Am. J. Hum. Genet. 41:189–201, 1987

Multipoint Gene Mapping Using Seriation. II. Analysis of Simulated and Empirical Data

KENNETH H. BUETOW AND ARAVINDA CHAKRAVARTI

Department of Biostatistics, Human Genetics Program, University of Pittsburgh, Pittsburgh

SUMMARY

Seriation methods provide an accurate and efficient means of constructing preliminary multilocus genetic maps. By using both simulated and previously published empirical data, multipoint mapping by seriation was critically evaluated. Analysis of the simulated data sets showed that the seriation methodology could accurately estimate order and interlocus distances. Application to the empirical data demonstrated that seriation could obtain results directly comparable with those of other multipoint mapping methods. Techniques such as seriation can produce preliminary genetic maps that may be used as starting points for more computer-intensive maximum-likelihood multipoint techniques.

INTRODUCTION

The rapid accumulation of data on DNA polymorphisms at a large number of loci has made it feasible to construct multipoint linkage maps of each human chromosome. The volume and diversity of sources of these data require that efficient means of analysis and summary be used in map construction and communication of results. In the preceding, companion paper (Buetow and Chakravarti 1987), a new method called seriation is proposed as one solution to this problem.

Currently, most linkage analyses are concerned with marker loci known to be located on a particular chromosome. Therefore, the primary objective of multipoint analysis is to obtain locus order and to estimate interlocus map

Received August 26, 1986; revision received February 23, 1987.

Address for correspondence and reprints: Dr. K. H. Buetow, Fox Chase Cancer Center, 7701 Burholme Avenue, Philadelphia, PA 19111.

© 1987 by the American Society of Human Genetics. All rights reserved. 0002-9297/87/4102-0010$2.00

distances. The preceding paper (Buetow and Chakravarti 1987) describes how seriation, together with least-squares techniques, can achieve these goals. In the present paper we evaluate, using both simulated and empirical data, the efficiency and utility of seriation for constructing multipoint linkage maps. We consider the following four different data sets: (1) Monte Carlo simulation of five- and 10-locus linkage groups, with simulation results given as estimated recombination values between all locus pairs, (2) family data on a four-locus linkage group simulated for the Genetic Analysis Workshop III (GAW III) (MacCluer et al. 1985), (3) published pairwise lod-score data for eight loci on human chromosome 1 (Robson and King 1984; Povey et al. 1985), and (4) family data on five markers on the short arm of chromosome 11. The analyses of these data allowed us to evaluate not only how seriation performed as a multipoint method but also how seriation performed in comparison with other multilocus methods applied to these data sets.

MATERIAL AND METHODS

Monte Carlo Simulation Data

Matrices consisting of recombination values were generated by means of Monte Carlo techniques. In these simulations, each value in the matrix was taken to be analogous to a recombination value between a pair of loci in a linkage group estimated from a set of families independent of those used for any other pair. Thus, each entry in the matrix could be simulated by drawing independent samples from a binomial distribution with parameters p (the recombination value for that locus pair) and n (the number of informative meioses for the sample). The following three maps were used for the macro simulations: (1) five-locus uniform map distance (5U), A-10-B-10-C-10-D-10-E; (2) five-locus nonuniform map distance (5N), A-5-B-1-C-9-D-16-E; and (3) 10-locus uniform map distance (10U), A-10-B-10-C-10-D-10-E-10-F-10-G-10-H-10-I-10-J, where A, B, C, . . . represent loci in the map and the numerical values between loci are interlocus map distances in centiMorgans (cM). The pairwise recombination values used as input for the simulations were obtained by transforming the map distance between any two loci by means of Haldane's mapping function (Haldane 1919). The recombination values so obtained for each map are presented in table 1.

In the first set of experiments with the Monte Carlo simulation data, we assessed the ability of seriation to recover the order of the above three maps. This ability was tested using 100 replicated sets of recombination matrices with samples of 20, 40, 60, 80, and 100 informative meioses. Seriation was used to derive a locus order for each replicate. The frequency with which a particular order could be obtained and the number of correct orders were calculated.

