A genetic map of the mouse with 4,006 simple sequence length polymorphisms

William F. Dietrich, Joyce C. Miller, Robert G. Steen, Mark Merchant, Deborah Damron, Robert Nahf, Alec Gross, Diane C. Joyce, Michael Wessel, Robert D. Dredge, Andre Marquis, Lincoln D. Stein, Nathan Goodman, David C. Page & Eric S. Lander

We have constructed a genetic map of the mouse genome containing 4,006 simple sequence length polymorphims (SSLPs). The map provides an average spacing of 0.35 centiMorgans (cM) between markers, corresponding to about 750 kb. Approximately 90% of the genome lies within 1.1 cM of a marker and 99% lies within 2.2 cM. The markers have an average polymorphism rate of 50% in crosses between laboratory strains. The markers are distributed in a relatively uniform fashion across the genome, although some deviations from randomness can be detected. In particular, there is a significant underrepresentation of markers on the X chromosome. This map represents the two-thirds point toward our goal of developing a mouse genetic map containing 6,000 SSLPs.

Whitehead Institute/MIT Center for Genome Research, Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA

Dense linkage maps are an invaluable tool for genetic and genomic analysis. They facilitate high resolution genetic mapping and positional cloning of monogenic traits, allow genetic dissection of polygenic traits, permit fine-structure linkage disequilibrium studies, assist in evolutionary comparisons and provide an ordered scaffold on which complete physical maps of genomes can be assembled. The power of genetic maps increases with their density. For key organisms such as the human and mouse, extremely dense genetic maps are essential.

The discovery of simple sequence length polymorphisms (SSLPs) or microsatellites has greatly accelerated genetic map construction[1–4]. SSLPs occur at high frequency throughout mammalian genomes, tend to be highly polymorphic, are easily assayed by the polymerase chain reaction (PCR), and can be disseminated simply by publishing the sequence of their PCR primers. Several projects are underway to build dense SSLP maps of the human genome, notably by Weissenbach and his colleagues at Généthon[3,5].

For the past three years, the Whitehead Institute/MIT Center for Genome Research (CGR) has been developing an SSLP map of the mouse[4,6–9]. In 1992, we reported an initial map consisting of 317 markers[4]. In October 1993, we reported a 1518 marker map, integrated into the mouse gene map[8,9]. Here, we report the construction of a genetic map of the mouse genome containing 4,006 SSLPs. Markers are distributed at an average spacing of 0.35 cM, corresponding to about 750 kb. The map represents the two-thirds point toward our goal of developing a mouse genetic map containing 6,000 SSLPs.

Construction of genetic map

The mouse genetic map was constructed essentially as before[4]. Briefly, random clones containing the simple sequence repeat $(CA)_n$ were identified by oligonucleotide hybridization from total mouse genomic libraries with size-selected inserts and their DNA sequence was determined by single-pass automated sequencing. Also, mouse DNA sequences containing a variety of simple sequence repeats were identified in known gene sequences, primarily from GenBank. A PCR assay encompassing each simple sequence repeat was designed based on computer analysis. The PCR assays were tested for polymorphism on 12 inbred mouse strains. Those that defined different alleles in the OB and CAST strains — about 92% of the total — were genotyped in 46 progeny from an (OB × CAST) F_2 intercross. Because the cross involves 92 meioses, there is a crossover every 1.1 cM on average and markers can be ordered to this resolution (see Methodology).

The genetic map contains 4,006 SSLPs, of which 3,783 were derived from anonymous clones and 223 were taken from known gene sequences reported in GenBank and elsewhere (Fig. 1; see end of paper). The map densely covers all 20 mouse chromosomes and has a total genetic length of 1409 cM (Table 1). Because the markers were genotyped in an F_2 intercross, the map represents sex-averaged genetic distance. The observed genetic distances agree reasonably well with previous consensus estimates based on data from various mouse crosses.

A full description of the markers — including primer sequences, complete locus sequence, allele sizes in characterized inbred strains and genotypes in the cross — would require more than 250 journal pages and is thus

Table 1 Genetic markers and genetic length by chromosome

Chromosome	No. of markers	No. of random markers	No. from GENBANK	'Consensus' genetic length[a]	Observed genetic length[b]
1	310	293	17	98	114.6
2	307	290	17	107	95.8
3	223	212	11	100	67.6
4	228	220	8	81	74.7
5	257	245	12	93	85.4
6	221	205	16	74	63.4
7	230	215	15	89	69.3
8	208	202	6	81	75.0
9	205	188	17	70	71.1
10	184	177	7	78	73.4
11	215	191	24	78	84.5
12	186	177	9	68	61.7
13	207	200	7	72	65.3
14	183	170	13	53	66.0
15	172	165	7	62	62.7
16	134	133	1	59	54.7
17	165	148	17	53	51.1
18	153	148	5	57	44.5
19	84	81	3	42	57.0
X	134	123	11	88	71.5
Total	4006	3783	223	1503	1409.2

[a]Based on 'consensus' genetic map in *Encyclopedia of the Mouse Genome* (1993).
[b]Distance between most proximal and most distal markers in the map reported here.

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

		Based on cytogenetic length[a]		
Chromosome	No. of random markers[b]	% of total length	Expected no. of markers[c]	Z-score[d]
Autosomes only				
1	293	7.7%	281.2 ± 16.1	0.73
2	290	7.4%	271.5 ± 15.9	1.17
3	212	6.4%	234.0 ± 14.8	–1.48
4	220	6.3%	230.1 ± 14.7	–0.69
5	245	6.1%	221.9 ± 14.4	1.60
6	205	5.9%	216.0 ± 14.3	–0.77
7	215	5.5%	202.7 ± 13.8	0.89
8	202	5.3%	194.1 ± 13.6	0.58
9	188	5.1%	187.1 ± 13.3	0.07
10	177	5.1%	185.1 ± 13.3	–0.61
11	191	5.0%	184.4 ± 13.2	0.50
12	177	5.2%	190.6 ± 13.4	–1.01
13	200	4.7%	171.1 ± 12.8	2.26
14	170	4.8%	174.2 ± 12.9	–0.33
15	165	4.3%	158.2 ± 12.3	0.55
16	133	4.1%	148.8 ± 11.9	–1.32
17	148	4.1%	150.8 ± 12.0	–0.23
18	148	4.1%	151.6 ± 12.1	–0.30
19	81	2.9%	106.6 ± 10.2	–2.52
Total	3660	100.0%	3660.0	
Autosomes versus X				
Autosomes	3660	93.7%	3556.9 ± 14.9	6.90
X	123	6.3%	226.1 ± 14.9	–6.90
Total	3783	100.0%	3783.0	

[a]Cytogenetic length taken from ref. 10.
[b]Only random markers are considered to avoid biases in chromosomal distribution of known genes.
[c]Mean ± standard deviation. For comparison of autosomes to X chromosome, the expectation reflects the fact that 10% of the random markers were derived from male DNA (thus under-representingthe X chromosome by a factor of two) and 90% from female DNA.
[d]Z-score = (observed–expected)/standard deviation. For the autosomes, none of the Z-scores are significant at the p=0.05 level after Bonferroni correction for multiple testing. For the comparison of autosomes to X chromosome, the Z-score is significant at $p < 0.0001$.

omitted. Instead, this information can be obtained for any subset of markers by sending queries to an automatic electronic mail server maintained by CGR. To obtain a query form, send electronic mail consisting of the single word "help" to genome_database@genome.wi.mit.edu. Over the Internet, queries are typically answered in under two minutes.

Our SSLP map has recently been integrated with the mouse gene map[8,9]. Over 250 SSLPs were genotyped in a B6×(B6×SPRET) backcross in which Copeland, Jenkins and colleagues have genetically mapped RFLPs for about 1,000 genes.

Distribution of genetic markers

A key issue in evaluating a map is the distribution of markers throughout the genome. Analysis of our initial 317 marker mouse map suggested that SSLPs were distributed in a relatively uniform fashion. With more than 4,000 markers on the current map, it is possible to reinvestigate this question with greater precision.

There are a variety of ways to study whether the distribution of markers is uniform. One approach is to ask whether the observed number of markers on each chromosome agrees with expectation assuming that markers are uniformly distributed with respect to cytogenetic length[10]. For the autosomes, the chromosomal distribution of the random markers agrees remarkably well with expectation (Table 2). There are no statistically significant deviations (after accounting for multiple hypothesis testing).

In contrast, the X chromosome shows a clear deficit of random markers (Table 2). In examining the proportion of markers from the X chromosome, a small correction is required inasmuch as the first 10% of the random markers were isolated from male DNA[4] while the remaining 90% were isolated from female DNA. After adjusting for this slight systematic underrepresentation, the X chromosome contains only 52% as many markers as expected under the assumption of a uniform distribution across cytogenetic length. Possible explanations for such a striking deficit include: (i) a lower density of $(CA)_n$ repeat sequences on the X chromosome or (ii) a lower rate of polymorphism among $(CA)_n$ repeats present on the X chromosome. (The trivial explanation that our "female" DNA source was actually from a male was excluded on two grounds. We confirmed the sex of the DNA sources by using PCR assays for the mouse *Zfx* and *Zfy* loci on the X and Y chromosomes, respectively[11]. We also noted that the proportion of markers on the X chromosome was twofold higher among those markers isolated from the female DNA than the known male DNA.)

Another way to study the distribution of genetic markers is to examine the occurrence of clusters of crossovers and clusters of markers in the map. In our data, the position of every marker relative to every crossover can be identified. By ordering all crossovers (occurring in any of the meioses studied) and all markers relative to one another, the map of each chromosome can be reduced by a long string of the form "mmmmccmmmcc. . .", where each m denotes the occurrence of a marker and each c denotes the occurrence of a crossover (in one of the meioses studied). The string above, for example, indicates a succession of a block of four markers which showed no recombination in the meioses studied, an interval of two crossovers, a block of three markers which showed no recombination, and so

Table 3 Distribution of number of crossovers between consecutive random markers[a]

Crossovers per interval	Observed No.	Observed Percentage	Expected[b] No.	Expected[b] Percentage	P (longest run ≥ n)[b]
0	2866	76.1%	2800.6 ± 26.8	74.4%	100%
1	641	17.0%	717.4 ± 24.1	19.1%	100%
2	176	4.7%	183.7 ± 13.2	4.9%	100%
3	52	1.4%	47.1 ± 6.8	1.3%	100%
4	19	0.5%	12.1 ± 3.5	0.3%	98%
5	8	0.2%	3.1 ± 1.8	0.1%	65%
6	2	0.1%	0.8 ± 0.9	<0.1%	24%
7	0	0.0%	0.2 ± 0.5	<0.1%	7%
8	0	0.0%	0.1 ± 0.2	<0.1%	2%
9	1	<0.1%	0.0 ± 0.1	<0.1%	0.5%
Total	3765				

[a]Only random markers are considered, to avoid biases in distribution of known genes.
[b]See methodology concerning calculation.

on. Runs of many consecutive c's correspond to large genetic intervals, while runs of many consecutive m's correspond to large blocks of recombinationally unseparated markers.

If genetic markers are uniformly distributed with respect to crossovers, then the string should correspond to tossing a coin with probability π_m of being "m" and probability π_c ($= 1-\pi_m$) of being "c". Here, $\pi_m = M/(M+C)$ where M is the total number of markers and C is the total number of crossovers. The expected proportion of genetic intervals containing ≥ i consecutive crossovers is easily seen to be π_c^i (that is, the probability that an m is followed by at least i consecutive c's). The distribution of the length of the longest genetic interval (the longest run of c's) can also be calculated[12] (see Methodology). Similarly, the expected proportion of blocks containing ≥ i recombinationally unseparated markers is π_m^i and the distribution of the longest such block can be calculated. To avoid bias due to the distribution of gene sequences, the analysis was performed using only the randomly generated markers.

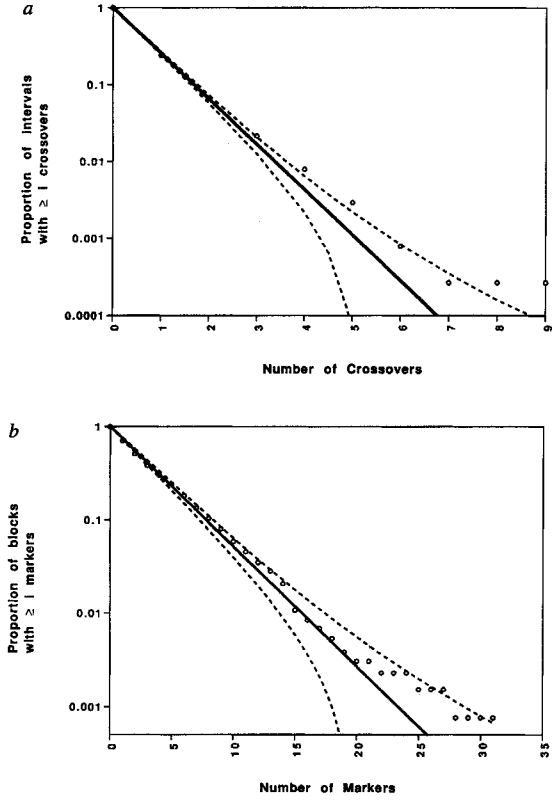


Fig. 2 *a*, Observed proportion of genetic intervals in the map having ≥ i crossovers compared to the expected proportion of π_c^i (where $\pi_c = 0.26$). *b*, Observed proportion of blocks in the map containing ≥ i recombinationally unseparated markers compared to the expected proportion of π_m^i (where $\pi_m = 0.74$). Data (*a* and *b*) are plotted on a logarithmic scale, for which the expected data fall on the solid lines. Observed data are plotted as points. Dotted lines indicate upper and lower confidence intervals corresponding to 2 standard deviations.

