TABLE 1

Genetic markers associated with restriction fragment length polymorphisms[a]

Marker	Probe	Alleles	Strains
D12Nyu1	M13φ7-97	a (T: 3.4 + 1.8)	BALB/cJ, AKR/J, STS/A
		b (T: 3.8 + 1.8)	C57BL/6J, O20/A
D12Nyu2	M13φ19-25	a (T: 4.1 + 2.7)	BALB/cJ, O20/A, C57L/J, DBA/2J, STS/A
		b (T: 3.9 + 3.6)	C57BL/6J
		c (T: 4.1 + 2.8)[b]	AKR/J
D12Nyu4	M13φ26-15	a (T: 7.7 + 1.8)	O20/A
		b (T: 9.2)	AKR/J, STS/A
D12Nyu5	M13φ30-3	a (M: 4.4 + 3.7)	BALB/cJ, C3H/HeJ, O20/A, STS/A
		b (M: 3.8)	C57BL/6J
D12Nyu6	pBRφ49-9	a (R: 6.9)	SWR/J, SJL/J
		b (R: 5.6)	C57BL/6J, O20/A, AKR/J, C57L/J, DBA/2J, C3H/HeJ, BALB/cJ, STS/A
D12Nyu7	pBRφ39L5	a (R: 3.8)	BALB/cJ, DBA/2J
		b (R: 4.1)	C57BL/6J, O20/A, AKR/J, C57L/J, SWR/J, C3H/HeJ, SJL/J, STS/A
D12Nyu8	pUCφ38H10	a (T: 2.5)	SWR/J, SJL/J
		b (T: 2.4)	C57BL/6J, AKR/J, C57L/J, DBA/2J, C3H/HeJ, BALB/cJ
D12Nyu9	pUCφ57UX17	a (R: 6.1 + 4.2)	AKR/J
		b (R: 6.1)	C57L/J, DBA/2J
D12Nyu10	pUCφ43G14	a (R: 3.0)	DBA/2J, BALB/cJ
		b (R: 2.8)	C57BL/6J, AKR/J, C57L/J, SWR/J, C3H/HeJ, SJL/J
Aat-1	pG3.5	a (M: —[c])	BALB/cJ
		b (M: 1.7 + 1.2)	STS/A
Mtv-9	pMtv9-5′c	a (M: 8.5)	BALB/cJ, SWR/J, DBA/2J, C3H/HeJ, SJL/J
		b (M: 3.7)	C57BL/6J, AKR/J, C57L/J
Igh-Ca	p3′α	a (T: 14 + 3.5 + 2.8)	BALB/cJ, STS/A
		b (T: 18 + 5.6)	C57BL/6J, SJL/J, O20/A
		c (T: 14 + 4.9)	AKR/J
		d (T: 14 + 3.7 + 2.9)	C57L/J
		e (T: 18 + 9.4)	SWR/J
		f (T: 18 + 14)	DBA/2J
		g (T: 5.4 + 3.1)	C3H/HeJ

[a] Liver DNA digested with *Eco*RI (R), *Msp*I (M), or *Taq*I (T) restriction endonucleases was analyzed by Southern blotting. Fragment sizes in kilobases were calculated from their mobilities relative to those of *Hin*dIII-digested λcI857 DNA fragments.

[b] The mobility difference that distinguished the smaller DNA fragment of alleles a and c was not detected previously (D'EUSTACHIO 1984).

[c] In addition to the fragments listed, DNA from both strains yielded approximately 10 invariant fragments.

done on a VAX/VMS computer at NYU Medical Center. Programs and documentation are available from P.D.

RESULTS

Previous studies suggested that the positions occupied by the four anonymous DNA markers *D12Nyu1, 2, 3* and *4*, the *Fos* proto-oncogene, *Igh* and *Aat-1* spanned most of chromosome *12* (D'EUSTACHIO 1984). To construct an accurate map of the chromosome, inheritance of these markers and *Mtv-9* (PETERS *et al.* 1986) was followed in a backcross between C57BL/6J and SWR/J mice. (C57BL/6J × SWR/J)F_1 females were mated to males of both inbred progenitor strains, and females of both strains were mated to F_1 males. Liver genomic DNA from each of 317 backcross individuals was typed by Southern blotting for inheritance of alleles at all eight markers. Within sampling fluctuation, alleles of each marker segregated 1:1 in all four backcrosses and recombination fractions measured in crosses to C57BL/6J inbred mice were the same as those measured in crosses to SWR/J inbred mice. Data from both crosses to inbred females were pooled as were data from both crosses to inbred males, to test the effect of F_1 parental sex on recombination fraction. While most recombination fractions measured in crosses to F_1 males were smaller than the corresponding fractions measured in crosses to F_1 females, only 4 of the 21 differences were of even marginal statistical significance ($0.05 > P > 0.01$).

To construct a linkage map of the chromosome, the

TABLE 2

Recombination fractions measured among markers of chromosome *12* in backcross experiments[a]

Marker	*D12Nyu2*	*D12Nyu1*	*D12Nyu3*	*Fos*	*D12Nyu4*	*Mtv-9*	*Aat-1*
D12Nyu1	36/317*						
D12Nyu3	39/317*	7/317*					
Fos	86/317*	54/317*	47/317*				
D12Nyu4	90/317*	58/317*	51/317*	4/317*			
Mtv-9	96/317*	64/317*	57/317*	10/317*	6/317*		
Aat-1	118/317°	90/317*	83/317*	36/317*	32/317*	26/317*	
Igh-Ca	158/317	140/317	133/317	90/317*	86/317*	80/317*	54/317*

[a] Each fraction is the (number of individuals recombinant for a pair of markers)/(the total number of individuals examined). Bayesian probabilities that the observed deviations from 1:1 segregation were due to chance fluctuation are indicated: ° $0.01 \geq P > 0.001$; * $0.001 \geq P$.

BAYLOC algorithm was applied to data pooled from all four backcrosses (Table 2). Figure 1 shows the 99% confidence interval for the localization of each marker in the maps generated as successive markers were added. Figure 2A shows the final map of the chromosome. To test the possibility that, although individual recombination fractions were not significantly reduced in backcrosses to F_1 males, there was a significant trend overall, maps were constructed using only F_1 male data (Figure 2B) and only F_1 female data (Figure 2C). The female map was 1.3 times the length of the male map. The *D12Nyu3–Fos* interval was unchanged from the combined map; all other intervals in the male map were reduced in length.