Next, we evaluated how well interference could be measured, by determining the efficiency with which Rao et al.'s (1977) mapping parameter, p, could be estimated from the simulated data by using the least squares method. For this analysis, 100 replicate sets of recombination matrices were generated, using the 5U map, at five different sample sizes (100, 200, 300, 400, and 500 informative meioses). For a given sample size, each of the replicate matrices was ordered

TABLE 1

MATRICES OF RECOMBINATION VALUES BETWEEN LOCUS PAIRS USED AS INPUT FOR DETERMINISTIC SIMULATION

A. 5U AND 10U

	B	C	D	E	F	G	H	I	J
A...	.091	.165	.226	.275	.316	.349	.377	.399	.417
B...		.091	.165	.226	.275	.316	.349	.377	.399
C...			.091	.165	.226	.275	.316	.349	.377
D...				.091	.165	.226	.275	.316	.349
E...					.091	.165	.226	.275	.316
F...						.091	.165	.226	.275
G...							.091	.165	.226
H...								.091	.165
I ...									.091

B. 5N

	B	C	D	E
A...	.048	.057	.130	.231
B...		.010	.091	.203
C...			.082	.197
D...				.137

according to the seriation algorithm; then map distances and p values were estimated by means of the least-squares procedure (Buetow and Chakravarti 1987). The results were evaluated by calculating the mean, mode, and mean square error of the estimated p value at each of the sample sizes. We examined how these results may be affected by interference levels higher than that assumed by the Haldane function. The recombination fractions were generated from the 5U map by using the Kosambi mapping function (Kosambi 1944), obtained by assuming $p = 0.5$ in Rao et al.'s (1977) generalized mapping function, and the above analysis was repeated.

GAW III Data

Simulated data, consisting of 150 pedigrees segregating for two four-locus linkage groups that were generated for GAW III (see MacCluer et al. 1985), were analyzed using the seriation methodology. The analysis was restricted to the linkage group consisting of the loci A, D, E, and G, since the second group was simulated with linkage disequilibrium. The linkage group used in the analysis was simulated under the assumptions of a 2:1 female:male recombination ratio and Kosambi-level interference ($p = 0.5$).

Pairwise recombination values were estimated for each of the six interlocus intervals by using the computer program LIPED (Ott 1974). Bivariate lod tables, $Z(\theta_M, \theta_F)$, were calculated for each family at the 64 grid points obtained by considering all pairwise combinations of the values 0, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, and 0.5. After the lod tables over all 150 pedigrees had been summed,

estimates of the recombination values and their SEs were obtained separately for males and females by means of quadratic interpolation that used a modified version of the program QUAD (Hodge et al. 1983). These estimates (data available on request) were then assembled into male and female matrices from which orders were determined by seriation and interlocus distances estimated using the least-squares method, as described in the preceding paper (Buetow and Chakravarti 1987).

Chromosome 1 Data

Published pairwise lod-score data for chromosome 1 markers from the Galton laboratory (Robson and King 1984) and from the Eighth International Human Gene Mapping Workshop (Povey et al. 1985) were used to construct male and female recombination-value matrices for five- and eight-locus linkage groups. The five-locus group contained the short-arm markers PGD, Rh, UMPK, Sc, and PGM, whereas the eight-locus group contained the markers PGD, FUCA, Rh, UMPK, PGM1, AMY, 1qh, and Fy. A complete description of each of the markers and its physical location has been presented by Povey et al. (1985).

The above subsets were chosen from the published data on 16 loci because they represented the largest set of loci for which all pairwise combinations were available. The five-locus subset was considered because of its similarity (as determined on the basis of the map of this region published by Sherman et al. [1984]) to our simulated five-locus maps (5U and 5N), thereby allowing an approximate evaluation of significance. An approximate sample size was estimated for this analysis by using Edwards's (1971) method for converting variances estimated from lod-score analysis into "equivalent number of meioses." Recombination-value estimates for the five- and eight-locus groups were then assembled into male and female matrices from which orders were determined by seriation and interlocus distances estimated by least squares.

Chromosome 11p Data

A map for five loci (listed in table 2) located on the short arm of chromosome 11 was constructed by using pedigree data on 29 nuclear families provided courtesy of Dr. H. H. Kazazian, Jr. (Johns Hopkins Hospital). These five loci (described in detail in Grzeschik and Kazazian 1985) have previously been mapped (Fearon et al. 1984; Kittur et al. 1985) using these same pedigrees but employing a different multipoint methodology. Univariate (male = female) recombination values were estimated for the 10 locus pairs from the 29 families by using the computer programs MLINK and ILINK (Lathrop et al. 1984, 1985). These estimates, together with their associated lod scores, are presented in table 2. With use of the variances of the recombination-value estimates as calculated by ILINK, the equivalent number of meioses (Edwards 1971) for these data, averaged over all locus pairs, was estimated. The recombination-value estimates were then assembled into a matrix from which orders were determined by seriation and interlocus distances estimated by least squares.

