapping of markers close to the centromere and telomere. The anchor map identifies the great majority of backcross progeny mice recombinant in any interanchor interval (excluding only those rare individuals that have double interanchor recombinants) and provides panels of mice for high-resolution mapping in each interanchor chromosome region. In total, 93 primary anchors were assigned (see Table 1). Subsequently, a large number of secondary anchors (principally microsatellite markers) were added to the map reducing the size of recombinant panels still further and allowing for rapid high-resolution mapping of markers in any chromosome region. Final interanchor intervals comprised panels of ~36 recombinant mice on average and thus corresponded to a genetic interval of ~4 cM.

Mapping Microsatellites at High-Resolution on the EUCIB Backcross

To develop the high resolution genetic map, a large number of microsatellites markers from the Whitehead/MIT map (Dietrich et al. 1996) were analyzed through the EUCIB backcross. The bulk of microsatellite mapping used a novel, high-throughput and semiautomated fluorescent dUTP genotyping approach (Rhodes et al. 1997)—2278 of the total of 3368 were added to the map by this approach. The remainder were mapped either by use of standard agarose gel electrophoresis or alternatively with an enhanced chemiluminescence approach (Vignal et al. 1993). Following the determination of the parental allele sizes (C57BL/6 and *M. spretus*), the appropriate recombinant panel of mice was genotyped. Given the limited resolution afforded by previous maps, it was not always apparent which interanchor interval a microsatellite would lie within and, therefore, which recombinant panel should be typed. Under these circumstances, appropriate adjacent panels were typed.

Although we tested all the Whitehead/MIT primers available during the period of map construction, inevitably a proportion failed to amplify product or proved problematic for reliable scoring and mapping. For the 4450 microsatellite markers tested by the semiautomated fluorescent dUTP genotyping approach, 51.2% amplified and were mapped successfully. Of the 48.8% that failed to be added to the map, 14.7% failed to give any product whatsoever (on either C57BL/6 or *M. spretus* DNA) and 31.1% produced some product but was not scoreable (e.g., multiple bands or variable product sizes). A small percentage, 3.0%, gave reliable, but identical, products between C57BL/6 and *M. spretus* DNA and were therefore not mappable.

Given the high throughput requirements of the project, we did not return to failed primer sets to optimize PCR conditions, and as a consequence, a proportion of microsatellites were not added to the map on each chromosome. Ultimately, of the Whitehead/MIT microsatellites available to us during map construction, 56% were added to the EUCIB map. For individual chromosomes, the proportion of Whitehead/MIT microsatellites placed on

Table 1. Summary of Markers and Map Statistics by Chromosome

Chromosome	Primary anchors	Microsatellites	Percent MIT	Totals	Bins	Length (cM)
1	4	226	48	230	153	90
2	5	331	70	336	197	95
3	4	176	52	180	128	77
4	5	208	64	213	158	84
5	4	225	60	229	137	92
6	6	195	58	201	139	78
7	4	201	62	205	148	71
8	3	172	52	175	127	62
9	6	143	48	149	107	73
10	6	142	53	148	105	68
11	5	142	45	147	97	72
12	5	149	58	154	116	60
13	4	178	61	182	116	68
14	5	130	54	135	101	53
15	3	144	58	147	91	62
16	3	110	55	113	77	56
17	4	117	52	121	90	59
18	7	115	56	122	67	55
19	4	67	57	71	54	54
X	6	104	52	110	94	69
Totals	93	3275	56	3368	2302	1398

The number of primary anchors and microsatellites mapped to each chromosome is indicated. Percent MIT indicates the proportion of Whitehead/MIT microsatellites available to us at the time of mapping that were added to the EUCIB genetic map. In addition, the numbers of marker bins for each chromosome is given along with the genetic map length (the distance between the most proximal and distal anchors mapped–see Methods).

the EUCIB map varied from 45% to 70% of Whitehead/MIT microsatellite markers (see Table 1).

The *MBx* Database—Construction of the EUCIB High-Resolution Microsatellite Map

The *MBx* database that supports the EUCIB program has been described previously (Breen et al. 1994). Genotypes were entered into the *MBx* database and genetic maps were constructed. Determining locus order rather than genetic distance was the primary consideration for the construction of genetic maps because this provides the most important enhancement to future physical maps that will be underpinned by the high-resolution genetic map. The order of microsatellite markers along each chromosome was determined by a haplotype analysis that minimizes the recombinants in any chromosomal region. The genetic distances displayed in the EUCIB Genetic Map and MultiMaps are calculated so that the marker order derived by haplotype analysis is maintained (see Methods); these displays (see Fig. 1) are available on the World Wide Web site:(URL:http://www.hgmp.mrc.ac.uk/MBx/MBxHomepage.html), which also includes genotype data for individual markers on the maps. Direct access to the *MBx* database to view haplotypes is also available (see Methods). The *MBx* database provides scrollable tables of haplotypes for each chromosome, identifying and highlighting all recombination events. It is possible to select and display all mice containing recombination events in a particular interval to assess the raw data and to evaluate how robust locus order is.