The observed distribution of interval lengths fits expectation reasonably well although there is a modest excess of larger intervals, suggesting some clustering of crossovers (Table 3 and Fig. 2*a*). The longest run of consecutive crossovers has expected length 5, with a 95% confidence interval of roughly 4–8. The single outlier is the distal interval on chromosome 19 between *D19Mit33* and *D19Mit76*, which had 9 crossovers or about 10 cM. The probability that such a large interval would occur by chance anywhere in the map is only 0.005. (Genotypes were carefully reconfirmed to ensure that the crossovers did not result from mistyping.) The data suggest the possibility of a recombinational hotspot near the telomere of chromosome 19, at least in (CAST × B6) F2 crosses. Interestingly, there does not appear to be enhanced recombination in a (SPRET × B6) × B6 backcross: Eicher and Shown[13] reported that the interval *D19Mit1–D19Mit33–D19Mit6* measured only 5 cM. This might be due to a difference between the strains or to enhanced recombination being present only in male meiosis (which contributes to recombination frequency in F2 intercrosses but not in *M. spretus* backcrosses in which only the female parent segregates for polymorphisms).

The distribution of the number of markers occurring between consecutive crossovers also shows some modest evidence of clustering (Table 4 and Fig. 2*b*). There are significantly more occurrences of consecutive crossovers without an intervening marker than expected by chance (387 versus 334.3 ± 15.8; Z-score = 3.4). These data are consistent with the presence of recombinational hotspots in some regions. Adjusting for this excess, the remainder of this distribution is not a bad fit to expectation. The largest block of recombinationally unseparated markers would be expected to contain about 23 markers, with a 95% confidence interval of 19–34. In fact, the largest observed block is 32 which falls within the expected range.

The map appears to provide convenient entry points for nearly the entire genome. Approximately 90% of the map lies within 1.1 cM of a marker and 98.8% lies within 2.2 cM. The coverage is only slightly less than the expectation for randomly spaced markers, which is 93% and 99.5%, respectively (based on expectations in Tables 3 and 4).

Polymorphism among mouse strains

SSLPs are particularly useful for mouse genetics because they are highly variable even among inbred laboratory mouse strains, making it possible to genotype virtually any cross of interest[4]. For the SSLPs in the map, allele sizes were determined in 12 inbred strains (ten laboratory strains, which are derived from *M. m. domesticus* and *M. m. musculus* progenitors; the different subspecies *M. m. castaneus*; and the different species *M. spretus*). The SSLPs

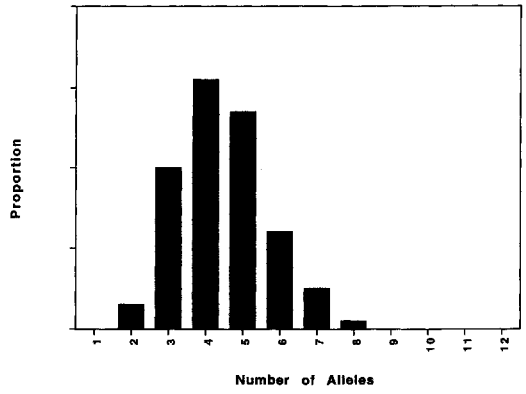


Fig. 3 Histogram showing number of distinct allele sizes among 12 strains characterized for SSLP markers reported here.

in the map are necessarily variant between OB and CAST, since this was a prerequisite for genetic mapping in the cross.

The average number of alleles per SSLP was 4.5 (Fig. 3). Conveniently, over 75% of the pairwise allele difference are ≥ 4 bp — making it possible to score the difference on high percentage agarose gels. The polymorphism rate between laboratory strains and the different species *M. spretus* or the different subspecies *M. m. castaneus* was about 94%, while the polymorphism rate among laboratory strains averaged about 50% (Table 5). In only five of 45 pairwise comparisons among the ten laboratory strains was the polymorphism rate below 44%: C3H-DBA (39%), C3H-BALB (37%), C3H-A (34%), BALB-A (32%), and OB–B6 (1%). (The last case is expected since OB and B6 are a congenic pair, with OB having been derived by repeated backcrossing to B6 with selection for the ob mutation.)

Interestingly, the polymorphism rate among laboratory strains was not constant across chromosomes (Table 6). The most extreme deviation was for the X chromosome, which showed a polymorphism rate of only 33% as compared to 50% for the autosomes. Also, chromosome 10 showed a significantly lower rate of polymorphism (36%). The rate of polymorphism between laboratory strains and *M. spretus* or *M. m. castaneus* did not vary significantly across chromosomes.

Table 4 Distribution of number of random markers occurring between consecutive crossovers[a]

Markers per block	Observed No.	Observed Percentage	Expected[b] No.	Expected[b] Percentage	P(longest run ≥ n)[b]
0	387	29.6%	334.8 ± 15.8	25.6%	100%
1	251	19.2%	249.0 ± 14.2	19.1%	100%
2	168	12.9%	185.2 ± 12.6	14.2%	100%
3	110	8.4%	137.8 ± 11.1	10.5%	100%
4	77	5.9%	102.5 ± 9.7	7.8%	100%
5	78	6.0%	76.2 ± 8.5	5.8%	100%
6	61	4.7%	56.7 ± 7.4	4.3%	100%
7	39	3.0%	42.2 ± 6.4	3.2%	100%
8	31	2.4%	31.4 ± 5.5	2.4%	100%
9	28	2.1%	23.3 ± 4.8	1.8%	100%
10	17	1.3%	17.4 ± 4.1	1.3%	100%
11	14	1.1%	12.9 ± 3.6	1.0%	100%
12	9	0.7%	9.6 ± 3.1	0.7%	100%
13	10	0.8%	7.1 ± 2.7	0.5%	100%
14	13	1.0%	5.3 ± 2.3	0.4%	100%
15	3	0.2%	4.0 ± 2.0	0.3%	100%
16	2	0.2%	2.9 ± 1.7	0.2%	100%
17	2	0.2%	2.2 ± 1.5	0.2%	100%
18	2	0.2%	1.6 ± 1.3	0.1%	99%
19	1	0.1%	1.2 ± 1.1	0.1%	97%
20	0	0.0%	0.9 ± 0.9	0.1%	93%
21	1	0.1%	0.7 ± 0.8	0.1%	85%
22	0	0.0%	0.5 ± 0.7	<0.1%	76%
23	0	0.0%	0.4 ± 0.6	<0.1%	66%
24	1	0.1%	0.3 ± 0.5	<0.1%	55%
25	0	0.0%	0.2 ± 0.5	<0.1%	45%
26	0	0.0%	0.2 ± 0.4	<0.1%	36%
27	1	0.1%	0.1 ± 0.3	<0.1%	28%
28	0	0.0%	0.1 ± 0.3	<0.1%	22%
29	0	0.0%	0.1 ± 0.3	<0.1%	17%
30	0	0.0%	0.0 ± 0.2	<0.1%	13%
31	0	0.0%	0.0 ± 0.2	<0.1%	10%
32	1	0.1%	0.0 ± 0.2	<0.1%	7%
Total	1307				

[a]Only random markers are considered to avoid biases in distribution of known genes.
[b]See methodology concerning calculation.

Discussion

The 4,006 marker genetic map of the mouse constructed here constitutes the densest SSLP map constructed in any organism so far. The total genetic length has not grown significantly with the addition of the last 2,500 markers, suggesting that the map covers essentially the entire mouse genome. Interestingly, the genetic length of 1409 cM measured in our (CAST × OB) F_2 intercross is significantly larger than the length of 1224 cM in a (SPRET × B6) backcross. (For this comparison, the genetic length of the SPRET × B6 backcross was recalculated using the Kosambi map function. The corresponding lengths are 1,436 and 1,344 cM with Haldane's map function[8].) The discrepancy is more striking than it may appear, as the F_2 intercross reflects sex-averaged genetic distance while the backcross measures female genetic distance, which is generally thought to be substantially larger than in males. The difference may reflect crossover-suppression caused by local inversions between laboratory mouse and the evolutionarily more distant *M. spretus*[8].

By a number of tests, the markers appear to be relatively uniformly distributed across the genome, although some modest evidence of clustering is present. There is only one suprisingly large gap, a 10 cM interval at the distal end of chromosome 19. Increased recombination in subtelomeric regions has been suggested for some human chromosomes[14]. More generally, there is a slight overall excess clustering of crossovers, which could reflect non-uniformity in the distribution of recombination or $(CA)_n$ repeats with respect to physical distance. Recombinational hotspots and coldspots are certainly known to exist in many organisms including the mouse[15], but the relative uniformity of marker distribution indicates that their effect is not dramatic on maps of this density and resolution. Studies involving much denser maps may reveal greater clustering of recombination at a finer level, while studies involving many more meioses might reveal greater clustering of markers.

An unexpected observation is the nearly twofold underrepresentation of markers on the X chromosome. The deficit could be due either to a deficit of $(CA)_n$ repeats on the X chromosome or a lower polymorphism rate among those $(CA)_n$ repeats on the X chromosome. In principle, these alternatives could be distinguished by determining the chromosomal distribution of the $(CA)_n$ repeats that were not polymorphic between OB and CAST. It is also striking that the rate of polymorphism among

Table 5 Rate of polymorphism for SSLP markers among 12 mouse strains[a,b]

	OB	B6	DBA	A	C3H	BALB	AKR	NON	NOD	LP	SPR	CAST
OB	–											
B6	1%	–										
DBA	55%	52%	–									
A	54%	53%	48%	–								
C3H	55%	52%	39%	34%	–							
BALB	54%	51%	46%	32%	37%	–						
AKR	54%	52%	48%	46%	45%	44%	–					
NON	55%	53%	52%	49%	49%	48%	49%	–				
NOD	57%	54%	50%	51%	50%	50%	48%	45%	–			
LP	57%	55%	52%	53%	50%	49%	52%	52%	52%	–		
SPR	93%	92%	92%	92%	92%	92%	92%	92%	92%	92%	–	
CAST	100%	98%	94%	94%	95%	94%	95%	95%	95%	94%	94%	–

[a]Strains designations are: OB, C57BL/6J-ob/ob; B6, C57BL/6J; DBA, DBA/2J; A, A/J; C3H, C3H/HeJ; BALB, BALB/cJ; AKR, AKR/J; NON, NON/Lt; NOD, NOD/MrkTacBr; LP, LP/J; SPR, SPRET/Ei; CAST, CAST/Ei.
[b]Standard error of the mean is approximately 0.8% for rates near 50% and 0.3% for rates near 95%.

laboratory strains was significantly lower on the X chromosome than for other chromosomes, although there was no difference for the rate of polymorphism between laboratory strains and the more distant CAST or SPR. In the human, the X chromosome has been reported to have a threefold lower rate of RFLP polymorphism[16]. The effect has been attributed to the different genetic and population genetic forces acting on the X chromosome as compared to the autosomes. For example, the mutation rate is thought to be higher in the male germline than the female germline. As X chromosomes pass through the male germline only 2/3 as often as do autosomes, the mutation rate may be correspondingly lower. Also, the fact that X chromosomes function in the haploid state in males implies that selection acts differently, which may diminish polymorphism. Our data suggest that the presence of reduced polymorphism on the X chromosome may be general, at least in mammals.

Table 6 Polymorphism rate for SSLP markers by chromosome

Chromosome strains[a,b]	Among lab SPR or CAST[b]	Lab strains versus
1	55%	94%
2	49%	94%
3	51%	95%
4	53%	93%
5	48%	95%
6	47%	94%
7	48%	93%
8	45%	94%
9	53%	94%
10	36%	97%
11	55%	94%
12	50%	93%
13	48%	94%
14	49%	94%
15	51%	94%
16	46%	94%
17	59%	92%
18	52%	95%
19	50%	92%
X	33%	94%
Genome-wide	49%	94%

[a]Pairwise comparisons of OB, B6, DBA, A C3H, BALB, AKR, NON, NOD, and LP.
[b]Standard error of the mean for each chromosome depends on number of markers studied, but is <1% in all cases.