TABLE 2

PAIRWISE UNIVARIATE RECOMBINATION VALUES FOR A FIVE-LOCUS LINKAGE GROUP ON CHROMOSOME 11p

	CAL	PTH	HBBC	INS	HRAS1
CAL.....		8.0	11.5	15.0	25.3
PTH.....	3.26		11.9	23.2	16.5
HBBC.....	3.52	4.84		15.2	9.3
INS.....	2.46	1.06	6.99		5.4
HRAS1.....	0.47	2.39	7.27	8.68	

NOTE.—Values above the diagonal are recombination fractions (in %); values below the diagonal are peak lod scores. A complete description of the loci has been given by Grzeschik and Kazazian (1985).

RESULTS

Monte Carlo Simulation Data

As explained in the companion paper (Buetow and Chakravarti 1987), when the pairwise recombination values cannot be reconciled with any map, seriation may not produce a locus order. Results of seriation of the simulated data, which are presented in table 3, show that this is a rare phenomenon and that the seriation algorithm can recover an order virtually 100% of the time even with very few (20–40) informative meioses (note that 20–40 meioses are designated as a "small sample size" because, while sufficient for determining linkage when θ is small, the sampling variance of the pairwise θ estimates is large). With these small sample sizes, however, the proportion of correct orders is not very high (30%–63%). The results are more satisfactory when sample sizes are larger. For 100-meioses samples of the 5U, 5N, and 10U maps, the algorithm recovered the correct order 93%, 54%, and 80% of the time, respectively.

Next, the cases in which the algorithm obtained an incorrect order were investigated. First, it was determined, by counting the number of "inversions"

TABLE 3

ORDERS DERIVED FROM SERIATION OF RECOMBINATION-VALUE MATRICES GENERATED FROM THREE MAPS BY DETERMINISTIC SIMULATION

No. of Informative Meioses	No. of Orders Derived/ No. of Orders Correct		
	5U	10U	5N
20.....	97/30	96/7	89/19
40.....	100/63	99/30	98/42
60.....	100/71	100/40	100/43
80.....	100/89	100/65	100/51
100.....	100/93	100/80	100/54

that occurred in the incorrect order, how an incorrect order deviated from the true order. Inversions were scored by assigning each locus its rank in the true order. Then, with respect to each locus in the derived order, a count of the number of loci with a lower rank and located to its right was made. These counts were then summed over all loci to obtain the total number of inversions. The results of this analysis are presented in table 4 and demonstrate that an incorrect order was usually different from the correct order by only a single inversion. This means that the order was inaccurate by the reversal of two adjacent loci. For the 5U and 10U maps, as expected, there was no order preferentially represented among the inversions given the uniform distance between loci. This was not the case for the 5N map, in which the order ACBDE, which has the inversion of the two loci separated by 1 cM (B-C), accounted for 67% (20 meioses) to 100% (100 meioses) of the single-inversion orders.

If one then redefines a "correct" order as one containing 0 or 1 inversions, at moderate sample sizes (60 meioses), 85% of the 5U and 10U maps will be correct. The 5N map approaches this level at 100 meioses (81%). With a sample size of 100 meioses, virtually 100% of the orders have 0 or 1 inversions in the 5U and 10U maps.

To evaluate whether the true order provided a better fit than the order obtained by the algorithm, the continuity index (CI), equation (1) in the companion paper (Buetow and Chakravarti 1987), was calculated for the true order and then compared with the value obtained for the derived order. If the CI was lower, it was judged that the algorithm had failed to achieve the best order for the observed data.

The results of this analysis (presented in table 5) for the five-locus maps showed such failures to be rare. The number of failures among the total number of orders obtained was well below 5% when moderate to large sample sizes (60–100 meioses) were used. The results of the 10-locus map were less encouraging, in that, in the same sample-size range (60–100 meioses), the algorithm failed, on average, 29% of the time. With 100 meioses the situation was slightly better, with only a 13% failure rate.

The second set of experiments, the results of which are presented in table 6, evaluated the estimation of the mapping parameter. These results showed that with sample sizes of 100 meioses, estimates of the parameter $p = 1.0$ fell within the interval (.9–1.0) 41% of the time. The empirical 95% confidence interval of p was 0.0–1.0. Accuracy improved as a function of increasing sample size, since the mean moved closer to the simulated p value, as evidenced by the steady decrease in the mean square error. Unfortunately, even with very large sample sizes (500 meioses), only 54% of the estimated values fell within the 0.9–1.0 range.