To test the effect of the order in which markers were added on the final map, order of addition was systematically varied. The analysis proceeded in two stages. First, five markers, *D12Nyu2*, *1* and *4*, *Aat-1* and *Igh-Ca*, were chosen because the observed recombination fractions indicated that each was separated from the other four by a recombination fraction of at least 0.1. Each of the ten possible initial pairs of markers was tested for each of the six possible orders of addition of the remaining markers. Twenty of the 60 paths to a map failed at the first step: *D12Nyu2* appeared unlinked to *Aat-1* and to *Igh-Ca* considered pairwise, and *D12Nyu1* appeared unlinked to *Igh-Ca* (Table 2). All remaining paths yielded the same five-locus map, and every marker was placed in its interval in every intermediate-stage map at >99% confidence. The second stage was to consider the three markers, *D12Nyu3*, *Fos* and *Mtv-9*, initially excluded from consideration because of their close linkage to one of the initial group of five. All possible orders of addition of these markers to the five-locus map were tested, and yielded the identical final map. Further, if *Mtv-9* was added first, every marker was again placed in its interval at >99% confidence in every intermediate stage map.

The linkage data were tabulated by haplotype, assuming the marker order generated by the BAYLOC analysis (Table 3). The gene order predicted by BAYLOC agreed qualitatively with the haplotype tabulation: chromosomes predicted to require double or multiple crossover events were observed infrequently. This point was tested quantitatively, using the method of BISHOP (1985), for two trios of closely linked markers, *D12Nyu1*, *2* and *3*, and *Fos*, *D12Nyu4* and *Mtv-9* (Table 4). In both cases, the order generated by the BAYLOC algorithm was found most probable by the BISHOP method.

Although the number of backcross animals studied was small, the length of the interval mapped and the number of markers localized within it were sufficient to allow effects of interference to be examined. The 317 backcross chromosomes revealed 180 recombination events (Table 3). If these events occurred randomly (*i.e.*, if there were no interference), the number of chromosomes with 0, 1, 2, 3, ... events would follow a Poisson distribution with a mean of 180/317 = 0.568. 179.7, 102.0, 29.0 and 6.4 chromosomes would be expected with 0, 1, 2 or 3+ events, respectively. The numbers of such chromosomes found were 148, 158, 11 and 0 (Table 3). The deviation between the two distributions is significant: $\chi^2 = 53.79$ with 2 d.f., $P < 0.001$ [SNEDECOR and COCHRAN (1967), pp. 223–226 and 236–238].

These interference effects could not be attributed to any particular region of the chromosome. Numbers of double crossover events found were consistently smaller than expected on the hypothesis of no interference, but the deviation from expectation, calculated as a binomial probability was significant at the 0.05 probability level only in three cases and significant at the 0.01 level in none.

Using the cloning strategy described previously (D'EUSTACHIO 1984), five additional polymorphic genomic DNA fragments from chromosome *12* were identified (Table 1). To localize these markers on the chromosome, their inheritance in RI strains of mice was determined (Table 5). *D12Nyu6* was indistinguishable by recombination from *D12Nyu3* (0 of 17 strains recombinant, $c_{95} < 4.8$ cM); *D12Nyu8* was indistinguishable from *Fos* (0 of 15 strains recombinant, c_{95}

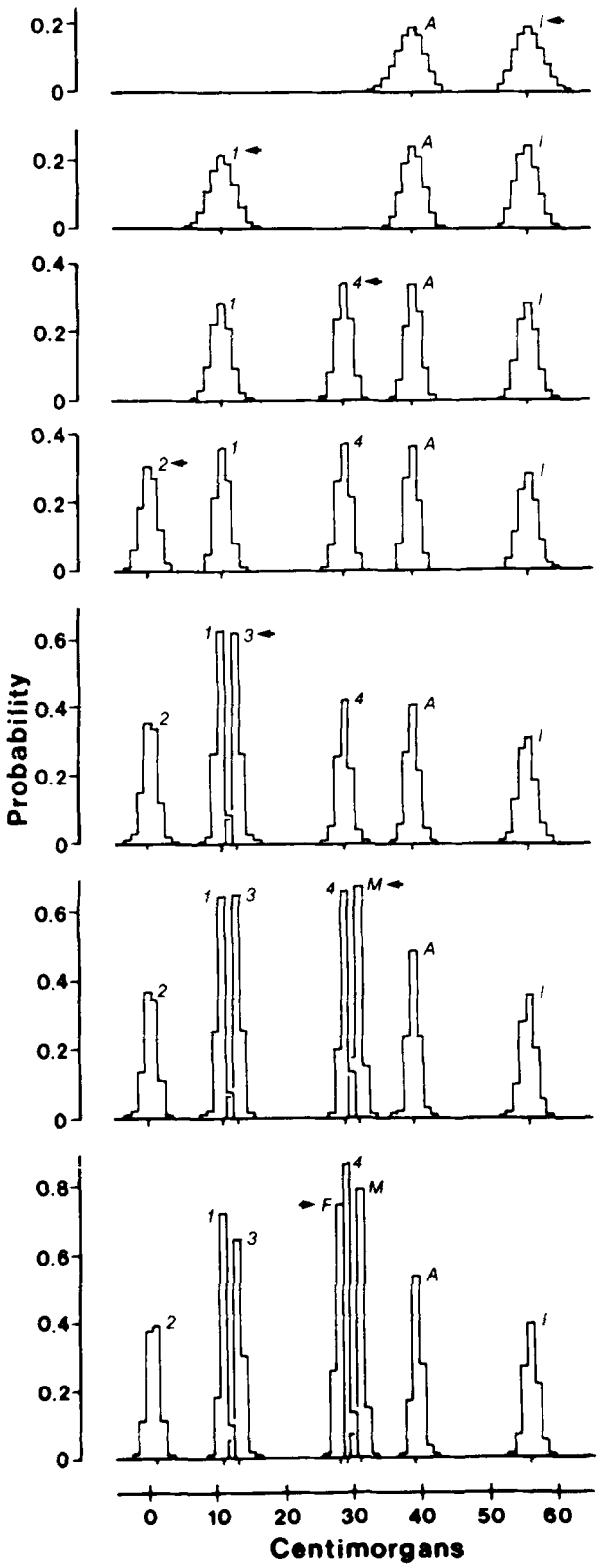

FIGURE 1.—Bayesian probability distributions for markers in successive maps of chromosome *12*. The probability distributions are shown for each marker, contingent on the observed recombination fractions and the placements of the other markers in the map. The marker newly added at each cycle, proceeding downwards in the figure, is indicated with an arrow. Marker names are abbreviated: *1*, *D12Nyu1*; *2*, *D12Nyu2*; *3*, *D12Nyu3*; *4*, *D12Nyu4*; *A*, *Aat-1*; *F*, *Fos*; *I*, *Igh-Ca*, *M*, *Mtv-9*.