The 4,006-marker SSLP map should facilitate a wide range of biological studies. For initial genetic mapping studies, one can select about 100 markers spaced at 15 cM intervals. It may be convenient to use polymorphisms that are easily resolved on agarose gels (about 75% of the total). To map a polygenic or quantitative trait[17], one would genotype each progeny for each marker, a task that might take a few months or less. To map a monogenic trait, one can proceed even more rapidly by using 'phenotyping pooling'[18] in which one initially genotypes only two samples — containing pooled DNA from affected progeny and unaffected progeny, respectively. The two samples should show similar proportions of the two parental alleles at markers unlinked to the trait, but quite different proportions for linked markers. In this manner, one can initially localize a trait with only about 200 PCR reactions, a task that can be accomplished in a few days. Once initial linkage is detected, individual progeny should be genotyped using all markers in the region to identify the closest flanking markers. Since a typical gene should lie at an average distance of 375 kb from a marker and since yeast artificial chromosome YAC libraries with average insert size of 700 kb are available[19,20], chromosomal walking to the gene should be rapid. In addition to its application in positional cloning, the map should be valuable for evolutionary studies as well[21].

For the purpose of constructing a physical map of the mouse genome with overlapping YACs, an even denser genetic map would be desirable. With a map consisting of 6,000 SSLPs, the average spacing between markers would be 500 kb and the typical gene would be at an average distance of 250 kb, both distances being smaller than the average size of current YACs. Given the presence of 50,000–100,000 $(CA)_n$ repeats in the mouse genome and the availability of streamlined methods for genetic map construction, such a goal should be feasible.

Methodology

Construction of genetic map. Briefly, (i) sequences containing simple sequence repeats (almost all $(CA)_n$) were obtained, either through sequencing of genomic clones that hybridize to $(CA)_{15}$ or $(GT)_{15}$, or by searching sequence databases; (ii) PCR primers flanking the simple sequence repeat were selected; (iii) the PCR assays were used to characterize allele sizes in 12 mouse strains: C57BL/6J-ob/ob, C57BL/6J, DBA/2J, A/J, C3H/HeJ, BALB/cJ, AKR/J, NON/Lt, NOD/MrkTacBr, LP/J (all laboratory strains, derived from *M. m. domesticus* and *M. m. musculus*), SPRET/Ei (a strain of the species *M. spretus*), and CAST/Ei (a strain of the subspecies *M. m. castaneus*); (iv) for those assays detecting variation between OB and CAST, 46 progeny from an (OB × CAST)F2 intercross were genotyped; and (v) genetic maps were constructed by using the MAPMAKER computer package[22], incorporating a mathematical error-checking procedure[23]. These steps were performed essentially as described[4], with the following modifications: the OB and B6 strains are a congenic pair, with OB having been constructed by repeated backcrossing to B6 with selection for the *ob* mutation.

Genomic libraries. The short-insert total genomic libraries were constructed by using a variety of different procedures: complete single digestion with *Mbo*I, *Alu*I, *Hae*III, and complete triple digests using *Alu*I, *Hae*III and *Rsa*I. Digests were fractionated on 4% NuSieve GTG agarose and fragments between 200 and 500 bp were selected for ligation into M13mp19. (Some libraries were also prepared by ligation into the plasmid pcDNAII, but this vector was eventually abandoned in favour of M13mp19 due to the superior sequence quality obtained from the single stranded template.) All libraries were transformed into XL1-Blue cells (Stratagene). DNA was prepared from M13 clones by using a magnetic bead miniprep, essentially as previously described[24].

Length screen of clones. In some proportion of clones, the $(CA)_n$ repeat is too close or too far from the cloning site to allow PCR primers to be selected on both sides. To avoid sequencing such clones, a preliminary screening step was used to determine the size of the insert and the position of the repeat relative to the vector sequence. Miniprepped DNA was diluted 30-fold in distilled, deionized H_2O, and 1 µl of this dilution was used as the template in a 15 µl PCR reaction using AmpliTaq DNA polymerase (Perkin-Elmer Cetus) set up according to the manufacturer's specifications. Three PCR reactions were performed on each miniprep: (1) with primers flanking the M13 cloning site ("Forward": 5'–TGTAAAACGACGGCCAGT–3' and "Reverse": 5'–CAGGAAACAGCTATGACC–3'); (2) with the "Forward" primer and a primer complementary to a $(CA)_n$-repeat (5'–CCCGGATCC$(GT)_9$–3'); and (3) with the "Forward" primer and a primer complementary to a $(GT)_n$-repeat (5'–CCCGGATCC$(CA)_9$–3'). Reaction 1 is designed to measure the length of the insert, while Reaction 2 or 3 is designed to measure the distance from the Forward primer to the repeat. Reaction 1 and a pool of reactions 2 and 3 were electrophoresed on a 2% Metaphor agarose gel (FMC Bioproducts). Clones with insert size less than 700 bp and with the repeat within a range of 50–500 bp from the Forward primer were sequenced. Length screening was carried out in high throughput in 96-well microtitre plates.

Duplicate checking. To avoid mapping previously encountered simple sequence repeats, a computer program was used to compare newly determined DNA sequences to previously sequenced clones. The proportion of duplicates remained in the range of 5–10% throughout the project, owing to periodic substitution of new libraries constructed with different restriction enzymes.

Genotyping. To genotype F_2 progeny for SSR polymorphisms, PCR reactions were performed with one radioactively labeled primer and one unlabelled primer and the products were visualized upon autoradiography of polyacrylamide gels. Primers were end-labelled with [γ-^{32}P]ATP (RediVue, Amersham) according to standard protocols[25]. A 20 ng aliquot of genomic DNA was amplified in a 10 µl PCR reaction using AmpliTaq DNA polymerase (Perkin-Elmer Cetus) according to manufacturer's specifications. The primer concentrations were 75 nM end-labelled forward primer, and 75 nM unlabelled reverse primer. The reactions were overlaid with 40 µl of light mineral oil (Sigma). Reactions were amplified on a TC1600 thermal cycler (Intelligent Automation Systems, Cambridge, MA) using the following protocol: 30 cycles of 92 °C for 30 s, 55 °C for 30 s, and 72 °C for 30 s. Gels and autoradiography were as previously described[4].

Analysis of clusters of crossovers and markers. As noted in the text, the assumption that markers are randomly distributed with respect to genetic distance implies that the sequence of markers and crossovers occurring in the map should follow the expected behaviour of coin flips. The expected behaviour of head runs in coin flipping has been well studied[12]. If the probability of heads is p, the expected proportion of tails followed by at least i consecutive heads is p^i. If R_n denotes the longest run of consecutive heads when the coin is flipped n times, the expected value of R_n is $\mu = \log_{1/p}((n-1)(1-p)+1)$ and the distribution of R_n is given approximately by $\text{Prob}(R-\mu > t) \quad \exp(-p^t)$. These formulas are used in computing the expectations in Tables 3 and 4.

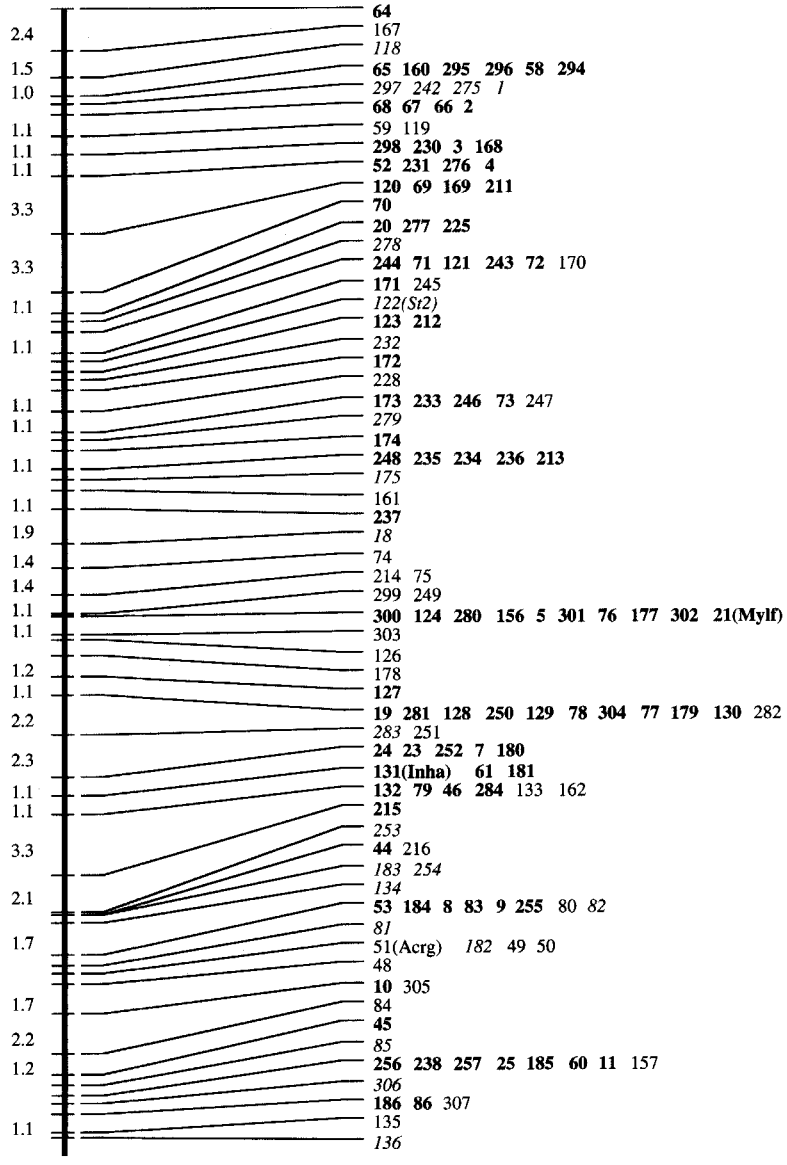
Acknowledgements

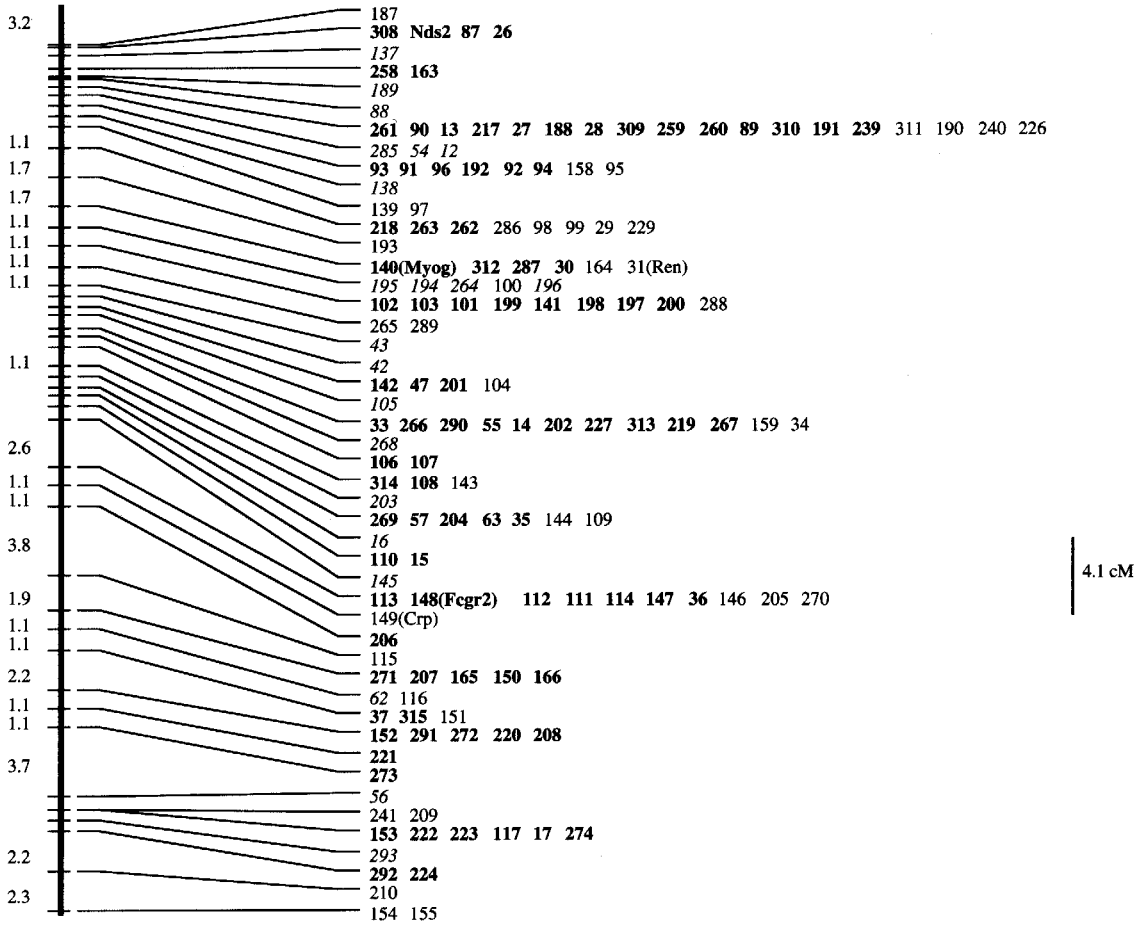
We thank laboratory aides Loden Wangchuk, Dawa Tsering and Gail Farino for excellent technical assistance; Robert Williams and Richelle Cutler (University of Tennessee), Lois Maltais (Jackson Laboratory) and Summa Dutta (MIT) for their help in ascertaining the official mouse gene names for SSLPs developed from gene sequences in GENBANK. This work was supported in part by grants from the National Center for Human Genome Research and the Markey Foundation (E.S.L.).