The results obtained when Kosambi-level interference ($p = 0.5$) was assumed are comparable with those obtained above. The mode of each of the distributions was $p = 0.5$, but the proportion of correct estimates (values in the range 0.45–0.55) was 13%–25%. At a sample size of 500 meioses, only 25% of the estimated p values were correct. Interestingly, at the same sample sizes,

TABLE 4

DISTRIBUTION OF THE NUMBER OF INVERSIONS OBSERVED FOR THE THREE SIMULATED MAPS

No. of Inversions	No. of Meioses														
	5U					10U					5N				
	20	40	60	80	100	20	40	60	80	100	20	40	60	80	100
0.....	30	64	71	88	92	7	28	44	59	80	19	42	43	51	54
1.....	24	23	21	12	7	14	33	40	31	19	33 (22)	23 (21)	25 (24)	28 (26)	27 (27)
2.....	9	4	1			15	23	11	6	1	10	11	7	7	7
3.....	15	4	5		1	4	7	5	3		7	5	11	9	5
4.....	12	1	1			10	3				13	12	7	3	5
5.....	1	2				8	1				7	5	7	1	1
6–22.....						31	4		1						

NOTE.—Numbers in parentheses are number of inversions accounted for by the order A-C-B-D-E.

TABLE 5

FREQUENCY WITH WHICH THE CORRECT ORDER FIT THE OBSERVED DISTANCE MATRIX BETTER THAN DID THE ORDER DERIVED

No. of Meioses	No. of Correct Orders with Better Fit/ Total No. of Orders Derived		
	5U	10U	5N
20	5/91 (.06)	48/96 (.50)	6/89 (.07)
40	7/98 (.07)	43/99 (.43)	4/89 (.04)
60	4/99 (.04)	43/100 (.43)	3/100 (.03)
80	0/100 (.00)	32/100 (.32)	2/99 (.02)
100	1/100 (.01)	13/100 (.13)	3/99 (.03)

uniformly smaller mean square errors were observed when interference was high ($p = 0.5$) as compared with when it was absent ($p = 1.0$).

GAW III Data Analysis

Application of the seriation methodology to the GAW III data produced the correct order D-A-G-E. The mapping parameter was estimated to be $p = 0.48$ in females, which was very close to the simulated value $p = 0.5$. On the other hand, the male estimate of 1.0 was quite different from the true value of $p = 0.5$. The failure to obtain the correct estimate of p in males was not altogether unexpected, given the short map distance (11.1 cM) across the linkage group. At this distance there is virtually no difference between the map distances generated by the various mapping functions. The estimated map distances for each interval, presented in table 7, were comparable with the simulated distances and well within the range of those observed by other investigators at the workshop (MacCluer et al. 1985, tables II and III). The average absolute differ-

TABLE 6

STATISTICAL PROPERTIES OF THE p ESTIMATES

p[a] and Sample Size	Mean $\hat{p}$	Mode	Mean Square Error	% Correct[b]
$p = 1.0$:				
100	0.57	1.0	35.655	41
200	0.73	1.0	29.563	53
300	0.81	1.0	10.667	54
400	0.83	1.0	8.191	58
500	0.85	1.0	5.618	58
$p = 0.5$:				
100	0.45	.5	12.339	13
200	0.46	.5	10.095	17
300	0.52	.5	6.803	16
400	0.51	.5	6.118	16
500	0.50	.5	3.466	25

[a] Estimated from 100 replicates of the 5U map, using $p = 1.0$ and $p = 0.5$.
[b] (No. of values observed in the interval containing the simulation p/total no. of replicates) × 100.