The EUCIB High-Resolution Genetic Map

In total, 3368 microsatellites and anchors have been mapped and ordered at high resolution on the EUCIB backcross (see Table 1). The total length of the high resolution EUCIB genetic map is 1398 cM as calculated from proximal–distal anchor distances on each chromosome (Table 1). On each chromo-

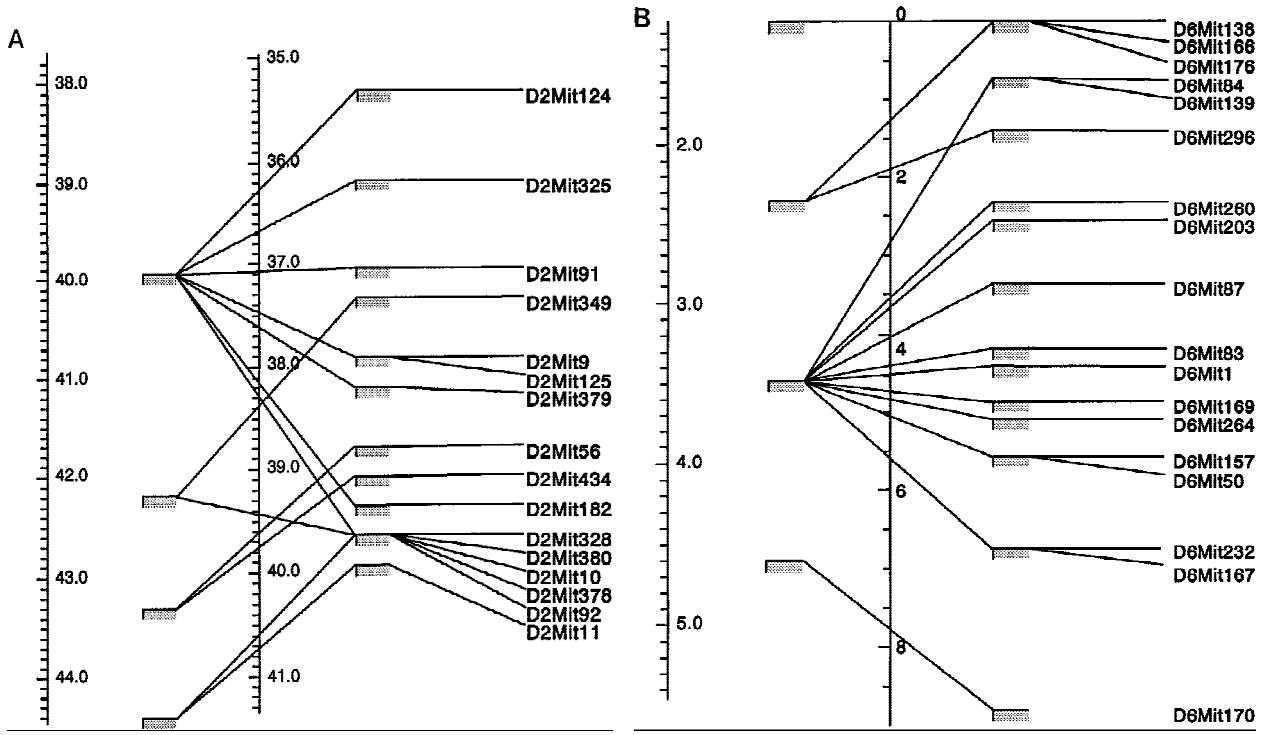


Figure 1 Multimap comparisons of regions of the Whitehead/MIT microsatellite map and the EUCIB map from chromosome 2 (*A*) and 6 (*B*) demonstrating increases in bin number and resolution by use of the EUCIB resource. Genetic map positions of Whitehead/MIT bins are shown at *left* with the genetic map position of the EUCIB bins at *right.*

some, a few microsatellites map beyond the most proximal or most distal anchors. However, for these markers it is not feasible to determine accurate anchor–marker distances (see Methods) and for this reason, these markers have not been included in the calculation of total genetic length. Thus, 1398 cM represents the minimum length of the genetic map. The total calculated length of the EUCIB map is very similar to that reported from the genome-wide Frederick C57BL/6 × *M. spretus* backcross map (Copeland and Jenkins 1991) whose genetic length (as reported in Dietrich et al. 1996) is 1385 cM.

The high resolution of the EUCIB map is reflected by the fact that the microsatellites were distributed among 2302 bins giving 1.46 markers per bin on average—a bin being defined as a minimally resolved genetic interval separated from adjacent bins by proximal and distal recombination breakpoints for all markers in that bin. Table 2 illustrates the distribution of markers per bin. Seventy-one percent of the bins contain only 1 marker, again reflecting the high resolution of the mapping process. Most notably, the average separation between bins is 0.61 cM. The average intermarker interval is 0.42 cM or ~800 kb.