< 5.7 cM); and *D12Nyu9* was indistinguishable from *Igh-C* (0 of 55 strains recombinant, c_{95} < 1.4 cM). c_{95} is the largest distance consistent with the observation of no recombinants at a Bayesian probability of 0.05. No recombinants were found between *D12Nyu8* and *Fos* in 46 backcross animals, confirming its localization.

Two markers, *D12Nyu7* and *10*, defined new genetic loci. They were closely linked to each other and to markers previously localized to the proximal region of the chromosome by analysis of RI strains (COBB, STOMING and WHITNEY 1987; D'EUSTACHIO 1984; HILGERS and ARENDS 1985; LUSIS *et al.* 1987; TRAINA, TAYLOR and COHEN 1981; VON DEIMLING and TAYLOR 1987). Attempts to build an unambiguously ordered linkage map of chromosome *12* from these RI data alone failed because of the small numbers of strains typed and large proportions of recombinant strains observed for most pairs of markers. We therefore constrained the five markers mapped in the backcross experiment and distinguishable in RI strains to lie in the order *D12Nyu2–D12Nyu1–Mtv-9–Aat-1–Igh-C* on a chromosomal segment spanning no more than 100 cM. The other markers could then be added to this group to yield the map of the proximal region of chromosome *12* shown in Figure 2D. *Es-25*, *D12Nyu10*, *D12Nyu7*, and *Apob* formed a cluster proximal to *D12Nyu2*. *Ah/Ly-18* (no RI strains are recombinant for these two loci—VON DEIMLING and TAYLOR 1987), and *D12Nyu5* formed a second cluster between *D12Nyu2* and *D12Nyu1*.

DISCUSSION

Data from backcross experiments and from the analysis of RI strains have been used to construct a detailed linear linkage map that spans most of the genetic length of mouse chromosome *12*. *Igh* is the most distal marker on the map. Significant interference effects were observed in the backcross experiment, and the overall length of a map derived from male meiotic data was substantially shorter than that of a map derived from female data.

Four points arise from the application of the BAYLOC algorithm to the analysis of chromosome *12*. First, this approach places each marker at the position that simultaneously maximizes the odds of finding the observed recombination fractions with all other markers in the linkage group. The use of this constraint to refine the position of a marker in a map is shown graphically in Figure 1. While the greatest reduction in the confidence interval associated with any marker occurred following the next few additions of markers to the map, all additions had some effect. Second, after a new marker is added to a map, each marker in the map is allowed in turn to move freely until all come to rest. This iterative procedure must converge.

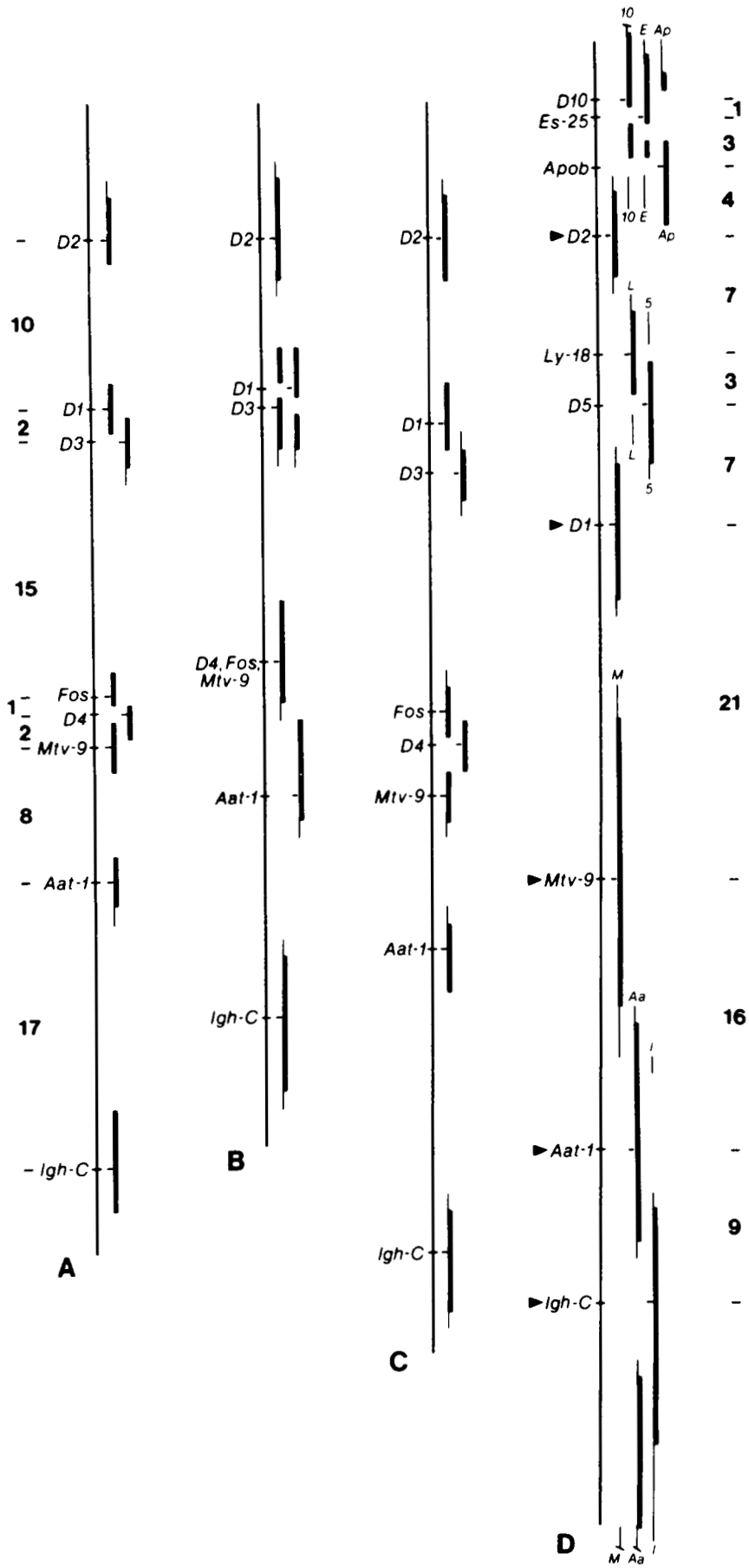

FIGURE 2.—Four views of mouse chromosome *12*. Linkage maps were assembled with the BAYLOC algorithm using (A) combined data from all backcrosses (Table 2); (B) data from backcrosses to F_1 males; (C) data from backcrosses to F_1 females; and (D) data from RI strains. The centromere is at the top of each map. The 95% (——) and 99% (——) confidence limits are shown for the placement of each marker, contingent on the observed recombination fractions and the placements of the other markers in the map. Distances between successive markers in cM are shown in panels A and D. The five markers in the RI map forced to lie in the same order as in the backcross map are indicated by arrowheads. The anonymous DNA markers *D12Nyu1*, *2*, *3*, *4*, *5*, *7* and *10* are labeled *D1*, *D2*, *D3*, *D4*, *D5*, *D7* and *D10* for simplicity. All other marker names are given in full.