Received 25 January; accepted 22 March 1994.

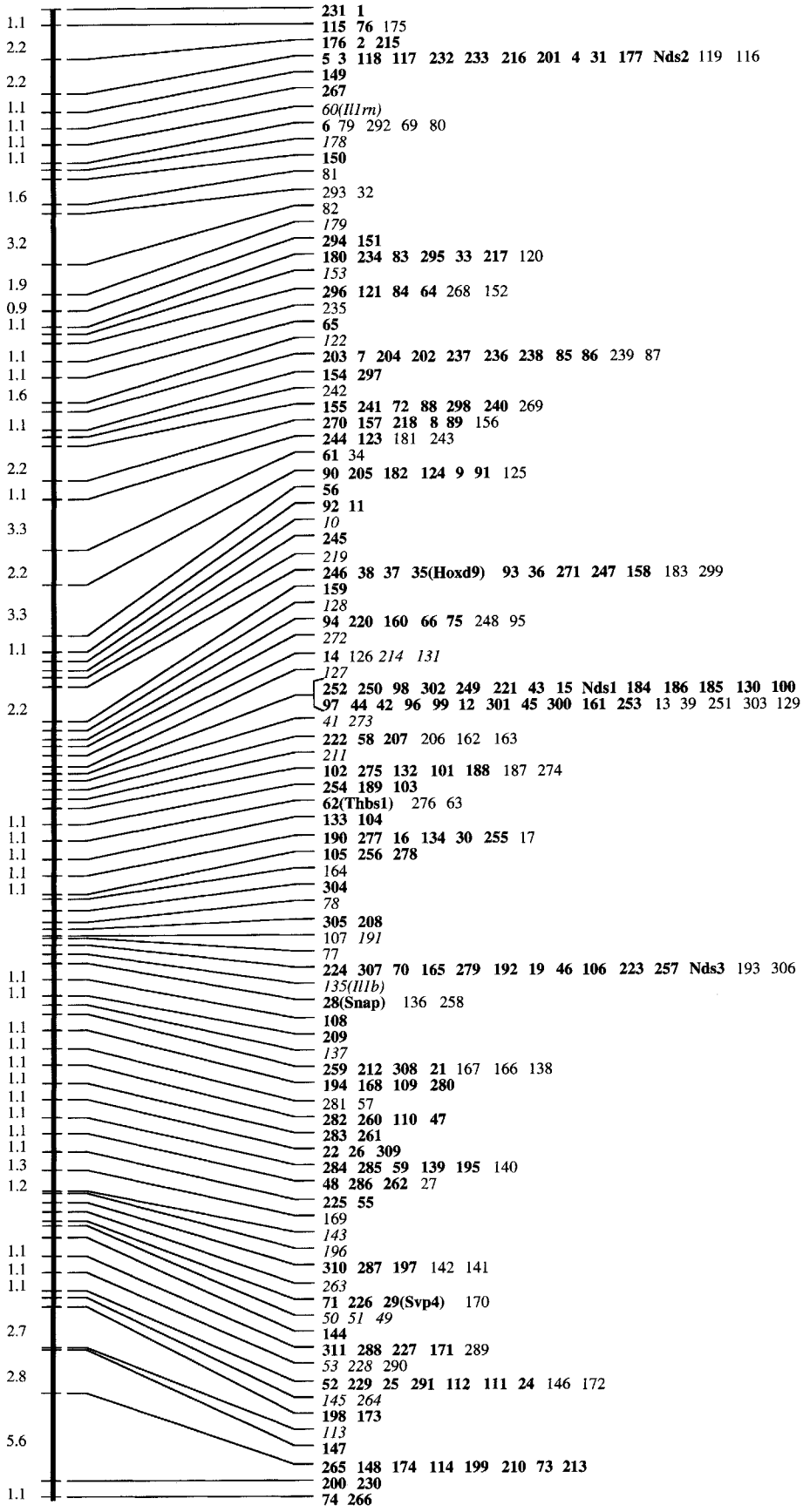
1. Weber, J.L. & May, P.E. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. hum. Genet.* **44**, 388–396 (1989).
2. Love, J.M., Knight, A.M., McAleer, M.A. & Todd, J.A. Towards construction of a high-resolution map of the mouse genome using PCR analysed microsatellites. *Nucl. Acids Res.* **18**, 4123–4130 (1990).
3. Weissenbach, J. *et al.* A second-generation linkage map of the human genome. *Nature* **359**, 794–801 (1992).
4. Dietrich, W.F. *et al.* A genetic map of the mouse suitable for typing intraspecific crosses. *Genetics* 131, 423–447 (1992).
5. Gypay, G. *et al.* The 1993–94 Généthon human genetic linkage map. *Nature Genet.* **7**, 246–339 (1994).
6. Dietrich, W.F. *et al.* in *Genetic Maps 1992* (ed. O'Brien, S.) 4.110–4.142 (Cold Spring Harbor Laboratory Press, New York, 1992).
7. Miller, J.C. *et al.* in *Genetic Variants and Strains of the Laboratory Mouse* 3rd edn (eds Lyons, M.F. & Searle, A.) (Oxford University Press, New York, in the press).
8. Copeland, N.G. *et al.* A genetic linkage map of the mouse: Current applications and future prospects. *Science* **262**, 57–66 (1993).
9. Copeland, N.G. *et al.* Genome Maps IV: The Mouse. *Science* **262**, 67–82 (1993).
10. Evans, E. in *Genetic Variants and Strains of the Laboratory Mouse* 2nd edn (eds Lyon, M.F. & Searle, A.) 576–578 (Oxford University Press, New York, 1989).
11. Mardon, G. *et al.* Mouse Zfx protein is similar to Zfy-2: each contains an acidic activating domain and 13 zinc fingers. *Molec. Cell Biol.* **10**, 681–688 (1990).
12. Arratia, R., Goldstein, L. & Gordon, L. Poisson Approximation and the Chen-Stein Method. *Stat. Sci.* **5**, 403–434 (1990).
13. Eicher, E.M. & Shown, E.P. Molecular markers that define the distal ends of mouse autosomes 4, 13, 19 and the sex chromosomes. *Mamm. Genome.* **4**, 226–229 (1993).
14. Tanzi, R.E. *et al.* A genetic linkage map of human chromosome 21: analysis of recombination as a function of sex and age. *Am. J. hum. Genet.* 50, 551–558 (1992).
15. Bryda, E.C., De Pari, J.A., Sant'Angelo, D.B, Murphy, D.B. & Passmore, H.C. Multiple sites of crossing over within the Eb recombinational hotspot in the mouse. *Mamm. Genome* **2**, 123–129 (1992).
16. Hofker, M.H. *et al.* The X chromosome shows less genetic variation at restriction sites than the autosomes. *Am. J. hum. Genet.* **3**, 438–451 (1986).
17. Lander, E.S. & Botstein, D. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**, 185–199 (1989).
18. Michelmore, R.W., Paran, I. & Kesseli, R.V. Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proc. natn. Acad. Sci. U.S.A.* **88**, 9828–32 (1991).
19. Larin, Z., Monaco, A.P. & Lehrach, H. Yeast artificial chromosome libraries containing large inserts from mouse and human DNA. *Proc. natn. Acad. Sci. U.S.A.* **88**, 4123 (1991).
20. Kusumi, K., Smith, J.S., Segre, J.A., Koos, D.S. & Lander, E.S. Construction of a large-insert yeast artificial chromosome (YAC) library of the mouse genome. *Mamm. Genome* **4**, 391–392 (1993).
21. Bruford, M.W. & Wayne, R.K. Microsatellites and their application to population genetic studies. *Curr. Op. Genet. Devel.* **3**, 939–944 (1993).
22. Lander, E.S. *et al.* MAPMAKER: An interactive computer package for constructing genetic linkage maps of experimental and natural populations. *Genomics* **1**, 174–181 (1987).
23. Lincoln, S.E. & Lander, E.S. Systematic detection of errors in genomic linkage data. *Genomics* **14**, 604–610 (1992).
24. Hawkins, T. M13 single-strand purification using a biotinylated probe and streptavidin coated magnetic beads. *DNA Seq.* **3**, 65–69 (1992).
25. Sambrook, J., Fritsch, E.F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* 2 edn (Cold Spring Harbor Laboratory, New York, 1989).

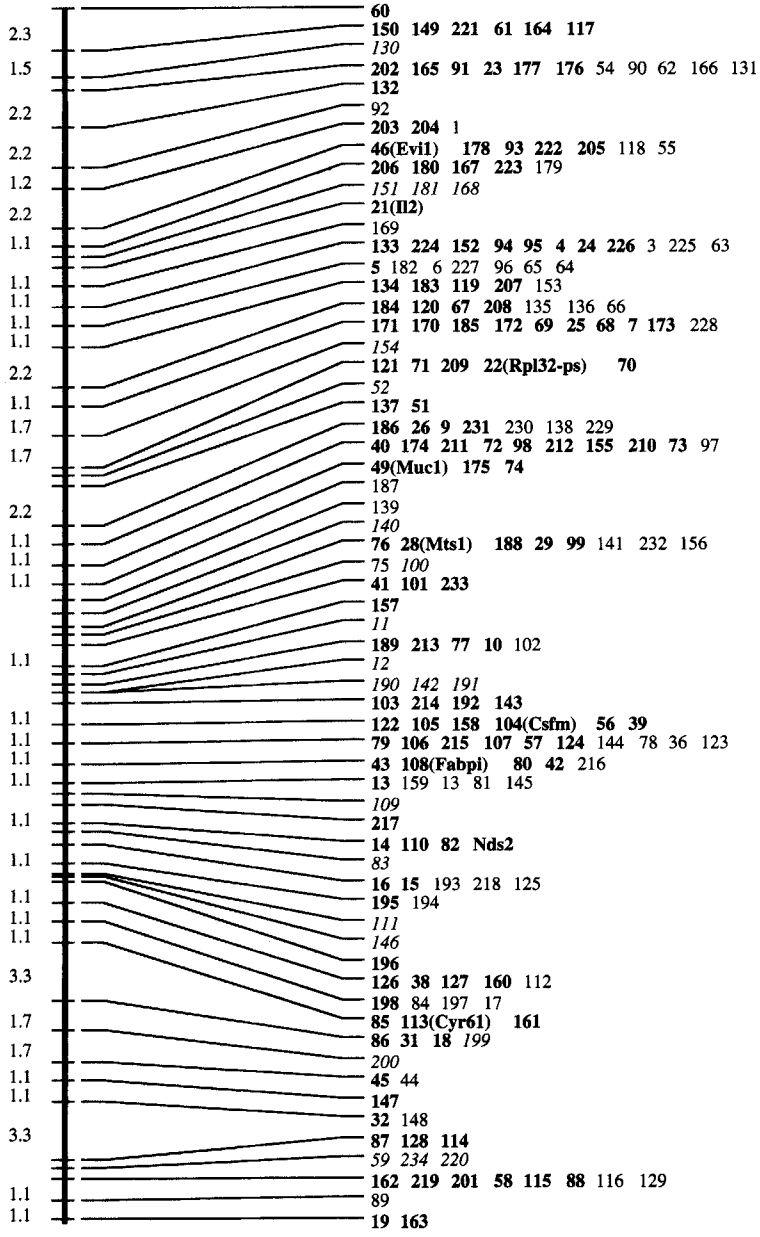
► Fig. 1 SSLP genetic map of the mouse. For the 99% of markers that were developed at the Whitehead Institute/MIT Center for Genome Research, formal locus names have been abbreviated. For example, the locus D7Mit3 is simply denoted by 3 on chromosome 7. For loci developed elsewhere, the laboratory designation is retained (for example, *D4Nds1* is denoted Nds1). For loci developed from genes for which a gene symbol has been assigned by the mouse nomenclature committee, the gene symbol is given in parentheses to aid in correlation with the mouse gene map. Linkage groups are represented by lines, with the centromere at the top. Each linkage group consists of markers that are linked to each other by a lod score of at least 5. The type-face of marker names indicates the statistical support for the genetic order shown. Markers whose order relative to the rest of the map is supported by a lod of 2.5 or higher are indicated in bold type; these are referred to as "framework" markers. Markers whose order is supported by a lod between 1.0 and 2.5 are indicated in plain face type. Markers in italics have a lod in support of order of 1.0 or less; these markers represent ambiguities with regard to the flanking markers only, but their placement lod score relative to the rest of the map is greater than 2.5. Ambiguities can arise either because genotypes were not obtained for a few progeny or because the marker is dominant rather than codominant (about 5% of the total). Distances in centiMorgans between markers are indicated to the left of the line and were calculated using Kosambi's map function. Where multiple markers did not recombine in the 92 meioses studied, they are listed together in a block. The scale is the same for all parts of Fig. 1; scale bar, 4.1 cM.