Figure 1 shows a multimap comparison of the EUCIB and Whitehead/MIT maps and illustrates the typical increase in resolution afforded by the EUCIB high-resolution mapping resource. It is important to

Table 2. Marker Distribution Across Bins

No. of markers per bin	No. of bins
1	1640
2	427
3	140
4	51
5	29
6	9
7	4
8	0
9	0
10	0
11	1
12	1

The numbers of markers in bins carrying 1, 2, 3, 4 markers, etc., is given.

note when comparing the EUCIB and Whitehead/MIT maps that for the Whitehead/MIT map (Dietrich et al. 1994) constructed from a limited number of F_2 intercross progeny, statistical support for order of a given marker could vary, either because of incomplete genotyping or because the marker is dominant rather than codominant. Thus multimap comparisons of Whitehead/MIT and EUCIB maps illustrate, not unsurprisingly, that markers lying in adjacent bins on the Whitehead/MIT map are sometimes found to interdigitate when their order is determined at high resolution on the EUCIB map (see Fig. 1). For this reason, we have chosen a relatively high cutoff point to assess the frequency of markers that deviate in position between the EUCIB and Whitehead/MIT maps. Proceeding systematically proximal to distal on each chromosome, we coaligned each EUCIB primary or secondary anchor marker with the corresponding locus on the Whitehead/MIT map, and then assessed each microsatellite lying in the following inter-anchor segment on the EUCIB map for significant deviations with the Whitehead/MIT map. (Only markers lying between the most proximal and distal anchors on each chromosome were subject to this analysis to avoid biases from the few poorly mapped markers lying outside these anchor loci.) Choosing a cutoff point of 10 cM, only 76 markers (2.3%) on the EUCIB map mapped 10 cM or more from their expected location on the basis of the Whitehead/MIT map. At 15 and 20 cM, this figure dropped to 1.1% and 0.8%, respectively. Overall, there is excellent agreement between Whitehead/MIT and EUCIB maps.

Distribution of Markers and Recombination Events

Only two large bins—one of 11 and one of 12 markers (see Table 2)—remain on the final EUCIB map. We examined these bins to see if they corresponded to any of the large bins on the Whitehead/MIT map that may be accounted for by regions of crossover suppression that are common between the two crosses used for mapping (the Whitehead/MIT map was constructed by use of a (C57BL/6J-*ob/ob* × *Mus castaneus*) F2 intercross (see Dietrich et al. 1996). The bin of 11 markers is found on chromosome 16 at position 6.34 cM and includes the microsatellites *D16Mit8, 32, 33, 79, 121, 122, 131, 142, 161, 180,* and *181*. The bin of 12 markers is found on chromosome 2 at position 32.23 cM and includes the microsatellites *D2Mit8, 72, 89, 155, 157, 240, 241, 298, 323, 373, 433,* and *471*. However, neither of these bins corresponded to large, unseparated bins of markers on the Whitehead/MIT map (Dietrich et al. 1996).

We have also examined the EUCIB maps to determine the largest genetic gaps. Only five gaps >5 cM were identified between the most proximal and distal anchors. On chromosome 1, a gap of 8.59 cM separates *D1Mit65* and *D1Mit118*. On chromosome 5, a gap of 7.38 cM was found to separate *D5Mit160* and *D5Nds6*, whereas on chromosome 9 two gaps between *D9Mit217* and *D9Mit58* (6.60 cM) and *D9Mit294* and *D9Mit42* (5.44 cM) were identified. On chromosome 18, a gap of 9.68 cM was identified between *D18Mit33* and *D18Mit8*. None of these gaps corresponded to any of the larger genetic intervals on the Whitehead/MIT map (Dietrich et al. 1996; see Discussion).

In total, 17,029 distinct recombination events were observed in the *MBx* database among the 982 progeny and distributed across all 20 chromosomes (see Table 3). On average, each mouse carries 17.3 recombination events. Forty-four percent of chromosomes did not show an observable recombination event. Forty percent of chromosomes (7863) showed a single recombinant, 7.2% of chromosomes demonstrated double recombinants, and 5.6% triple recombinants. We have used a goodness-of-fit test to analyze the distribution of recombinant classes on each chromosome for fit to Poisson. None of the 20 chromosomes fit a Poisson distribution, differing significantly in all cases (see Table 3). As observed for the original EUCIB anchor map (Breen et al. 1994), and in agreement with other reports (Ceci et al. 1989; Saunders and Seldin 1990; Nadeau et al. 1991; Reeves et al. 1991, 1997), there is a general over-representation of single recombinants, and double recombinants are underrepresented on all chromosomes—probably because of crossover suppression. In general, triple recombinant classes and classes carrying larger numbers of recombinants are over-represented—probably largely because of genotyping errors (see below). In total, 10,991 chromosomes (56%) carried one or more crossovers. In general, there was a broad relationship between a chromosome's genetic length and the total number of recombinant events observed for that chromosome. Overall, however, the relationship between numbers of recombinants per chromosome and genetic length across all 20 chromosomes was not significant (see legend to Table 3). A binomial test to identify those chromosomes that contribute significantly fewer or greater recombinants than expected indicates that chromosomes 1, 7, 8, and 18 contribute significantly more recombinants than expected, whereas chromosomes 3, 11,