TABLE 3

Haplotypes observed for chromosome *12*[a]

CROSS	$F_1 \times S$	$F_1 \times B$	$S \times F_1$	$B \times F_1$	ALL
HAPLOTYPE					
NO CROSSOVERS were found in 148 chromosomes:					
b b b b b b	25	18	22	13	148
s s s s s s	18	22	13	17	
ONE CROSSOVER was found in 158 chromosomes:					
in the *D12Nyu2–D12Nyu1* interval					
s b b b b b b b	3	7	1	2	27
b s s s s s s s	1	6	5	2	
in the *D12Nyu1–D12Nyu3* interval					
s s b b b b b b	0	1	1	0	5
b b s s s s s s	2	1	0	0	
in the *D12Nyu3–Fos* interval					
s s s s b b b b	10	6	2	4	45
b b b b s s s s	4	8	8	3	
in the *Fos–D12Nyu4* interval					
s s s s b b b b	2	1	0	0	4
b b b b s s s s	1	0	0	0	
in the *D12Nyu4–Mtv-9* interval					
s s s s s b b b	1	2	0	0	6
b b b b b s s s	3	0	0	0	
in the *Mtv-9–Aat-1* interval					
s s s s s s b b	3	7	1	1	24
b b b b b b s s	4	2	3	3	
in the *Aat1–Igh-Ca* interval					
s s s s s s s b	10	9	4	3	47
b b b b b b b s	7	5	5	4	
TWO CROSSOVERS were found in 11 chromosomes:					
in the *D12Nyu2–D12Nyu1* and *D12Nyu1–D12Nyu3* intervals					
b s b b b b b b	1	0	0	0	2
s b s s s s s s	1	0	0	0	
in the *D12Nyu2–D12Nyu1* and *Mtv-9–Aat-1* intervals					
b s s s s b b b	0	1	0	0	2
s b b b b s s s	1	0	0	0	
in the *D12Nyu2–D12Nyu1* and *Aat-1–Igh-Ca* intervals					
b s s s s s s b	0	1	1	0	5
s b b b b b b s	1	2	0	0	
in the *D12Nyu3–Fos* and *Aat-1–Igh-Ca* intervals					
b b b s s s s b	1	0	0	0	2
s s s b b b b s	0	0	1	0	
THREE OR MORE CROSSOVERS were found in no chromosomes					

[a] Haplotypes consist of the alleles found in the chromosome inherited from the heterozygous parent, in the order *D12Nyu2, D12Nyu1, DC12Nyu3, Fos, D12Nyu4, Mtv-9, Aat-1* and *Igh-C*. "b" indicates an allele of C57BL/6J origin, and "s" an allele of SWR/J origin. Thus, bssssss denotes a chromosome carrying the C57BL/6J allele of *D12Nyu2* and the SWR/J alleles of all other markers.

In fact, it does so rapidly: all maps stabilized in less than six cycles; most did so in one-three cycles. Third, if the constraint is applied that a new marker must fix in an interval at high confidence (here, summed probability >0.99), every order of addition of markers that leads to any map at all leads to the same one. The BAYLOC algorithm, applied to closely spaced groups of markers, yielded orders and associated confidence estimates comparable to those yielded by BISHOP's (1985) method (Table 4). It has the advantage over the latter of not being restricted, practically, to the analysis of small numbers of markers. Finally, our algorithm considers the genome as a series of discrete segments of equal size, and segment size can be changed. In the case of chromosome *12*, a segment of 1 cM is smaller than the confidence intervals associated with the marker positions in the map (Figure 2). A smaller segment size appears unlikely to provide any real gain in resolution except for data sets much larger than the one analyzed here.

The empirical criteria for choosing an order in which to add markers to a map are straightforward. Initial pairs of markers should be ones typed in large numbers of individuals and should yield a recombination fraction implying a large (>10 cM) distance between them. Markers added early in the process should likewise show large recombination fractions with those already in the map. Markers that show infrequent recombination with ones already placed should be added late in the process. All of the paths

TABLE 4

Maximum likelihood analysis of gene orders for trios of markers[a]

Order	Log likelihood	LOD for order
D12Nyu1, *D12Nyu2* and *D12Nyu3*:		
D12Nyu2–D12Nyu1–D12Nyu3	−63.3204	
D12Nyu2–D12Nyu3–D12Nyu1	−65.9378	2.6174
D12Nyu3–D12Nyu2–D12Nyu1	−100.0622	36.7418
D12Nyu4, *Fos* and *Mtv-9*:		
Fos–D12Nyu4–Mtv-9	−22.2406	
D12Nyu4–Fos–Mtv-9	−28.6065	6.3659
Fos–Mtv-9–D12Nyu4	−32.2027	9.9621

[a] For each group of three markers, likelihood scores were calculated for all three orders using data pooled from all backcrosses (BISHOP 1985). Their base-10 logarithms are shown. The most likely order is listed first in each group. The difference between its $\log_{10}$ value and that for each of the other orders is the LOD for order shown in the last column.

to a complete map of chromosome *12* that failed, it should be noted, violate these criteria. Otherwise, the order of marker addition appears to have little effect on the efficiency with which map building proceeds and none on the final map.

RI typing data alone were insufficient to allow the construction of an unambiguous linkage map of chromosome *12*. When five of the markers were provisionally located, based on their positions in the backcross map, the RI data could be used without further constraints to build a map of the chromosome. The RI map showed good overall agreement with the backcross map (Figure 2, A and D). However, if only RI data are considered, alternative orders cannot be excluded at high confidence. This observation probably explains previous failures to deduce a correct map of the distal portion of the chromosome from RI data (*e.g.*, D'EUSTACHIO 1984; OWEN, RIBLET and TAYLOR 1981).

At the same time, this result suggests that if an ordered group of markers is available, RI data can reliably be interpolated into it. Here, it allowed the orientation of the map with respect to the centromere of the chromosome simply by noting the tight linkage between *Apob* and a polymorphism associated with the nucleolar organizer region of chromosome *12* (LUSIS *et al.* 1987).

Three-point crosses (IKEDA, SATO and OKADA 1981; OKADA *et al.* 1981) have established the marker order *Igh–Aat-1–Fv-4*. The recombination fraction between *Aat-1* and *Igh* measured in these crosses (13/123) is not significantly different from ours (54/317—Table 2). Construction of a multilocus map with data combined from both sets of backcrosses suggested a localization of *Fv-4* very near the *Fos–D12Nyu4–Mtv-9* cluster. Its order with respect to the cluster is uncertain, but the close linkage of the integration sites of two proviruses, *Fv-4* (GARDNER, DANDEKAR and CARDIFF 1986) and *Mtv-9* (PETERS *et al.* 1986) and a proto-oncogene is intriguing. Further analysis will be necessary to define the relationship among these markers more precisely, and to investigate the possibility that some feature of this region of the chromosome makes it a target for proviral integration.