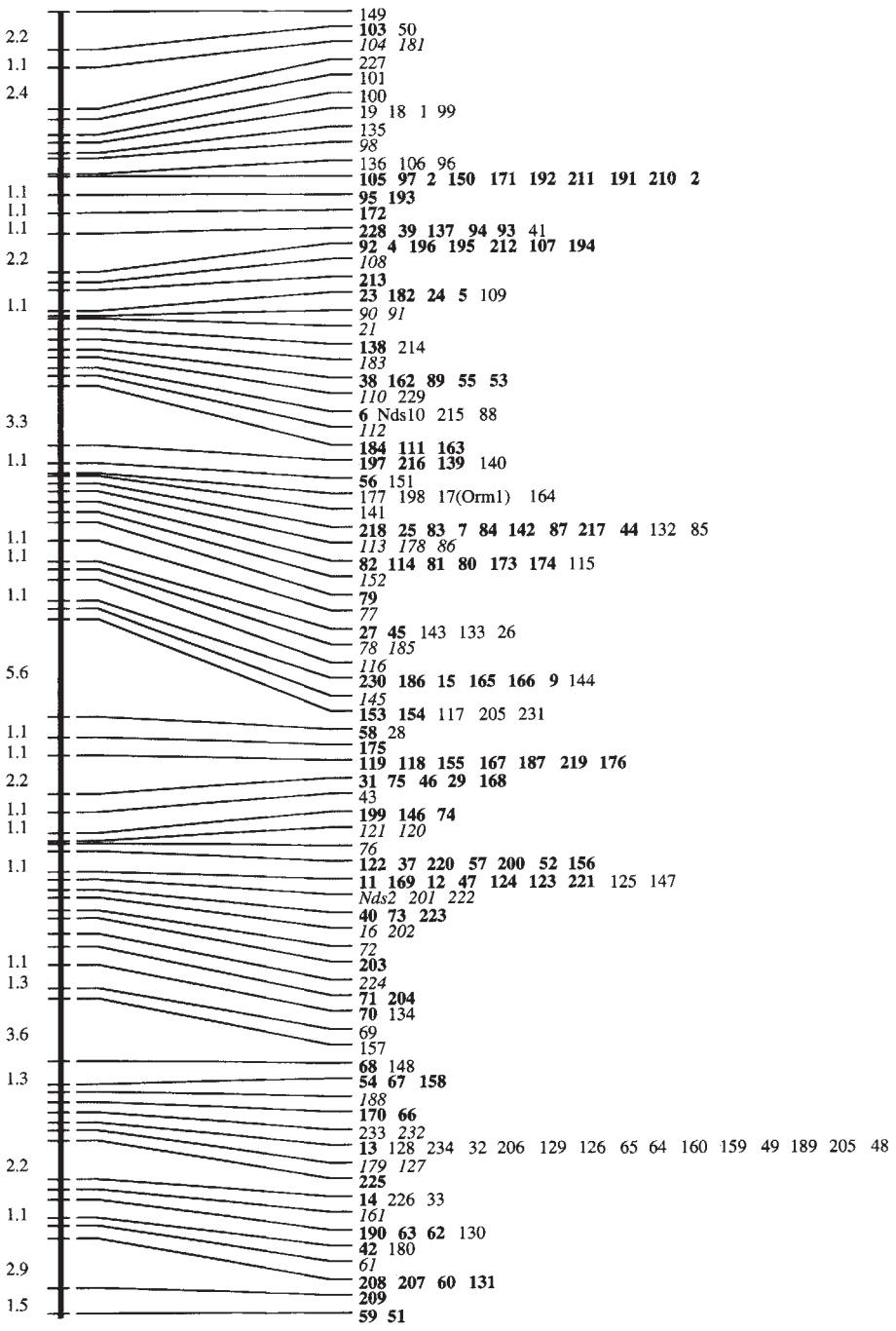


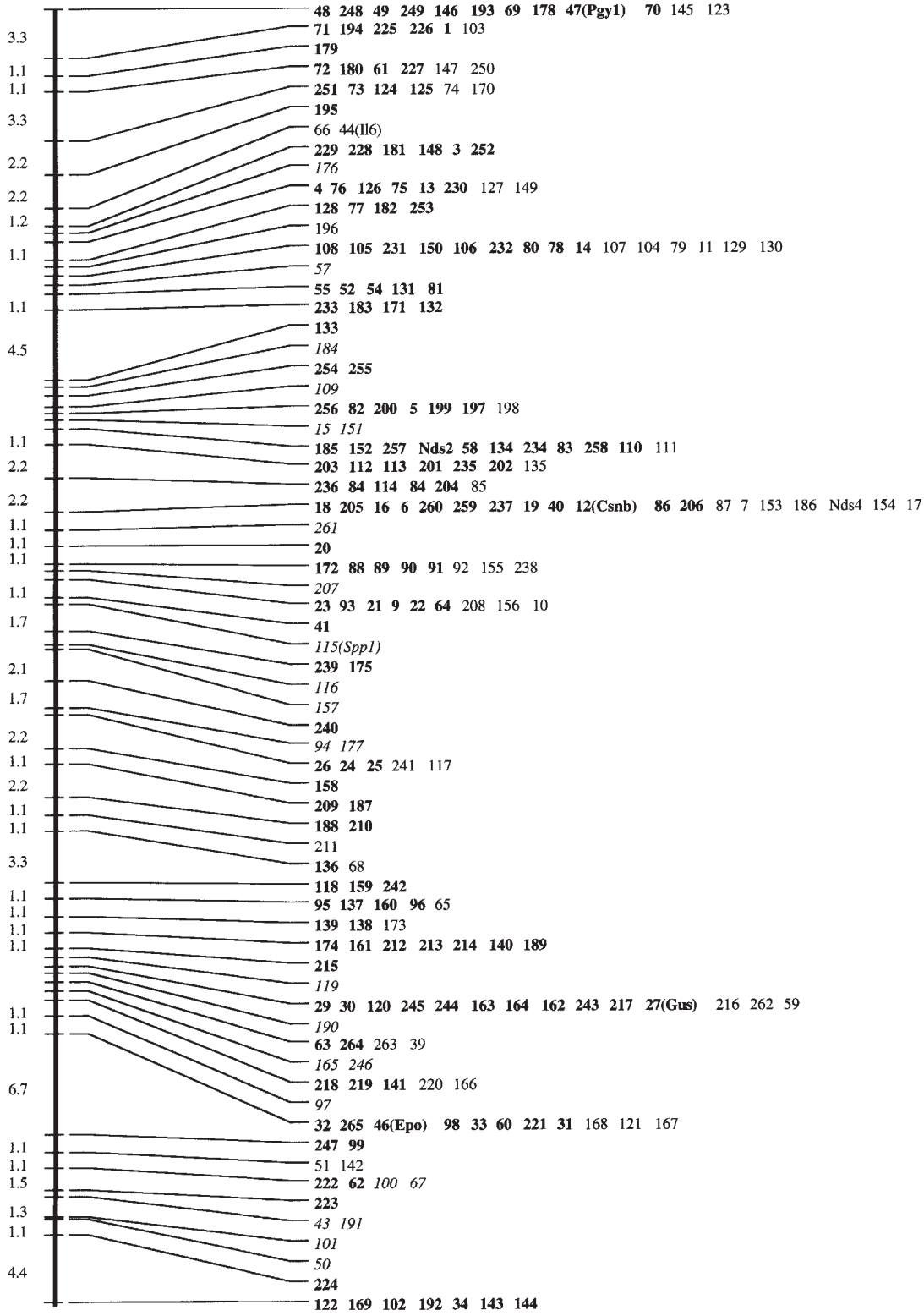
Chromosome 2



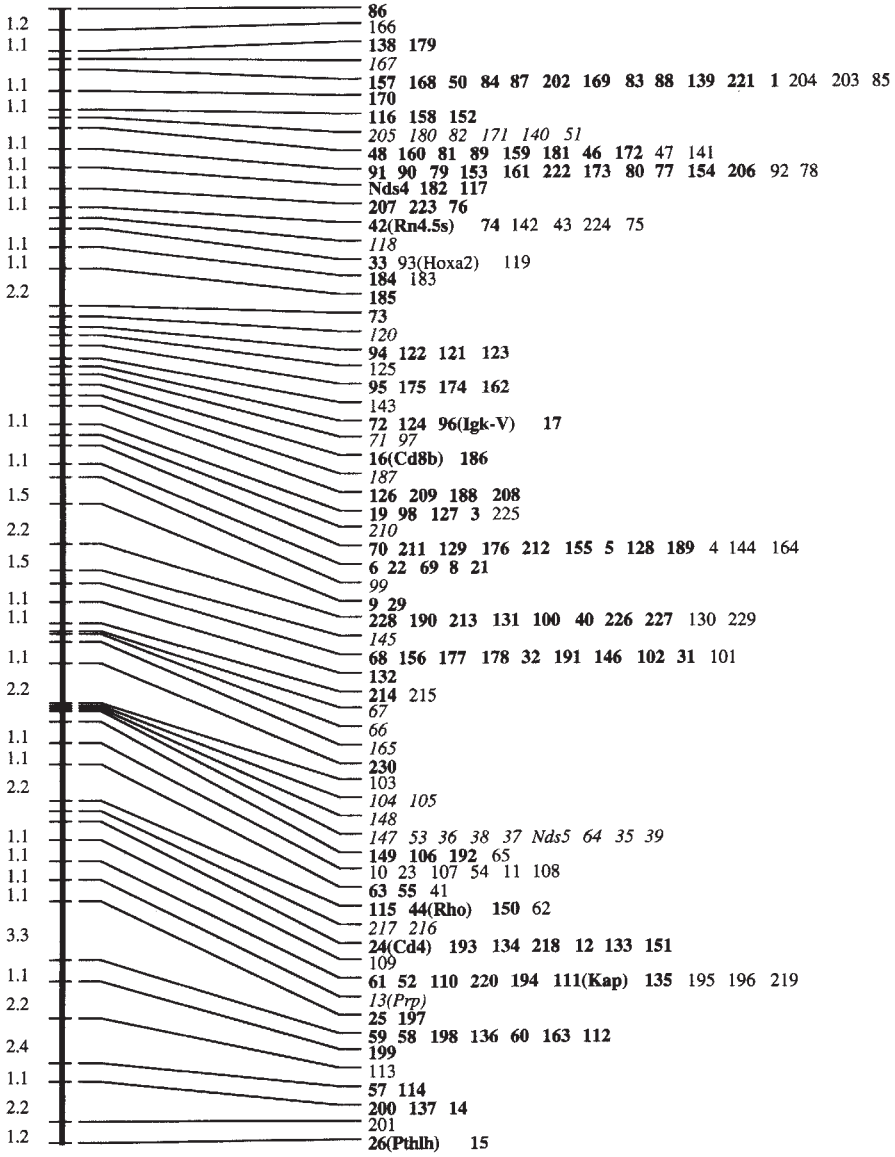


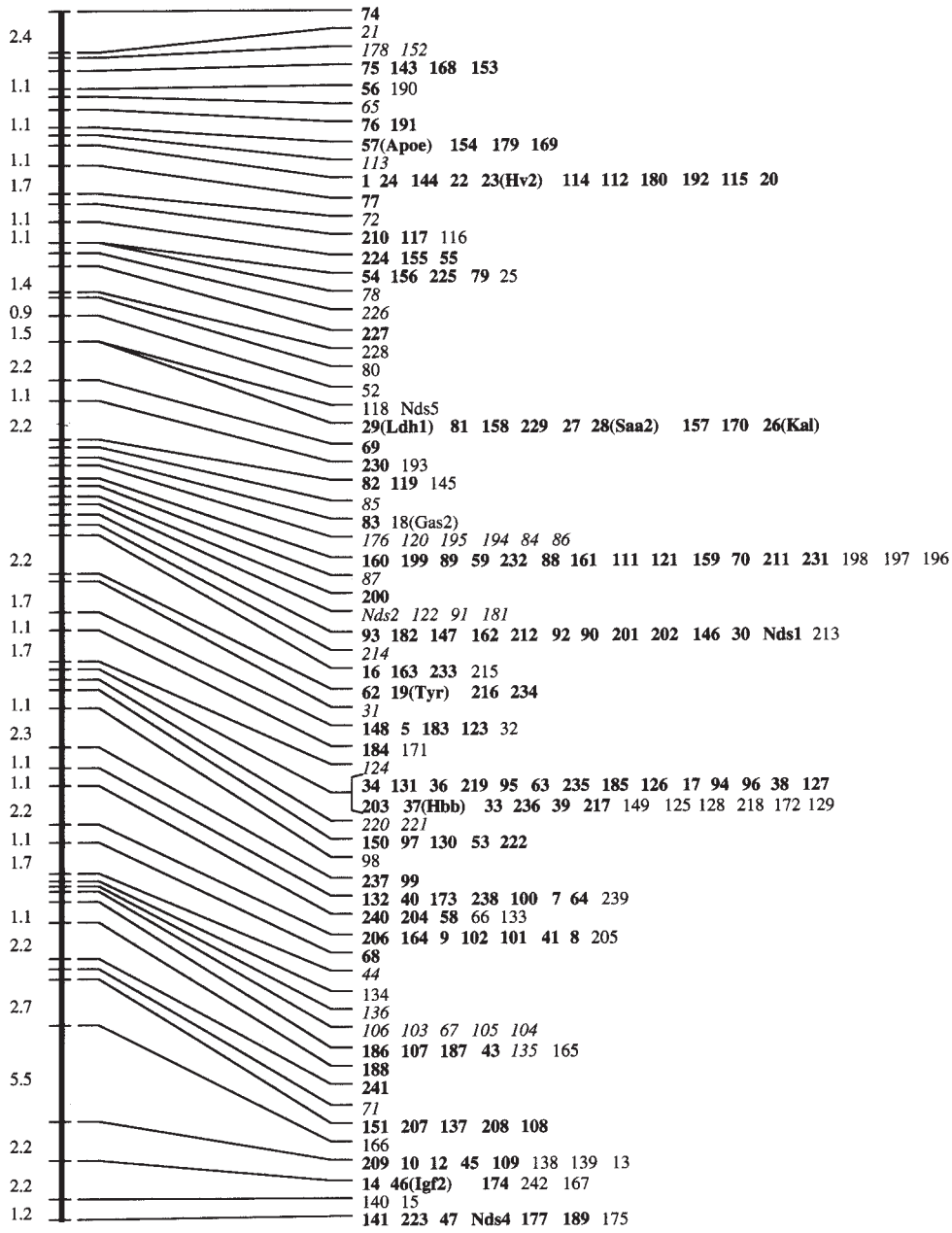
Chromosome 4