Table 3. Number and Distribution of Recombinants by Chromosome

Recomb. class	Chromosome										
	1	2	3	4	5	6	7	8	9	10	11
0	316 (282)	332 (281)	447 (418)	350 (318)	312 (291)	376 (342)	443 (371)	395 (347)	433 (414)	448 (400)	438 (490)
1	378 (352)	352 (352)	369 (367)	382 (359)	407 (354)	395 (361)	344 (361)	385 (361)	401 (358)	366 (359)	437 (341)
2	141 (220)	157 (220)	80 (152)	119 (202)	130 (215)	89 (190)	69 (176)	84 (188)	60 (154)	69 (161)	88 (118)
3	81 (91)	67 (92)	57 (43)	71 (76)	72 (87)	71 (67)	76 (57)	66 (65)	50 (44)	56 (48)	17 (27)
4	33 (29)	42 (29)	19 (9)	40 (21)	28 (27)	27 (18)	22 (14)	15 (17)	22 (10)	19 (11)	0 (5)
5	19 (7)	16 (7)	4 (2)	10 (5)	16 (7)	14 (4)	19 (3)	24 (4)	10 (2)	17 (2)	1 (1)
6	8 (2)	11 (2)	3	7 (1)	10 (1)	5 (1)	5	6 (1)	3	2	0
7	4	1	1	2	2	2	2	3	3	4	0
8	1	2	1	1	2	2	1	2	0	1	0
9	0	0	1	0	1	0	0	1	0	0	0
10	0	1	0	0	2	0	1	1	0	0	0
11	1	0	0	0	0	1	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	1
14	0	0	0	0	0	0	0	0	0	0	0
15	0	1	0	0	0	0	0	0	0	0	0
χ^2	101.38	98.9	70.03	89.07	123.72	150.85	282.99	317.6	125.38	128.19	46.81
Total recomb.	1225	1229	838	1107	1194	1035	955	1023	848	881	682
Length (cM)	90	95	77	84	92	78	71	62	73	68	72
Binomial test	16.151*	4.78	11.269*	7.302	5.14	8.032	9.9*	99.353*	2.015	3.524	45.725*

Table 3. (*Continued*)

Recomb. class	Chromosome 12	13	14	15	16	17	18	19	X	Total chr.	Total recombs.	Percent Total Chr.
0	462 (475)	475 (450)	578 (547)	503 (552)	507 (557)	474 (530)	467 (457)	483 (505)	410 (425)	8649	0	44.038
1	406 (345)	383 (351)	329 (320)	429 (318)	417 (316)	445 (327)	397 (350)	412 (336)	429 (356)	7863	7863	40.036
2	50 (125)	42 (137)	12 (94)	21 (92)	38 (90)	31 (101)	18 (134)	36 (112)	72 (149)	1406	2812	7.159
3	56 (30)	58 (36)	48 (18)	24 (18)	18 (17)	29 (21)	93 (34)	42 (25)	47 (42)	1099	3297	5.596
4	4 (6)	8 (7)	2 (3)	4 (3)	0 (2)	3 (3)	0 (7)	3 (4)	17 (9)	308	1232	1.568
5	2 (1)	10 (1)	9	0	2	0	5 (1)	4 (1)	4 (2)	186	930	0.947
6	1	3	3	0	0	0	1	2	1	71	426	0.362
7	1	1	1	1	0	0	0	0	2	30	210	0.153
8	0	1	0	0	0	0	0	0	0	14	112	0.071
9	0	0	0	0	0	0	1	0	0	4	36	0.020
10	0	0	0	0	0	0	0	0	0	5	50	0.025
11	0	1	0	0	0	0	0	0	0	3	33	0.015
12	0	0	0	0	0	0	0	0	0	0	0	0.000
13	0	0	0	0	0	0	0	0	0	1	13	0.076
14	0	0	0	0	0	0	0	0	0	0	0	0.000
15	0	0	0	0	0	0	0	0	0	1	15	0.088
χ^2	78.53	114.52	168.93	101.41	66.55	100.31	208.27	85.23	73.88			
Total recomb.	713	767	575	566	557	606	752	654	822		17029	
Length (cM)	60	68	53	62	56	59	55	54	69			
Binomial test	0.456	4.77	8.023	49.609*	23.913*	18.445*	10.46*	0.022	0.428			

The number of nonrecombinant and recombinant classes observed for each chromosome are listed along with the numbers of classes expected according to a Poisson distribution (in parentheses). Where the expected number is <1, no figure is given. χ^2 values from goodness-of-fit tests applied to each chromosome are given. All chromosomes differ very significantly from a Poisson distribution at $p < 10^{-9}$. Although there appears to be a broad relationship between numbers of total recombinants per chromosome and genetic length (cM), the observed distribution deviates significantly from that expected if the number of recombinants varies in proportion to genetic length ($\chi^2 = 313.64$, $P < 10^{-10}$; 19 *df.*). A binomial test of significance has also been applied to each chromosome to identify those chromosomes that depart most significantly from expected, and the χ^2 values for each chromosome are listed. We have taken account of the likelihood of false positives, i.e., chromosomes that depart from expected by chance. Monte Carlo simulations indicate that when adopting a significance level of 0.05, one false-positive chromosome will arise with a frequency of 0.6317. Reducing the false-positive rate to 0.05 the binomial test needs to be applied at a significance level of 0.002. At $\alpha = 0.002$ chromosomes 1, 3, 7, 8, 11, 15, 16, 17 and 18 show a significant departure from expected (indicated by asterisk).

15, 16, and 17 contribute significantly fewer (see legend to Table 3).