The use of multilocus analysis to detect trends in a set of linkage data is illustrated by the analysis of the effects of sex and interference on the map of chromosome *12*. Sex differences in recombination fractions are common in the mouse, but the extent and even the direction of the differences vary with the marker pair examined (DUNN and BENNETT 1967; DAVISSON and RODERICK 1981). We likewise found no significant difference in individual recombination fractions between male and female meioses. A difference between the male and female data emerged only when they were considered jointly for the entire chromosome, by using each data set to build a linkage map (Figure 2, B and C). This result agrees with those of DONIS-KELLER *et al.* (1987) for multiple human linkage groups. We speculate that every mouse autosome, when analyzed as a whole, will show a reduced male meiotic length.

Interference effects that were insignificant interval by interval were revealed when the linkage group was considered as a whole. Indeed, interference appears nearly complete: measured recombination fractions (Table 2) equal map distances (Figure 2A) for all intervals in which linkage could be detected. This degree of interference is greater than that assumed in the KOSAMBI or CARTER-FALCONER mapping functions, which are widely used for the analysis of mouse linkage data (ROBINSON, 1972, pp. 243–248). The generality of this deviation is unclear, however: one chromosome has been examined in one backcross and most of the pairwise recombination fractions are small.

The linkage relationships of *Fos*, *Aat-1* and *Igh* have been examined previously with inconsistent results. Somatic genetic analysis of *12;15* translocation chromosomes derived from mouse plasmacytomas supported the gene order Cen–(*Fos*, *Aat-1*)–*Igh-C–Igh-V* (ERIKSON *et al.* 1985; ERIKSON, MUSHINSKI and CROCE 1986; WIRSCHUBSKY *et al.* 1985). Analysis of a three-point cross informative for *Igh*, *Aat-1* and the *T(5; 12)31H* translocation breakpoint suggested the order Cen–*Igh–Aat-1–T31H* (MEO *et al.* 1980). Both analyses require the assumption that translocation chromosomes have undergone no secondary rearrangements, and the number of informative animals typed in the three-point cross was very small.

The backcross analyzed here indicates a distal location for the *Igh* gene complex. Two crucial issues remain open, however. First, additional typing of these and other backcross mice will be needed to

TABLE 5

Inheritance of RFLP marikers of chromosome *12* in RI strains of mice[a]

Marker	AKXD strain																												
	1	2	3	6	7	8	9	10	11	12	13	14	15	16	17	18	20	21	22	23	24	25	26	27	28
D12Nyu10	D	D	D	D	D	D	A	D	D	A	A	A	A	D	–	D	D	A	A	A	D	D	–	–	–
Mtv-9	D	A	D	A	A	A	D	A	D	A	D	A	A	D	D	A	A	D	A	A	A	D	D	A	A
D12Nyu9	D	A	D	A	D	D	A	A	A	A	A	A	A	D	A	D	D	D	A	A	A	D	D	A	A
Igh-Ca	D	A	D	A	D	D	A	A	A	A	A	A	A	D	A	D	D	D	A	A	A	D	D	A	A
	AKXL strain					1	1	1	1	1	1	2	2	2	2	2	3	3											
	5	6	7	8	9	2	3	4	6	7	9	1	4	5	8	9	7	8											
D12Nyu1	A	L	L	L	A	A	L	A	A	A	L	L	A	L	A	A	L	L											
D12Nyu9	A	L	L	A	A	A	L	L	L	A	L	L	A	A	A	L	A	A											
	BXD strain						1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	3	3	3				
	1	2	5	6	8	9	1	2	3	4	5	6	8	9	0	1	2	3	4	5	7	8	9	0	1	2			
D12Nyu10	B	D	B	B	D	D	D	D	B	D	B	B	B	B	D	D	D	D	D	B	D	B	B	B	B	D			
D12Nyu7	B	D	B	B	D	D	D	D	B	D	B	B	B	B	–	–	D	D	D	B	D	B	B	B	B	D			
	BXH strain							1	1	1	1	1																	
	2	3	4	6	7	8	9	0	1	2	4	9																	
D12Nyu5	B	B	B	B	H	B	H	B	H	H	H	H																	
D12Nyu9	H	B	H	H	H	B	H	B	B	H	B	H																	
	CXB strain																												
	D	E	G	H	I	J	K	N	O	P	Q	R																	
D12Nyu7	B	C	C	C	C	B	C	–	–	–	–	–																	
D12Nyu2	B	C	C	C	C	B	B	B	B	C	B	B																	
D12Nyu5	B	C	C	C	C	B	C	–	–	–	–	–																	
D12Nyu1	C	C	C	C	C	B	C	B	B	C	C	C																	
Aat-1	B	B	C	C	B	C	C	C	C	C	C	B																	
Igh-Ca	B	B	C	B	B	C	B	–	C	C	–	B																	
	CXJ strain					1	1	1					SWXL strain													BXJ strain		
	1	3	6	8	9	0	1	5								1	1	1			1	1				1	2		
													4	7	2	4	5		6	7									
D12Nyu7	C	C	C	J	J	C	J	C					*	*	*	*	*		*	*						*	*		
D12Nyu6	C	C	J	C	J	J	J	C					S	S	S	L	S		S	L						J	B		
Mtv-9	C	C	C	C	J	J	J	J					L	L	L	S	L		S	L						B	B		
D12Nyu8	C	C	C	C	J	J	J	J					L	L	L	L	S		S	L						B	B		
	OXA strain									1	1	1	1	1															
	1	2	3	4	5	6	7	8	9	0	1	2	3	4															
D12Nyu2	O	A	O	A	A	O	O	O	A	O	O	O	A	O															
D12Nyu1	O	A	O	A	O	O	A	O	A	O	O	O	A	O															
D12Nyu4	O	O	A	O	O	O	O	O	A	O	O	A	O	O															
Igh-Ca	O	A	A	O	O	O	O	A	A	O	A	A	A	O															
	CXS strain									1	1	1	1	1															
	1	2	3	4	5	6	7	8	9	0	1	2	3	4															
D12Nyu7	C	C	T	C	C	C	C	C	C	T	C	T	T	C															
Aat-1	C	T	C	T	C	C	C	T	T	T	C	T	T	C															