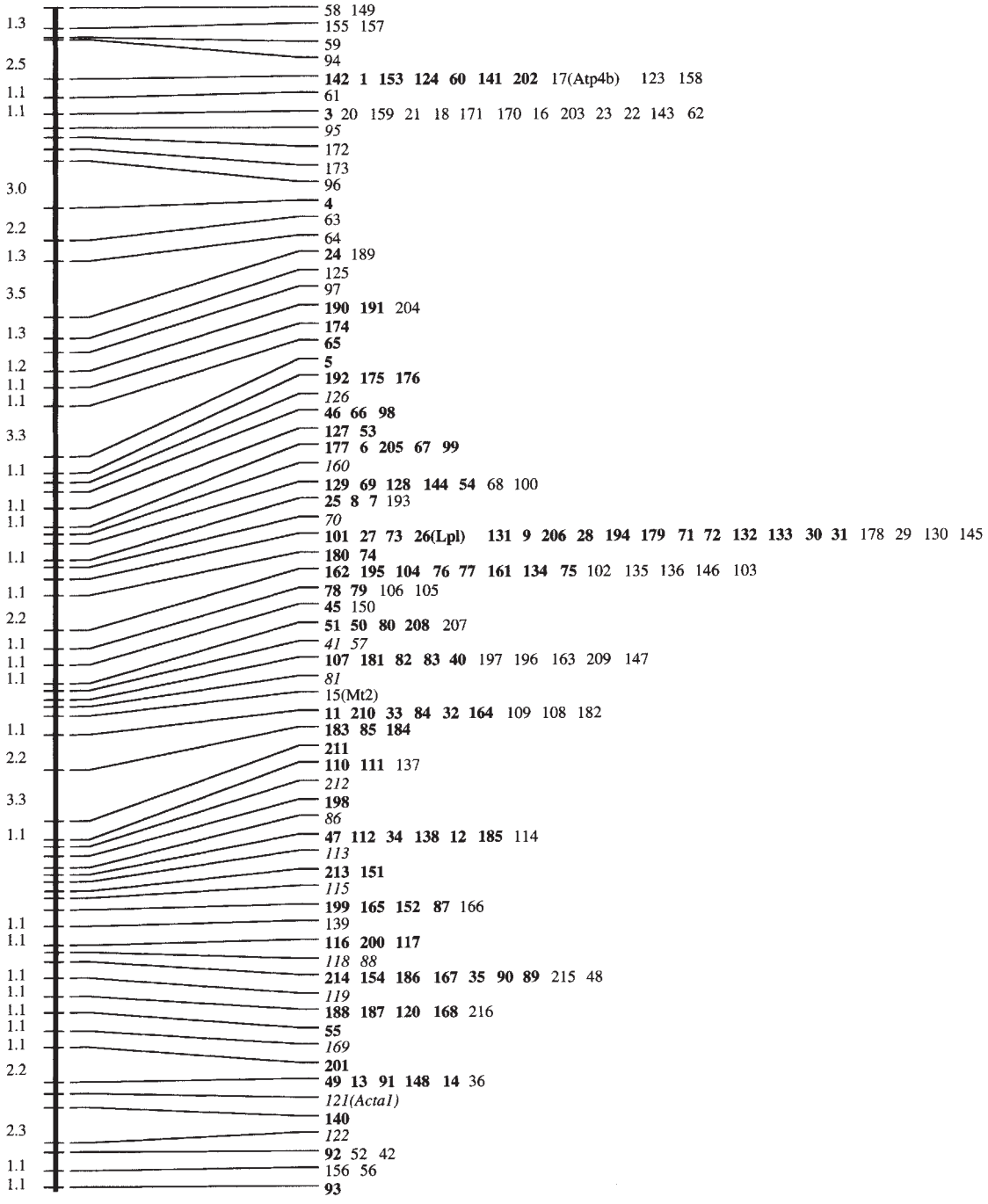


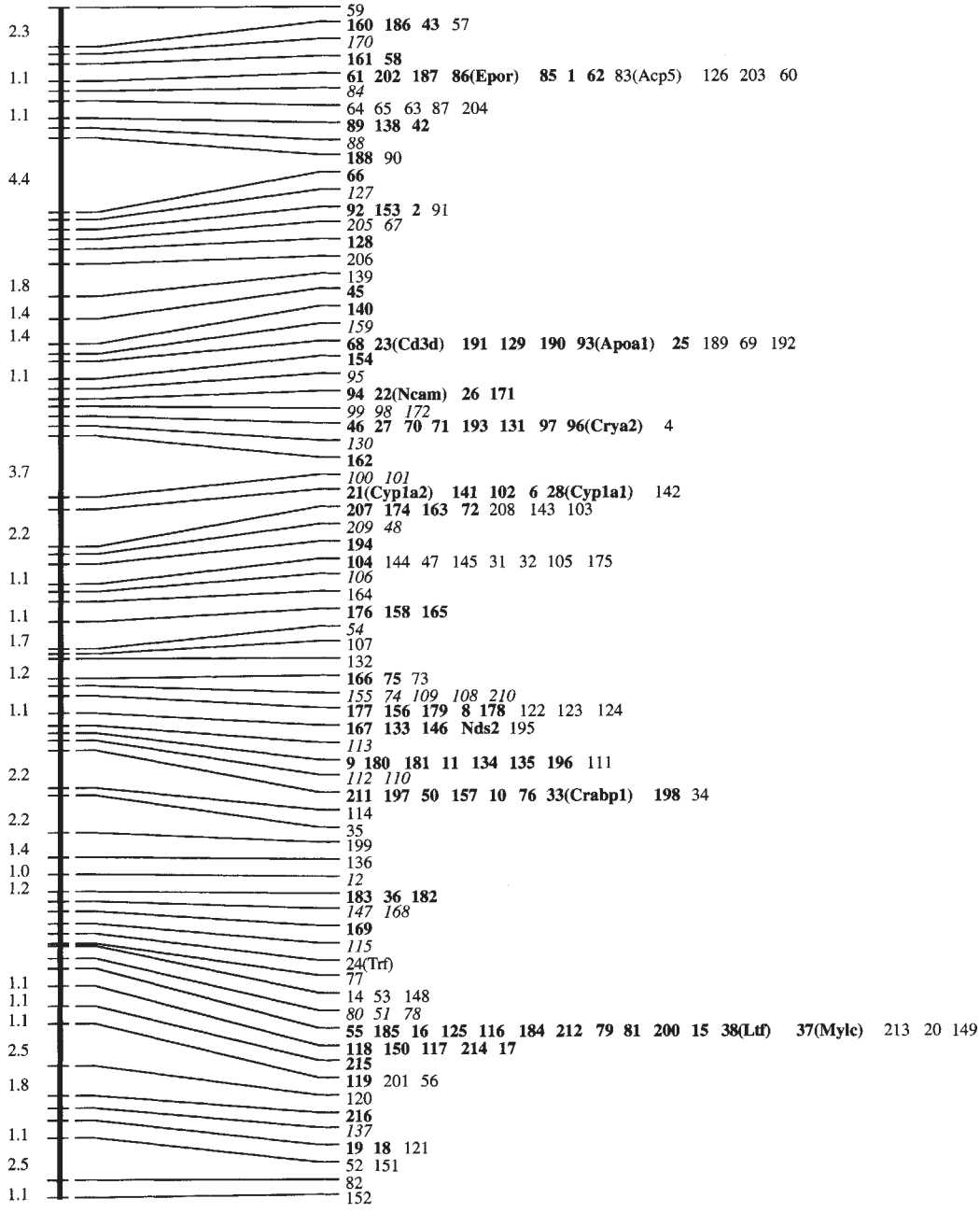
Chromosome 6



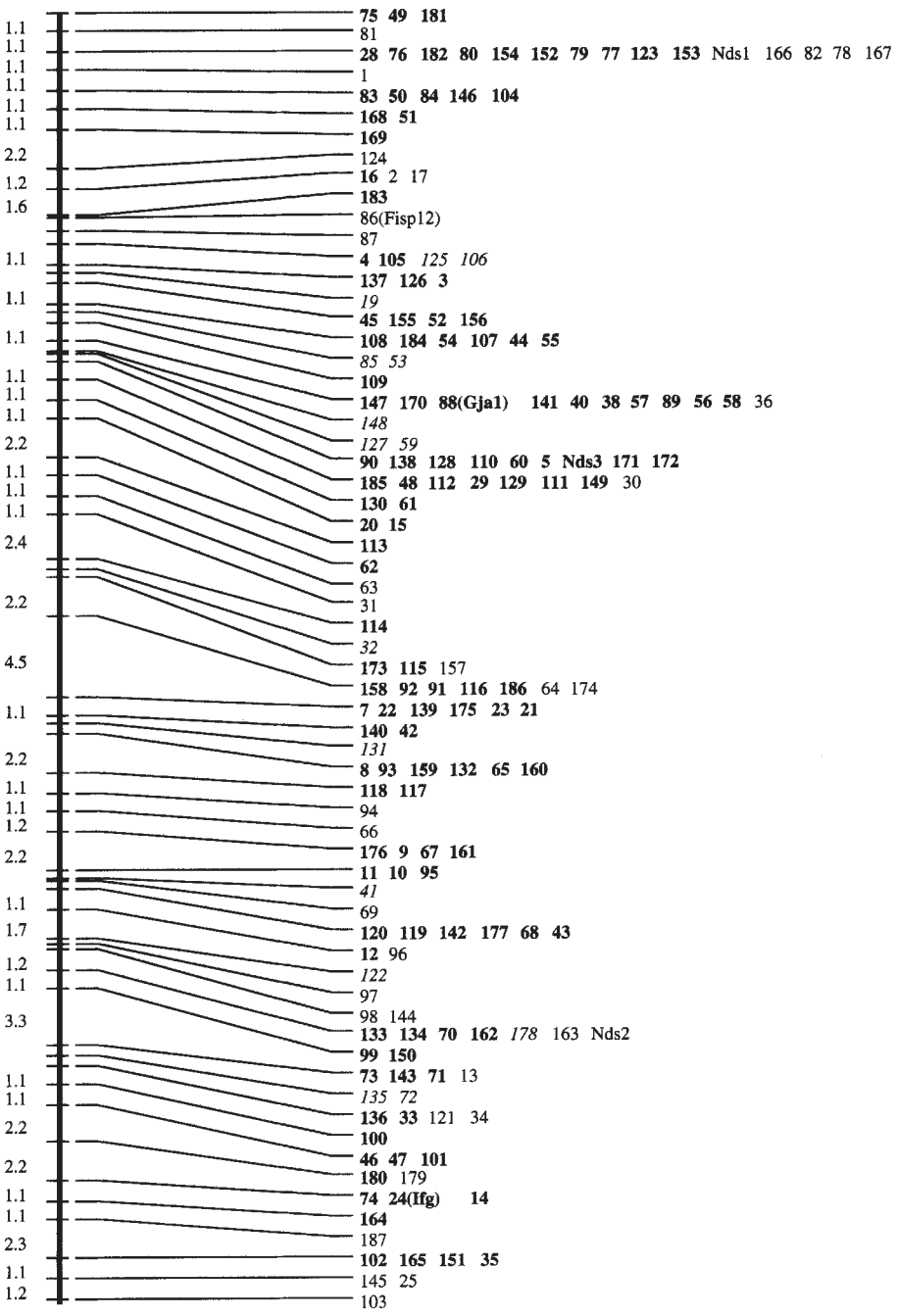


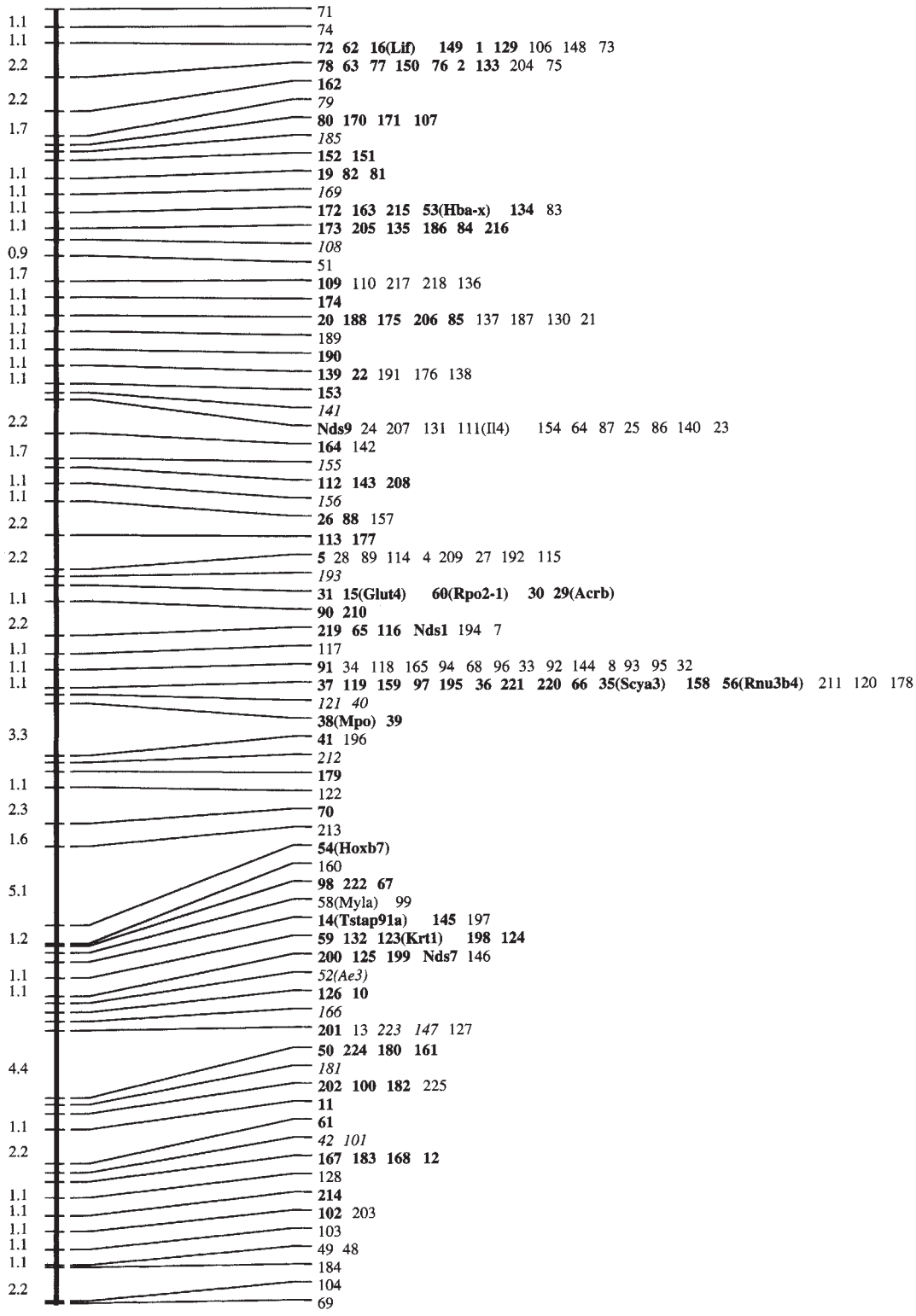
Chromosome 8