Error Correction

Multiple recombinants observed on an individual chromosome (see above) can be characteristic of genotyping errors. For this reason, we have sought to identify those multiple recombinants that may be most indicative of genotyping errrors to reach some assessment of the overall level of genotyping error in the EUCIB dataset. Error correction on each chromosome has been proceeded by the identification of haplotypes 1(.)2(.)1 and 2(.)1(.)2 for adjacent markers within the database where 1 represents a homozygote, 2 a heterozygote and (.) represents a variable number of intervening markers for which genotype information might not be available in any haplotype (see legend to Table 4). These apparently double recombinant haplotypes for very closely linked markers would be expected to occur rarely, if at all, and contribute to a proportion of double and triple recombinant chromosomes and to chromosomes carrying larger numbers of recombinants (see above). Subsequently, having identified these double recombinant haplotypes, primary data entry is checked, or in some cases scorings are repeated. Eventually, when error correction is complete, reordering of markers is carried out by *MBx*. Nevertheless, following error correction and reordering, some aberrant haplotypes remain (see Table 4 for the total number of spurious double recombinant haplotypes per chromosome). In total, 1158 double recombinants of the form 1(.)2(.)1 or 2(.)1(.)2 remain. Across the whole dataset, 41% of these double recombinants are of the form 121 or 212 with no intervening unscored markers. Sixty-six percent of double recombinants are of the form 121, 212 and 1(.)2(.)1 or 2(.)1(.)2 in which only a single intervening marker (.) is unscored. The high frequency of these apparently closely spaced double recombinants is very suggestive of genotyping errors rather than true recombination events.

In general, if we assume that all aberrant haplotype events represent genotyping errors, then the average genotyping error rate across the whole genome is ~0.01. The rate varies from chromosome to chromosome, therefore, on some chromosomes, for example, chromosome 16, it is as low as 0.002. However, we have undertaken additional analyses to empirically estimate the residual error rate for genotypes within the EUCIB dataset. Additionally, we have attempted to estimate the overall error rate for marker order across the genome.

Table 4. Number and Chromosome Distribution of Aberrant Double and Triple Recombinants

Chromosome	Double recombinants 1(.)2(.)1 2(.)1(.)2	Triple recombinants 11(.)2(.)1(.)22 22(.)1(.)2(.)11
1	83	8
2	47	4
3	44	18
4	111	16
5	65	1
6	86	19
7	120	13
8	106	9
9	63	10
10	73	7
11	16	2
12	52	6
13	46	4
14	57	5
15	15	10
16	8	0
17	28	2
18	28	2
19	46	7
X	64	17
Totals	1158	160

Double recombinants of the form 1(.)2(.)1 and 2(.)1(.)2 (1, homozygote; 2, heterozygote) that may represent genotyping errors (see text) were identified for each chromosome from the *MBx* database. (.) represents a variable number of intervening markers with no genotypes for a haplotype. Up to 10 intervening markers for which there was no genotype information were permitted. Thus, at the limit, double recombinants of genotype 1 (10 markers, no genotypes) 2 (10 markers, no genotypes) 1, and 2 (10 markers, no genotypes) 1(10 markers, no genotypes) 2 were identified. For chromosomes 1 and 5, a further round of error checking of double recombinant genotypes and reordering took place for these chromosomes to assess the underlying error rate (see text). Triple recombinants of the form 11(.)2(.)1(.)22 or 22(.)1(.)2(.)11, which may represent local misorders (see text), were also identified from *MBx*. Up to 10 intervening markers (.) for which there is no genotype information were again permitted.

Genotyping Error Rate

Following the final rounds of data production, error checking, and ordering, we chose two chromosomes—1 and 5—and identified all remaining 1(.)2(.)1 and 2(.)1(.)2 haplotypes for adjacent markers from within the dataset. Some 1(.)2(.)1 and 2(.)1(.)2 haplotypes will occur as part of more

complex triple recombinant haplotypes [e.g., 22(.)1(.)2(.)11] among adjacent markers. These triple recombinant haplotypes can arise because of errors in typing, or more likely, errors in ordering in which the inversion of the central two markers would remove the triple recombinant haplotype and substitute a single recombinant haplotype in its place (see below). Nevertheless, for chromosomes 1 and 5, the aberrant genotypes were rescored in all cases. This involved repeating the appropriate PCR reactions under identical reaction conditions. Following retyping, these chromosomes were reordered. Table 5 gives the reduction in the number of 1(.)2(.)1 and 2(.)1(.)2 haplotypes observed on each chromosome following second rounds of error checking and ordering and, therefore, a more accurate figure of the genotyping error rate. Taking both chromosomes together, of the original aberrant haplotype genotypings, 48% (134) were found to be incorrect, giving an error rate in genotyping for these two chromosomes of 0.008 that is, in general, in agreement with the genome-wide figure quoted above. Extrapolating to the whole genome, if ~50% of the observed aberrant double recombinants represent genotyping errors, then the overall error rate is ~0.005 or 1 in 200 genotypes in the database. What is notable is that a significant number of 1(.)2(.)1 and 2(.)1(.)2 haplotypes remain on each chromosome despite this second round of error checking.