[a] Spleen or liver genomic DNA from RI strains of mice was typed for RFLPs by Southern blotting. Every RI strain was homozygous for one of the progenitor alleles of each marker as indicated by the letters: A, AKR/J-like; B, C57BL/6J-like; C, BALB/cJ-like; D, DBA/2J-like; H, C3H/HeJ-like; J, SJL/J-like; L, C57L/J-like; O, O20-like; S, SWR/J-like; T, STS/A-like. *, not informative; –, not determined.

determine the order of V and C genes within the complex. Second, the constancy of this map from strain to strain of mouse must be tested. Several subspecies of wild mice have probably contributed to the gene pool from which laboratory strains of mice derive (BLANK, CAMPBELL and D'EUSTACHIO, 1986). In the case of chromosome *12*, the C57BL/6J-specific alleles of both the proximal *D12Nyu2* marker and the distal *Igh* gene complex appear to derive from *Mus molossinus*, for example (POTTER 1978; P. D'EUSTACHIO, unpublished data). At a molecular level, the *Igh-V* gene complexes of BALB/c and C57BL/6 mice show numerous differences in gene number and organization (*e.g.*, BOTHWELL 1984). It thus seems plausible to speculate that the map of chromosome *12* (and probably of all chromosomes) will differ in detail from strain to strain of mouse. Further breeding experiments will be needed to test this notion. The availa-

bility of multilocus map building procedures such as BAYLOC should facilitate detection of such differences and further analysis of the biological consequences of alterations in gene organization.

We thank S. COPPERSMITH, M. DAVISSON, E. EICHER, E. GREEN, R. KARESS, H. L. KLEIN, L. SCHULMAN and P. R. SMITH for numerous helpful discussions, and J. HILGERS, G. PETERS and J. STAVNEZER for gifts of material. We thank P. R. SMITH as well for providing access to the NYU Cell Biology Department VAX 11/750 computer. This work was supported by Grant GM32105 from the National Institutes of Health and by research grant 1-1018 from the National Foundation/March of Dimes. R.B. was supported in part by an award from the National Institutes of Health Medical Scientist Training Program.

LITERATURE CITED

BISHOP, D., 1985 The information content of phase-known matings for ordering genetic loci. Genet. Epidemiol. **2:** 349–361.

BLANK, R. D., G. R. CAMPBELL and P. D'EUSTACHIO, 1986 Possible derivation of the laboratory mouse genome from multiple wild *Mus* species. Genetics **114:** 1257–1269.

BOTHWELL, A. L. M., 1984 The genes encoding anti-NP antibodies in inbred strains of mice. pp. 19–34. In: *The Biology of Idiotypes*, Edited by M. I. GREENE and A. NISONOFF. Plenum, New York.

COBB, R. R., T. A. STOMING and J. B. WHITNEY, 1987 The aryl hydrocarbon hydroxylase (*Ah*) locus and a novel restriction fragment length polymorphism (RFLP) are located on mouse chromosome 12. Biochem. Genet. **25:** 401–413.

DAVISSON, M. T., and T. H. RODERICK, 1981 Recombination percentages. pp. 283–313. In: *Genetic Variants and Strains of the Laboratory Mouse*, Edited by M. C. GREEN. Gustav Fischer Verlag, Stuttgart.

D'EUSTACHIO, P., 1984 A genetic map of mouse chromosome 12 composed of polymorphic DNA fragments. J. Exp. Med. **160:** 827–838.

D'EUSTACHIO, P., S. JADIDI, R. C. FUHLBRIGGE, P. W. GRAY and D. D. CHAPLIN, 1987 Interleukin-1 α and β genes: linkage on chromosome 2 in the mouse. Immunogenetics **26:** 339–343.

DONIS-KELLER, H., P. GREEN, C. HELMS, S. CARTINHOUR, B. WEIFFENBACH, K. STEPHENS, T. P. KEITH, D. W. BOWDEN, D. R. SMITH, E. S. LANDER, D. BOTSTEIN, G. AKOTS, K. S. REDIKER, T. GRAVIUS, V. A. BROWN, M. B. RISING, C. PARKER, J. A. POWERS, D. E. WATT, E. R. KAUFFMAN, A. BRICKER, P. PHIPPS, H. MULLER-KAHLE, T. R. FULTON, S. NG, J. W. SCHUMM, J. C. BRAMAN, R. G. KNOWLTON, D. F. BARKER, S. M. CROOKS, S. E. LINCOLN, M. J. DALY and J. ABRAHAMSON, 1987 A genetic linkage map of the human genome. Cell **51:** 319–337.

DUNN, L. C., and D. BENNETT, 1967 Sex differences in recombination of linked genes in animals. Genet. Res. **9:** 211–220.

ERIKSON, J., D. A. MILLER, O. J. MILLER, P. W. ABCARIAN, R. M. SKURLA, J. F. MUSHINSKI and C. M. CROCE, 1985 The *c-myc* oncogene is translocated to the involved chromosome 12 in mouse plasmacytoma. Proc. Natl. Acad. Sci. USA **82:** 4212–4216.

ERIKSON, J., J. F. MUSHINSKI and C. M. CROCE, 1986 The locus for the serum prealbumin is proximal to the heavy chain locus on mouse chromosome $12q^+$. J. Immunol. **136:** 3137–3139.

GARDNER, M., S. DANDEKAR and R. CARDIFF, 1986 Molecular mechanism of an ecotropic MuLV restriction gene Akvr-1/FV-4 in California wild mice. Contemp. Top. Microbiol. Immunol. **127:** 338–345.

GREEN, E. L., 1981 *Genetics and Probability in Animal Breeding Experiments.* Oxford University Press, New York.

HILGERS, J., and J. ARENDS, 1985 A new series of recombinant inbred strains between the BALB/cHeA and STS/A mouse strains. Contemp. Top. Microbiol. Immunol. **122:** 31–37.

HILL, R. E., AND N. D. Hastie, 1987 Accelerated evolution in the reactive centre regions of serine protease inhibitors. Nature **326:** 96–99.

IKEDA, H., H. SATO and T. OKADA, 1981 Mapping of the *Fv-4* mouse gene controlling resistance to murine leukemia viruses. Int. J. Cancer **28:** 237–240.

LATHROP, G. M., J. M. LALOUEL, C. JULIER and J. OTT, 1984 Strategies for multilocus linkage analysis in humans. Proc. Natl. Acad. Sci. USA **81:** 3443–3446.

LUSIS, A. J., B. A. TAYLOR, D. QUON, S. ZOLLMAN and R. C. LEBOEUF, 1987 Genetic factors controlling structure and expression of apolipoproteins B and E in mice. J. Biol. Chem. **262:** 7594–7602.