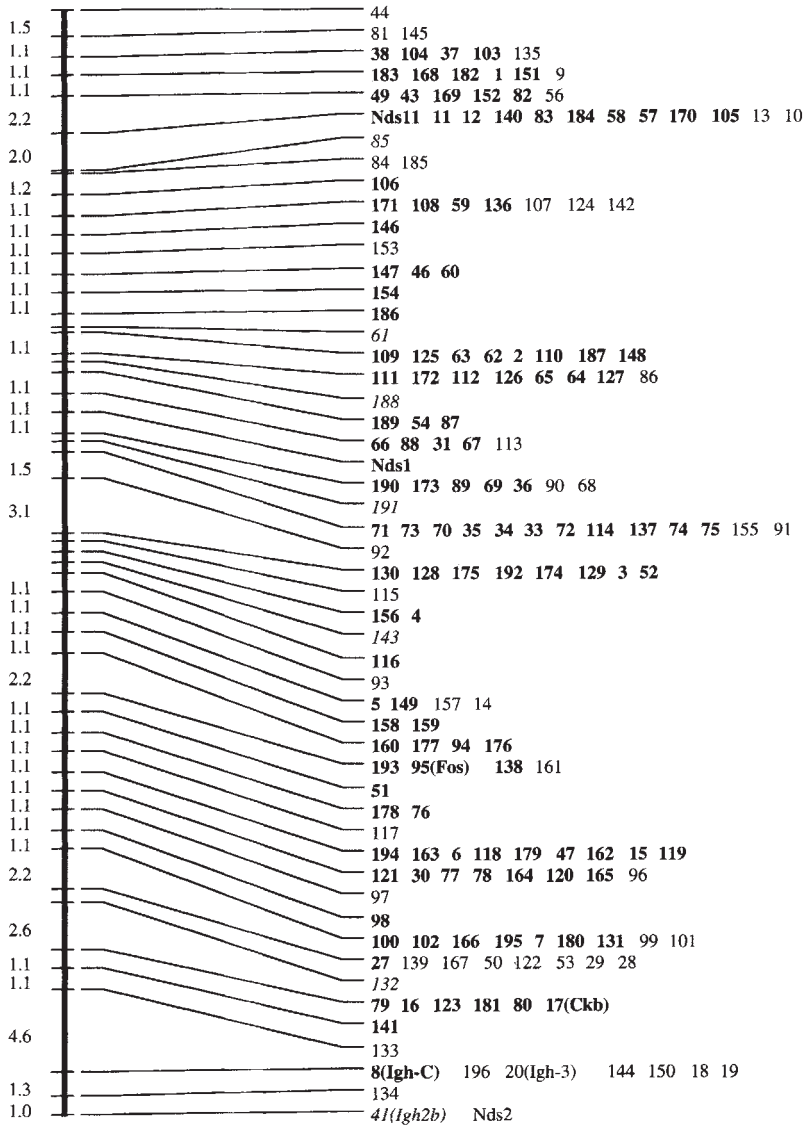


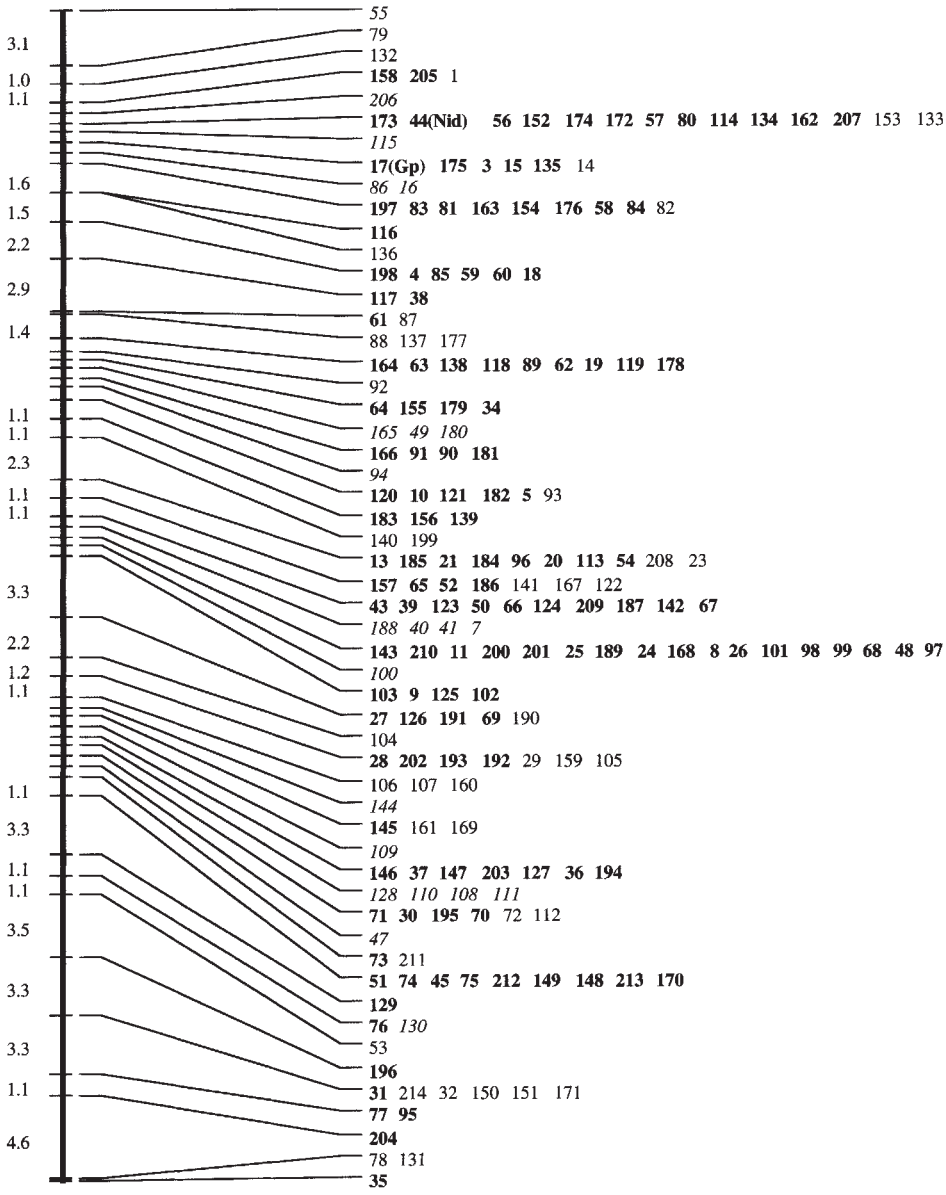
Chromosome 10

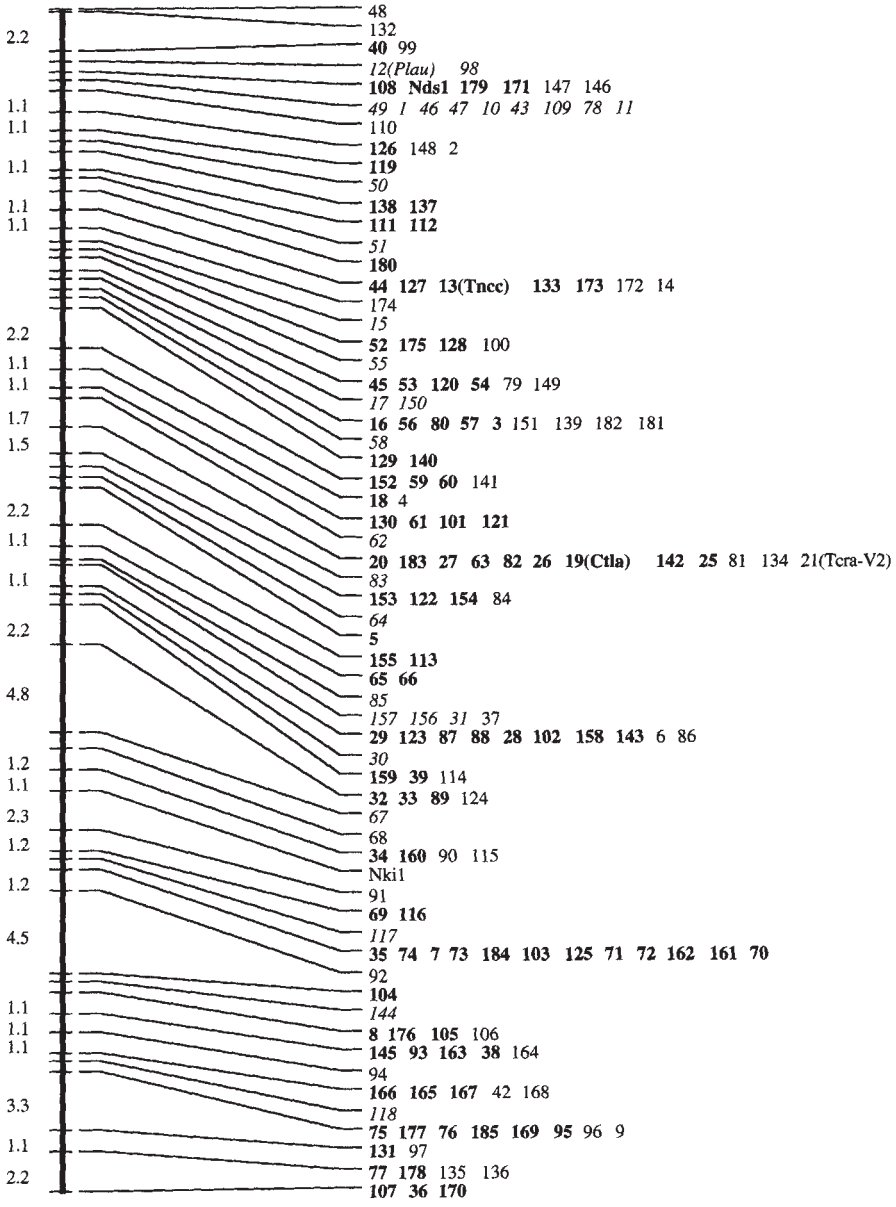


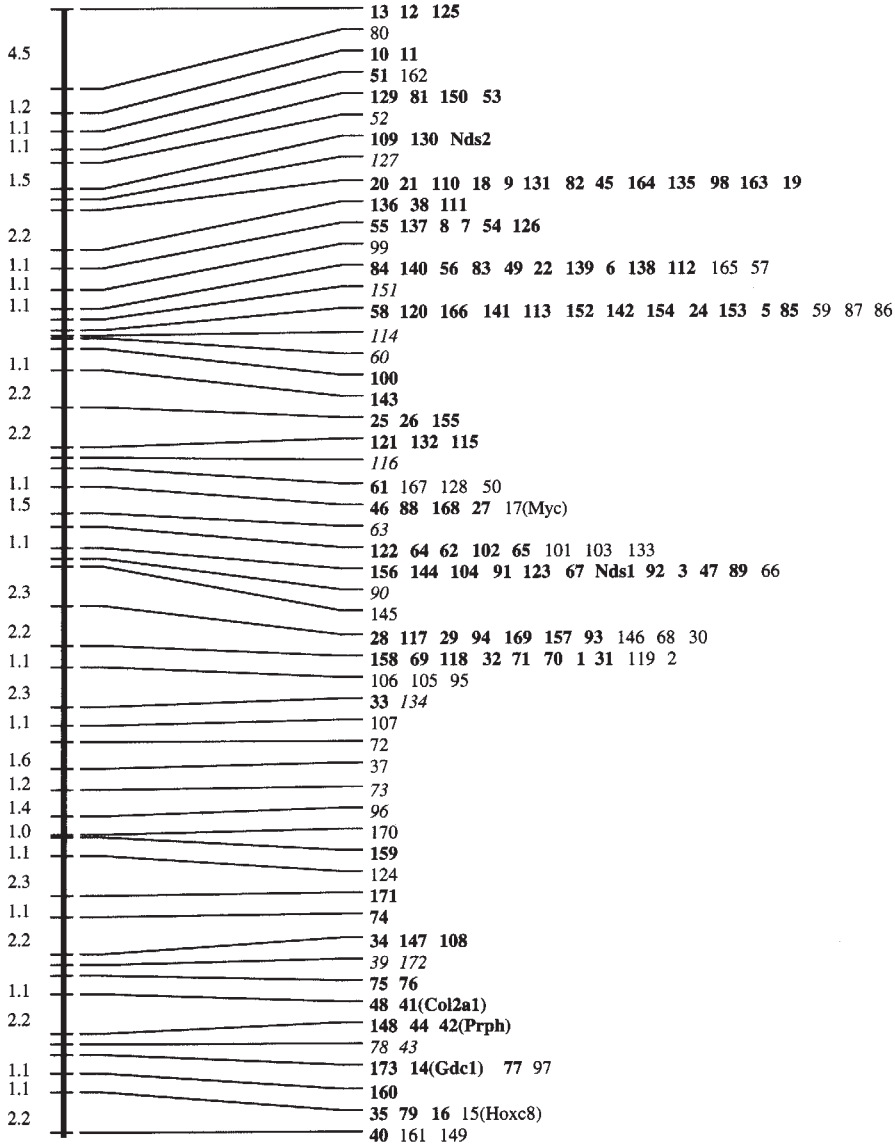


Chromosome 12