By and large, these aberrant haplotypes do not result from misorders because the level of triple recombinants is very low (see below). We have also considered the possibility that some or all of these aberrant double recombinants arise because of residual heterozygosity within the *M. spretus* mice used in establishing the backcross. That part of the EUCIB backcross performed in London used *M. spretus* animals from a colony that had not been systematically inbred, whereas *M. spretus* animals used in Paris were from the SEG/Pas colony that is moderately inbred after 20 generations of unrelaxed brother-sister matings (Breen et al. 1994). Apparent double recombinant chromosomes of the form SSBSS (where S and B are the *M. spretus* and BL/6 alleles, respectively) could arise if there is residual heterozygosity in the *M. spretus* parents with a B rather than an S allele present at the supposed double recombinant locus in some members of the parent *M. spretus* population. In the backcross to BL/6, an SSBSS haplotype inherited from the F1 would be scored as a 2(.)1(.)2 haplotype in *MBx*—which we have designated a B1 haplotype. Conversely, in the backcross to *M. spretus*, an SSBSS haplotype would be scored as 1(.)2(.)1 in *MBx*, designated a S2 haplotype. Both B1 and S2 classes might be expected to be in excess if residual heterozygosity was a significant factor. However, residual heterozygosity from the *M. spretus* parent population could not account for haplotypes of the form BBSBB. In the backcross to BL/6, an BBSBB haplotype inherited from the F1 would be scored as a 1(.)2(.)1 haplotype in *MBx*—which we have designated a B2 haplotype. Conversely, in the backcross to *M. spretus*, an BBSBB haplotype would be scored as 2(.)1(.)2 in *MBx*—designated a S1 haplotype. Overall, we find that there are in total 606 double recombinant haplotypes in the B1 + S2 class, whereas there are 552 haplotypes in the B2 + S1 class. Residual heterozygosity does not, therefore, appear to be a major factor in the appearance of aberrant double recombinants.

Table 5. Assessing Genotype Error Rates in EUCIB

Chromosome	Double recombinants 1(.)2(.)1/2(.)1(.)2		
1	125[a]	86[b]	83[c]
5	157[a]	84[b]	65[c]

Reduction in the numbers of double recombinants following further rounds of regenotyping and reordering on chromosomes 1 and 5 are given.
[a]After data production, error checking, and ordering.
[b]Remaining double recombinants following regenotyping.
[c]Remaining double recombinants following reorder.

Marker Order Error Rate

Following retyping and reordering, we also identified on every chromosome all triple recombinant haplotypes that remained—11(.)2(.)1(.)22 and 22(.)1(.)2(.)11—and that potentially represent local misorders. The numbers of triple recombinant haplotypes remaining on each chromosome are also given in Table 4. Some chromosomes had no detectable triple recombinants and in total, across the genome, we found 160 haplotypes representing the likely total number of locally misordered markers.

DISCUSSION

We have constructed the first high-resolution genetic map for a mammalian species. The EUCIB high-resolution microsatellite map has allowed us to order markers to 2302 bins providing a bin sepa-

ration at ~0.6 cM and a robust framework on which to complete the physical map of the mouse genome. Seventy-one percent of bins contain only one marker and only two large bins of 11 and 12 markers remain among the 3368 microsatellites mapped. Five gaps in the genetic map >5 cM remain on chromosomes 1, 5, 9, and 18. These do not correspond to regions of poor coverage on the EUCIB map because many of the Whitehead/MIT microsatellites from these regions were added to the map. In addition, the gaps in the EUCIB maps do not correspond to any of the larger genetic intervals remaining on the Whitehead/MIT genetic map (Dietrich et al. 1996). Overall, it would appear that these gaps potentially correspond to recombination hotspots on the EUCIB map.

Error checking confirmed the robustness of the genetic maps constructed. Most importantly, the total number of locally misordered markers identified by aberrant triple recombinants (see Results) was small—only 160 in the total dataset. To empirically estimate the genome-wide genotyping error rate remaining in the EUCIB dataset, we reanalyzed for chromosomes 1 and 5 all double recombinant 1(.)2(.)1 and 2(.)1(.)2 haplotypes for adjacent markers from within the dataset (see Results and Table 5). Approximately 50% of the original aberrant haplotypes were found to be incorrect by repeating the aberrant genotypings. Extrapolation to the whole genome gives an error rate for genotyping of ~0.005. Given the low numbers of aberrant triple recombinants and the good agreement between the EUCIB and Whitehead/MIT maps, it would appear that a genotyping error rate of ~0.005 or lower will enable the reliable construction of high-resolution ordered maps from mouse backcrosses.

We were unable to eliminate a significant number of the aberrant double recombinant haplotypes for adjacent markers. As indicated above, the low numbers of aberrant triple recombinants and the general agreement of the EUCIB and Whitehead/MIT maps, gives us confidence that they do not all arise from local misordering. In addition, residual heterozygosity in the *M. spretus* parent population does not seem to be a major factor in the appearance of aberrant double recombinants (see Results). However, retesting of aberrant genotypes was carried out without altering the basic PCR reaction conditions used. Under these circumstances, it remains to be determined by more extensive examination of the supposed remaining double recombinants whether they have failed to be eliminated because of technical reasons, or truly represent aberrant nonmendelian events.