MEO, T., J. JOHNSON, C. V. BEECHEY, S. J. ANDREWS, J. PETERS and A. G. SEARLE, 1980 Linkage analyses of murine immunoglobulin heavy chain and serum prealbumin genes establish their location on chromosome 12 proximal to the *T(5;12)31H* breakpoint in band 12F1. Proc. Natl. Acad. Sci. USA **77:** 550–553.

OKADA, T., H. IKEDA, H. YOSHIKURA, K. MORIWAKI and S. SUZUKI, 1981 *Fv-4*: gene controlling resistance to NB-tropic Friend murine leukemia virus. Distribution in wild mice, introduction into genetic background of BALB/c mice, and mapping of chromosomes. JNCI **67:** 1123–1127.

OWEN, F. L., R. RIBLET and B. A. TAYLOR, 1981 The T suppressor cell alloantigen Tsu^d maps near immunoglobulin allotype genes and may be a heavy chain constant-region marker on a T cell receptor. J. Exp. Med. **153:** 801–810.

PETERS, G., M. PLACZEK, S. BROOKES, C. KOZAK, R. SMITH and C. DICKSON, 1986 Characterization, chromosome assignment, and segregation analysis of endogenous proviral units of mouse mammary tumor virus. J. Virol. **59:** 535–544.

POTTER, M., 1978 Comments on the relationship of inbred strains to the genus *Mus*. pp. 497–509. In: *Origins of Inbred Mice*, Edited by H. C. MORSE III. Academic Press, New York.

ROBINSON, R., 1972 *Gene Mapping in Laboratory Mammals*, Part B. Plenum Press, New York.

SILVER, J., and C. E. BUCKLER, 1986 Statistical considerations for linkage analysis using recombinant inbred strains and backcrosses. Proc. Natl. Acad. Sci. USA **83:** 1423–1427.

SNEDECOR, G. W., and W. G. COCHRAN, 1967 *Statistical Methods*, Ed. 6. Iowa State University Press, Ames, Iowa.

TAYLOR, B. A., 1978 Recombinant inbred strains: use in gene mapping. pp. 423–428. In: *Origins of Inbred Mice*, Edited by H. C. MORSE III. Academic Press, New York.

TAYLOR, B. A., 1981 Recombinant inbred strains. pp. 397–407. In: *Genetic Variants and Strains of the Laboratory Mouse*, Edited by M. C. GREEN. Gustav Fischer Verlag, Stuttgart.

TAYLOR, B. A., D. W. BAILEY, M. CHERRY, R. RIBLET and M. WEIGERT, 1975 Genes for immunoglobulin heavy chain and serum prealbumin protein are linked in the mouse. Nature **256:** 644–646.

TRAINA, V. L., B. A. TAYLOR and J. C. COHEN, 1981 Genetic mapping of endogenous mouse mammary tumor viruses: locus characterization, segregation, and chromosomal distribution. J. Virol. **40:** 735–744.

VON DEIMLING, O. H., and B. A. TAYLOR, 1987 Esterase-25 (*Es-25*): identification and characterization of a new kidney arylesterase of the house mouse, genetically linked to *Ly-18* on chromosome 12. Biochem. Genet. **25:** 639–646.

WIRSCHUBSKY, Z., S. INGVARSSON, A. CARSTENSSEN, F. WEINER, G. KLEIN and J. SÜMEGI, 1985 Gene localization on sorted chromosomes: definitive evidence on the relative positioning of genes participating in the mouse plasmacytoma-associated typical translocation. Proc. Natl. Acad. Sci. USA **82:** 6975–6979.

Communicating editor: N. A. JENKINS

APPENDIX

Construction of genetic maps using bayesian location scores

The experimental data underlying a linkage map are recombination fractions measured between pairs of genetic markers in breeding experiments. These fractions are binomially distributed variables and can thus be used to estimate map distances between pairs of markers and confidence intervals for those distances (GREEN 1981). Any two markers are far more likely *a priori* to be unlinked than linked, so a Bayesian approach to the calculation of confidence intervals associated with recombination fractions is appropriate.

Multilocus approaches to the analysis of linkage data provide a powerful approach for extracting map information from limited amounts of genetic data. Here, we describe an algorithm which incorporates Bayesian statistical analysis (SILVER and BUCKLER 1986) in the calculation of location scores for backcross and RI data. The algorithm allows the assignment of explicit probabilities to arbitrarily small intervals on a chromosome, direct summation of confidence intervals surrounding the best location for a test marker, and efficient assessment of alternative orders for the loci.

Theory: In a breeding experiment, inheritance of a test locus is scored with respect to one or more marker loci. Map positions are assumed for the marker loci; that of the test locus is unknown. For each marker j, the experimental data are the number of recombinants found between the test locus and the marker, R_j, and the number of individuals scored for that pair of loci, N_j.

Bayes theorem can then be applied to the analysis of this experiment:

$$P(L|D) = P(D|L)P(L)/P(D). \tag{1}$$

Here, L = location, D = data, $L|D$ = location conditional on data, $D|L$ = data conditional on location, and $P(X)$ = probability of X. The data sets D are the marker loci, their assumed map positions, and the numbers of informative chromosomes scored and recombinants found. That is, $P(L|D)$, the probability that the test locus resides at some location L given the data and hypothesis as to marker positions, can be expressed in terms of $P(D)$, $P(D|L)$, and $P(L)$. Expressions are needed for each of these three probabilities.

First, the genome may be considered to be composed of G locations of equal genetic length and probability $1/G$. The quantity $P(L)$ is then a constant, $1/G$. That is, in the absence of any data, each $1/G$ interval of the genome is equally probably the location of a test marker.

Next, for any location L_i, $P(D|L_i)$ is calculated using an algorithm based on that of LATHROP *et al.* (1984). The markers are placed at assumed positions and a chromosome is defined, extending from 50 cM to the left of the leftmost marker to 50 cM to the right of the rightmost marker. For each $1/G$ interval i, the map distance separating L_i from each marker j is calculated. The likelihood, $Y_{i,j}$, of observing the actual number of recombinant (R) and total individuals (N) is calculated using the binomial distribution and assuming that recombinant fraction is equal to map distance in cM:

$$Y_{i,j} = \binom{N}{R} p^R q^{(N-R)}, \tag{2}$$

where $\binom{N}{R}$ is the binomial coefficient, p is the map distance separating L_i and marker j divided by 100 and q is 1-p. All locations L_i more than 50 cM from a marker j, including unlinked markers, are assumed to recombine at 50% with regard to the marker j, *i.e.*, $p = q = 0.5$.

Mapping functions such as those of HALDANE or KOSAMBI can be incorporated at this point, using their inverses (GREEN 1981) to calculate transformed map distances corresponding to the observed recombinant fractions. For recombination data from sets of recombinant inbred strains of mice, the mapping function $p = R/(4N - 6R)$ (TAYLOR 1978) is used. Values of p and q corresponding to the transformed distances are used in the calculation of $Y_{i,j}$.