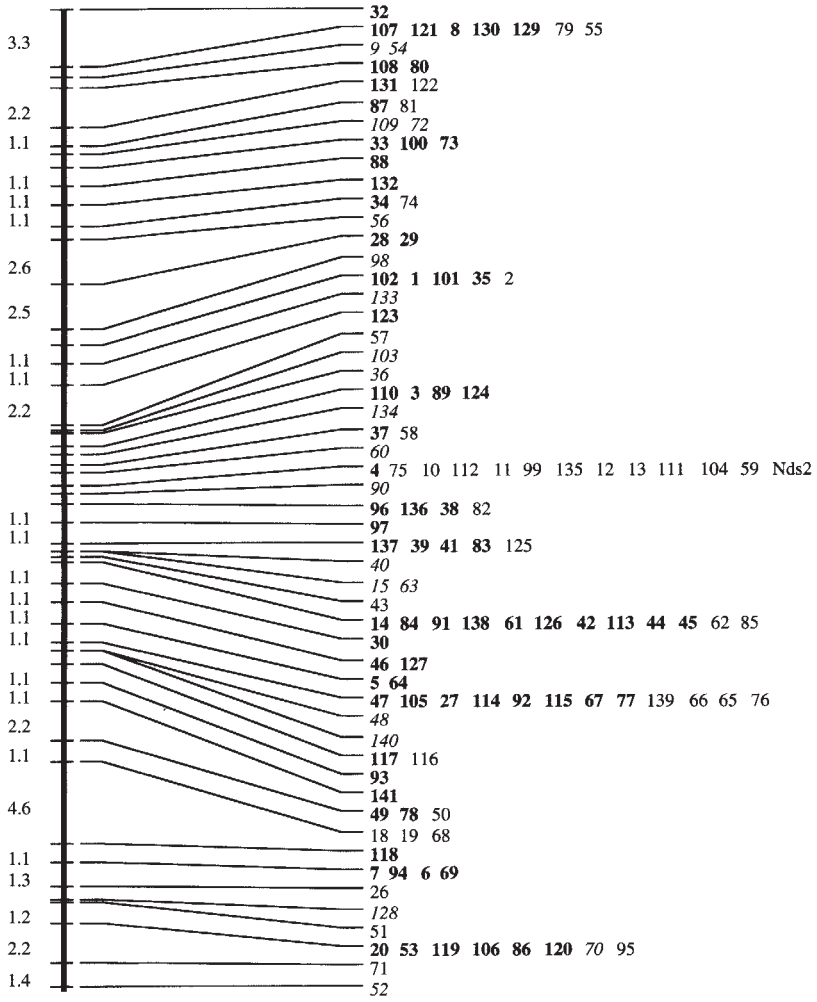




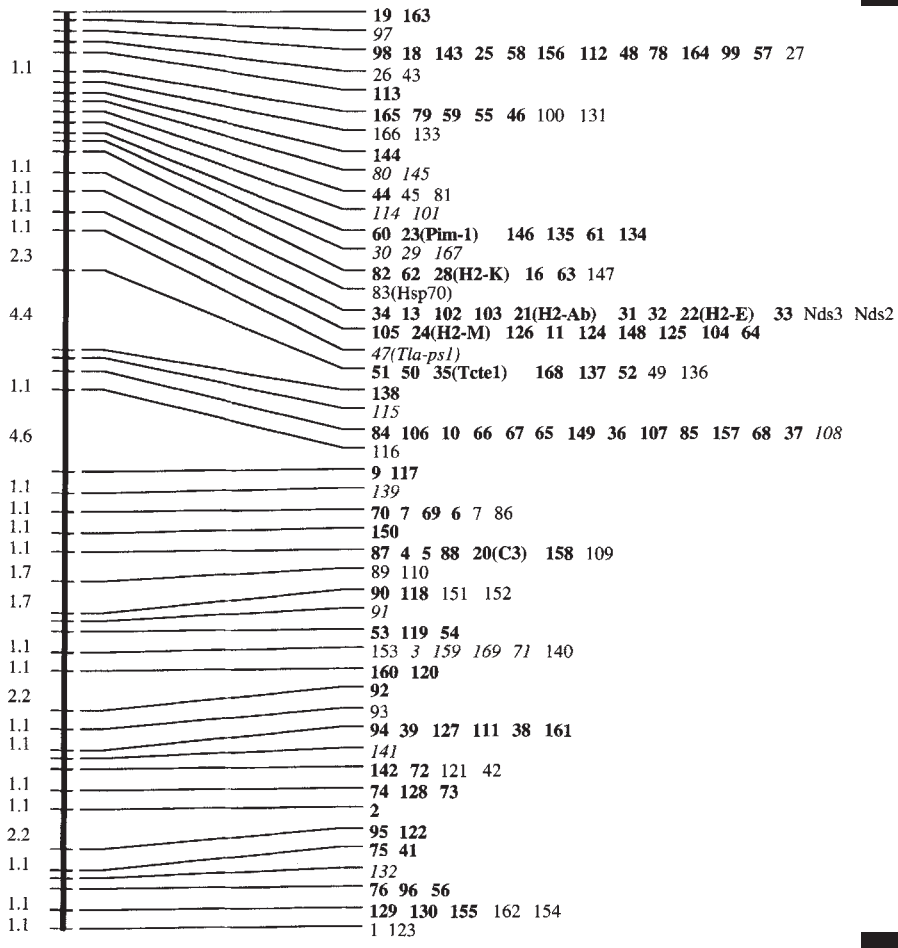




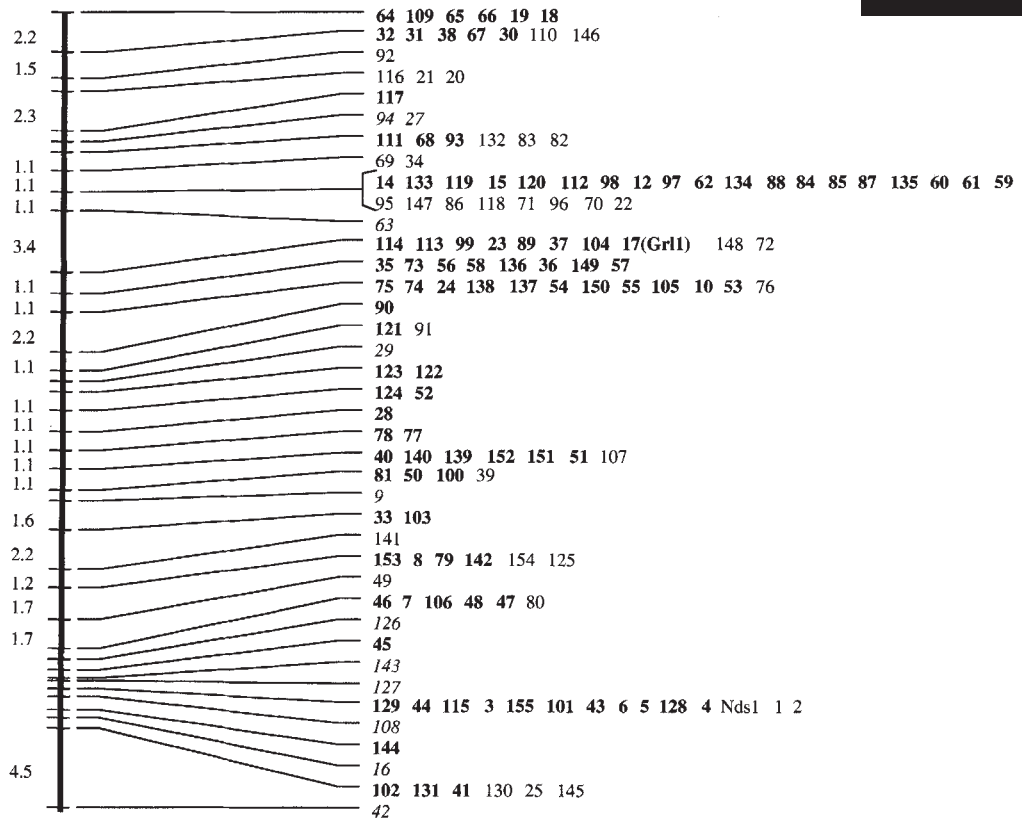
Chromosome 16



Chromosome 17



Chromosome 18



Chromosome 19