The EUCIB resource will be used in two ways. Firstly, the provision of definitive order for many microsatellite markers at high resolution across the genome will assist in the verification of physical maps as well as aid in orientating contigs and contig closure. Secondly, EUCIB can be used to provide further resolution and robustness to the construction of genetic and physical maps in any chromosome region. It is important to recognize that the total number of recombinants in the EUCIB resource—17, 029—far exceeds the minimum of 2302 recombinants required to separate and order the 2302 bins that form the EUCIB high-resolution map. Thus, the resolution of the EUCIB resource is not exhausted and clearly further recombinants can be exploited in any and every chromosome region to increase the resolution of genetic maps and to further enhance the construction of physical maps.

For a number of species (pig, cow, rat, and zebrafish) a number of genetic approaches with a variety of marker types are currently underway to develop comprehensive genome-wide microsatellite maps at intermediate resolution (Barendse et al. 1994; Postlethwait et al. 1994; Archibald et al. 1995; Jacob et al. 1995; Knapik et al. 1996). Local high-resolution genetic maps of STSs that have been constructed, either in the mouse or in other species, have proved enormously helpful in the construction of regional but robust physical maps. As for the mouse, there will be considerable value in ultimately developing genome-wide high-resolution maps in these other species.

METHODS

The EUCIB Backcross and *MBx* Database

The European Collaborative Interspecific Backcross (EUCIB) resource comprizes 982 DNAs derived from a C57BL/6–*M. spretus* backcross and has already been described in detail (Breen et al. 1994). The *MBx* database, URL http://www.hgmp.mrc.ac.uk/MBx/MBxHomepage.html, which holds all mouse, marker, and genotyping data and computes and displays high resolution maps has also been described (Breen et al. 1994). *MBx* can abstract and list all mice carrying single, double, and triple recombinant chromosomes (and chromosomes carrying larger numbers of apparent recombination events) across any chromosome region. In addition, *MBx* is able to provide a summary of haplotypes and their frequencies for any chromosome region. For access to the primary database contact: support@hgmp.mrc.ac.uk.

Details of map functions for the construction of genetic maps (see text) can be found at the *MBx* web site. Briefly, anchors and microsatellites are ordered on any chromosome using an algorithm that minimizes recombinants. This order is strictly maintained in constructing the genetic maps. The centromeric primary anchor on each chromosome is assigned a genetic map position according to consensus map data. In-

teranchor genetic distances for the primary anchors are then calculated for each chromosome and the remaining primary anchors assigned to the genetic map. Secondary anchors and microsatellites are subsequently incorporated into the genetic map maintaining genetic order as derived from the haplotype analysis. For each primary anchor interval, the total cumulative number of recombinants separating anchor and microsatellite markers in that interval is derived. The genetic distance separating a marker from any other marker or anchor in each primary anchor interval can then be calculated from the primary anchor genetic distance on the basis of the following ratio:

No. of recombinants separating marker from adjacent marker or anchor/Total cumulative no. of recombinants separating anchors and markers in the interval

Microsatellites mapping beyond the most proximal or distal primary anchors according to the haplotype analysis are added to the genetic map separately. However, it is likely that only a fraction of the relevant recombinants separating these markers and the primary anchor have been tested and genetic distances determined are, therefore, inaccurate.

Genotyping

Recombinant panels in any chromosome region were genotyped by use of a high-throughput, semiautomated fluorescent dUTP genotyping approach that has been described recently (Rhodes et al. 1997).

ACKNOWLEDGMENTS

This work was supported by the Medical Research Council, UK and partly by the European Commission (grant no. GENE-CT-93-0046). We thank the EUCIB steering committee for their advice and support and comments on earlier drafts of this paper. We also thank David Papworth for statistical advice.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Andersson, L., A. Archibald, M. Ashburner, S. Audun, W. Barendse, J. Bitgood, C. Bottema, T. Broad, and S.D.M. Brown. 1996. Comparative genome organization of vertebrates (First International Workshop). *Mamm. Genome* 7: 717–734.

Archibald, A.L., C.S. Haley, J.F. Brown, S. Couperwhite, H.A. McQueen, D. Nicholson, W. Coppieters, A. Van der Weghe, A. Stratil, A.-K. Wintero et al. 1995. The PiGMaP consortium linkage map of the pig (*Sus scrofa*) *Mamm. Genome* 6: 157–175.

Avner, P., L. Amar, L. Dandalo, and J.L. Guenet. 1988. Genetic analysis of the mouse using interspecific crosses. *Trends Genet.* 4: 18–23.

Barendse, W., S.M. Armitage, L.M. Kossarek, A. Shalom, B.W. Kirkpatrick, A.M. Ryan, D. Clayton, L. Li, H.L. Neibergs, N. Zhang et al. 1994. A genetic linkage map of the bovine genome. *Nature Genet.* 6: 227–235.

Breen, M., L. Deakin, B. Macdonald, S. Miller, R. Sibson, E. Tarttelin, P. Avner, F. Bourgade, J.-L. Guenet, X. Montagutelli et al. 1994. Towards high resolution maps of the mouse and human genomes—a facility for ordering markers to 0.1 cM resolution. *Hum. Mol. Genet.* 3: 621–627.