$Y_{i,j}$ is calculated for each marker j and the joint probability of the scoring data for all markers assuming residence of the test locus at L_i is given by:

$$P(D|L_i) = \prod_{j=1}^{\text{all markers}} Y_{i,j}. \tag{3}$$

No order for the test locus relative to the markers is assumed in calculating the joint probability for the observed recombination data. Different numbers of individuals may be scored for recombination between that locus and each of the markers j. That is, N_j need not be constant for different j's. The joint likelihood can be interpreted as a probability conditional on the assumption of specific map positions L_i.

Third, we find an expression for $P(D)$. For a genome of G equal intervals, $P(D)$ is the sum of all the $P(D|L_i)$ terms divided by G:

$$P(D) = \sum_{i=1}^{G} P(D|L_i)P(L_i). \tag{4}$$

Substituting each of these expressions back into Equation 1 we can calculate the probability of any specific location L_x conditional on the data by

$$P(L_x|D) = \frac{1/G \prod_{j=1}^{\text{total markers}} Y_{x,j}}{1/G \sum_{i=1}^{G} \prod_{j=1}^{\text{total markers}} Y_{i,j}}. \tag{5}$$

Equation 5 expresses the probability of obtaining particular numbers of recombinants between the test locus and each marker if the test locus resides "here" for each "here" in the genome. We call this probability the BOD score, for Bayesian ODds score. The sum of the $P(L_i|D)$ values for all positions is 1.

Marker loci are linked, however, at recombination fractions specified in the hypothesis, reducing the new information contributed by each successive marker to $P(L_x|D)$. The terms for successive $Y_{x,j}$ terms must therefore be reduced by some factor to correct for this interdependence and loss of information. The identical factor applies to each corresponding $Y_{i,j}$ term, however, and the correction factors thus cancel out in Equation 5. There is a crucial exception: if two successive markers are indistinguishable by recombination, then by this reasoning $Y_{x,j}$ and $Y_{i,j}$ both take on 0 values for the second marker and Equation 5 is undefined.

The $1/G$ interval with the highest BOD score is taken to be the position of the test locus. BOD scores are explicit probabilities, so a confidence interval about this position is calculated by summing BOD values for adjoining $1/G$ intervals. If the sum of BOD scores over the whole interval bounded by the adjacent markers exceeds the confidence criterion, the order of the test locus in the map may be inferred to be established at that confidence level. Failure to exceed the criterion demonstrates ambiguity as to order.

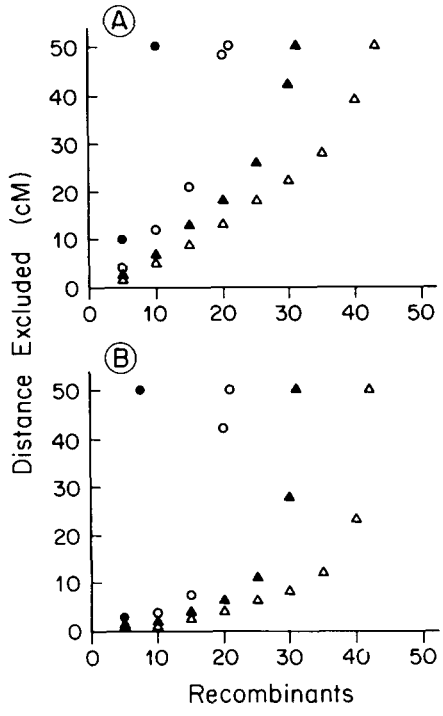

FIGURE 3.—Map distances excluded at 95% confidence as a function of numbers of recombinants found. Excluded distances are shown for 25 (●), 50 (○), 75 (▲), and 100 (△) total informative chromosomes scored in a backcross (A) or in RI strains (B).

The curve of BOD scores parallels that of LATHROP *et al.*'s (1984) location scores. $P(D|L)$ is a likelihood, as discussed above, and is calculated in the same manner as the location score. For any data set, $P(L_i)/P(D)$ is a constant whose value is determined by genome size, interval size, and the nature of the data. In general, the larger the number of informative individuals, the smaller the value of this constant. The use of Bayes' theorem in BAYLOC thus normalizes the BOD score distribution over the entire genome to a sum of 1 while preserving the shape of the likelihood curve generated by the location score.

Computation: These calculations are carried out by the Fortran program BAYLOC. BAYLOC first accepts the number of marker loci, their positions, and the numbers of informative individuals scored and recombinant between each marker and the test locus. Next, it calculates the size of the 'chromosome' tested by assigning the left end to the location 50 cM to the left of the leftmost marker and the right end to the location 50 cM to the right of the rightmost marker. For the test locus and each marker locus, the likelihood of observing the actual number of recombinant mice in the number of individuals scored is calculated for every map distance from 0 to 50 cM. If a mapping function has been invoked, the map distances are recombinant fractions as calculated by the appropriate formulas. For each position on the chromosome, the map distance separating it from each marker locus is calculated, and the appropriate likelihoods retrieved from memory and multiplied. This product is multiplied by $1/G$ to give the numerator of BOD for that position. The numerators are summed over all the positions on the chromosome and the value is recorded as P(Data | Test Locus Linked). The likelihood product for all markers at a recombination fraction of 50% is calculated and multiplied by (G − chromosome size)/G, giving P(Data | Test Locus Unlinked). The sum of P(Data | Test Locus Linked) and P(Data | Test Locus Unlinked) gives the denominator of BOD. BOD values are calculated for each position on the chromosome and the BOD scores are summed to give P(Test Locus Linked | Data).

Using the BAYLOC algorithm to calculate maximum numbers of recombinants consistent with the hypothesis of linkage for various total numbers of informative chromosomes yielded values in agreement with those of SILVER and BUCKLER (1986) using a Bayesian approach but a different computational algorithm.

Excluding linkage between two loci at a given confidence level is often of biological interest. BOD scores can be used to calculate intervals excluded at a given confidence level for given numbers of informative individuals and recombinants in a backcross (Figure 3A) or an RI strain survey (Figure 3B).

BOD probability values are insensitive to variations in genome size between 800 and 4000 cM. For genomes of less than ~400 cM the map distances spanned by a given number of individuals tested for a pair of loci increases markedly with decreasing genome size. This result has two consequences. First, when there is no prior information concerning location of a test locus, Bayesian probability of linkage calculations will not be substantially affected by even a twofold error in estimating genome size. Second, when prior data from, *e.g.*, *in situ* hybridization experiments or analysis of somatic cell hybrids allow the effective size of the genome to be reduced below ~400 cM, the number of animals needed to reach a conclusion at a given confidence limit is significantly reduced.