Brown, S.D.M. 1994. Integrating maps of the mouse genome. *Curr. Opin. Genet. Dev.* 4: 389–394.

———. 1996. Mouse genome. In *Encyclopedia of molecular biology and molecular medicine* (ed. R.A. Meyers), Vol. 4, pp. 120–128. VCH Publishers, New York, NY.

Brown, S.D.M and J. Peters. 1996. Combining mutagenesis and genomics in the mouse—closing the phenotype gap. *Trends Genet.* 12: 433–435.

Ceci, J.D., L.D. Siracussa, 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.

Copeland, N. and N.A. Jenkins. 1991. Development and applications of a molecular genetic linkage map of the mouse genome. *Trends Genet.* 7: 113–118.

Copeland, N.G., N.A. Jenkins, D.J. Gilbert, J.T. Eppig, L.J. Maltais, J.C. Miller, W.F. Dietrich, A. Weaver, S.E. Lincoln, R.G. Steen et al. 1993. A genetic linkage map of the mouse: Current applications and future prospects. *Science* 262: 57–66.

Dietrich, W.F., H. Katz, S.E. Lincoln, H.S. Shin, J. Friedman, N.C. Dracopoli, and E.S. Lander. 1992. A genetic map of the mouse suitable for typing intraspecific crosses. *Genetics* 131: 423–447.

Dietrich, W.F., N.G. Copeland, D.J. Gilbert, J.C. Miller, N.A. Jenkins, and E.S. Lander. 1995. Mapping the mouse genome: Current status and future prospects. *Proc. Natl. Acad. Sci.* 92: 10849–10853.

Dietrich, W.F. et al. 1996. A comprehensive genetic map of the mouse genome. *Nature* 380: 149–152.

Hudson, T.J., L.D. Stein, S.S. Gerety, J. Ma, A.B. Castle, J. Silva, D.K. Slonim, R. Baptista, L. Kruglyak, S.-H. Xu et al. 1996. An STS-based map of the human genome. *Science* 270: 1945–1954.

Jacob, H.J., D.M. Brown, R.K. Bunker, M.J. Daly, V.J. Dzau, A. Goodman, G. Koike, V. Kren, T. Kurtz, A. Lernmark et al. 1995. Genetic linkage map of the laboratory rat, *Rattus norvegicus*. *Nature Genet.* 9: 63–69.

Knapik, E.W., A. Goodman, O. Scott Atkinson, C.T. Roberts, M. Shiozawa, C.U. Sim, S. Weksler-Zangen, M.R. Trolliet, C. Futrell, B.A. Innes et al. 1996. A reference cross DNA panel for zebrafish (*Danio rerio*) anchored with simple sequence length polymorphisms. *Development* 123: 451–460.

Nadeau, J.H., B. Herrman, M. Bucan, D. Burkart, J.L. Crosby, M.A. Erhart, M. Kosowsky, J.P. Kraus, F. Michiels, A. Schnattinger et al. 1991. Genetic maps of mouse chromosome 17 including 12 new anonymous DNA loci and 25 anchor loci. *Genomics* 9: 78–89.

Postlethwait, J.H., S.L. Johnson, C.N. Midson, W.S. Talbot, M. Gates, E.W. Ballinger, D. Africa, R. Andrews, T. Carl, J.S. Eisen et al. 1994. A genetic linkage map for the zebrafish. *Science* 264: 699–703.

Reeves, R.H., M.R. Crowley, W.S. Moseley, and M.F. Seldin. 1991. Comparison of interspecific to intersubspecific backcrosses demonstrates species and sex differences in recombination frequency on mouse Chromosome 16. *Mamm. Genome* 1: 158–164.

Reeves, R.H., E.E. Rue, M.P. Citron, and D.E. Cabin. 1997. High-resolution recombinational map of mouse chromosome 16. *Genomics* 43: 202–208.

Rhodes, M., A. Dearlove, R. Straw, S. Fernando, A. Evans, M. Greener, T. Lacey, M. Kelly, K. Gibson, S.D.M. Brown, and C. Mundy. 1996. High throughput microsatellite analysis using fluorescent dUTPs for high resolution genetic mapping of the mouse genome. *Genome Res.* 7: 81–86.

Saunders, A.M. and M.F. Seldin. 1990. A molecular genetic linkage map of mouse chromosome 7. *Genomics* 8: 525–535.

Schuler, G.D., M.S. Boguski, E.A. Stewart, L.D. Stein, G. Gyapay, K. Rice, R.E. White, R. Rodriguez-Tome, A. Aggarwal, E. Bajorek et al. 1996. A gene map of the human genome. *Science* 274: 540–546.

Vignal, A., G. Gyapy, J. Hazan, S. Nguyen, C. Dupraz, N. Cheron, N. Becuwe, M. Tranchant, and J. Weissenbach. 1993. Non radioactive Multiplex Procedure for Genotyping of Microsatellite Markers. In *Methods in molecular genetics* (ed. Kenneth W. Adolph), Vol. 1, pp. 211–221. Academic Press, San Diefo, CA.

Received August 4, 1997; accepted in revised form February 9, 1998.