TABLE 7

MAP-DISTANCE ESTIMATES FROM CURRENT STUDY AND GAW III (MACCLUER ET AL. 1985, TABLES II AND III)

SOURCE	MAP-DISTANCE ESTIMATES (cM)					
	Male			Female		
	D-A	A-G	G-E	D-A	A-G	G-E
Simulation	10.9	3.8	8.4	21.7	6.0	18.9
Present study	13.6	4.4	12.3	21.8	8.3	20.0
	(11.3)	(3.8)	(10.2)			
Donald et al. 1985	10.0	6.5	9.7	19.1	7.9	16.9
Fain and Goldgar 1985	10.4	2.6	13.9	28.2	5.4	21.1
Ott 1985	10.3	3.9	8.9	18.2	8.0	21.0
Risch 1985	11.3	4.2	9.3	16.9	7.6	20.5
Sherman and Morton 1985	9.8	4.0	8.8	17.4	7.9	19.4

NOTE.—Numbers in parentheses are distance estimates using $p = 0.48$, as estimated from the data for females.

ence between the simulated and estimated map distances were 3.5 cM and 7.2 cM for females and males, respectively. In females this difference was smaller than that obtained by other investigators. If the mapping parameter estimated from the female map had been used for the male map, the male average absolute difference would have been reduced to 2.2 cM, making the combined difference observed in both males and females smaller than that obtained by any other group.

Chromosome 1 Analysis

Seriation of the five-locus group produced, in both sexes, the order PGD-Rh-UMPK-Sc-PGM1. Estimation of the equivalent number of meioses (Edwards 1971) from all locus pairs showed that, if the Rh-PGM1 comparison was excluded as being an outlier (2,161 meioses), the average sample size in males was 122. On the basis of the results of the 5N simulation, the order selected by the CI would be correct (0 or 1 inversions) 81% of the time. On the other hand, seriation of the 5U map produced a correct order 99% of the time at a sample size of 100 meioses. This suggested that the derived order had an 81%–99% chance of being correct.

The order obtained in this analysis was compared with orders of this five-locus group obtained by other investigators. Wedd (1984), using a maximum-likelihood ordering technique, obtained the same order as the present study. The order that Sherman et al. (1984), using location scores, derived for these loci differed from the above order by a single inversion involving the loci Sc and UMPK.

Seriation of the eight-locus linkage group resulted in an order for the male data only. This was due to the large number of 50% recombination values (9 of 28) observed in the female matrix, which made it impossible to mutually order

TABLE 8

COMPARISON OF MAPS OBTAINED BY VARIOUS STUDIES OF FIVE CHROMOSOME 1 p-arm MARKERS

MAP	MAP DISTANCE (cM)					
	Present Study		Wedd 1984		Sherman et al. 1984	
	Male	Female	Male	Female	Male	Female
PGD	20.4	27.2	22.9	42.6	22.0	39.3
Rh	11.2	22.8	13.5	26.3	14.0	23.0
UMPK	6.6 (30.5)	23.0 (70.6)	4.1 (32.8)	11.2 (59.0)	2.7 (33.5)	4.8 (54.3)
Sc	12.7	34.8	15.2	21.5	16.8	26.5
PGM1						
Total map length	50.9	97.8 (1.92)	55.7	101.6 (1.82)	55.5	93.6 (1.69)
Mapping parameter	0.06	0.58	0.5 (fixed)		0.351 (fixed)	

NOTE.—Numbers in parentheses are male:female map-distance ratios.

several loci. As based on the male recombination values, the best eight-locus order was PGD-FUCA-Rh-UMPK-PGM1-AMY-1qh-Fy. Inspection of this order shows that the orientation of the loci that overlap with those used in the five-locus analysis (overlapping loci underlined) was the same. This order also shows good correspondence with those obtained by Sherman et al. (1984) and Wedd (1984). It differs from each by only a single inversion. In the case of Sherman et al. (1984), this inversion occurs between the tightly linked pair Rh-FUCA (θ = .01). Wedd (1984) considered this pair to be a single locus in his analysis. Comparison of the rest of Wedd's (1984) order to ours shows identity, with the exception of an inversion between Fy and 1qh. Both orders are compatible with the chromosomal localization of Fy to bands 1p21–1q23 (Povey et al. 1985).

Given the difficulty in obtaining an order with the female data, the large number of 50% recombination values, and the unknown effects of recombination across the centromere, it was decided that map distances would only be estimated for the five-locus short-arm linkage group. These estimates and the comparable values from Sherman et al. (1984) and Wedd (1984) are presented in table 8. The map distances obtained from the different analyses are roughly comparable. Although a great deal of variability was observed for any given interval, the total map lengths were approximately equal and the male:female map-distance ratio was virtually constant across the various maps. In general, the male maps tended to agree more closely than the female maps. This was most likely due to the higher lod scores associated with the male values.

Chromosome 11p

Seriation and least-squares estimation of the 11p data produced the linkage map CAL-3.8-PTH-10.7-HBBC-12.3-HRAS1-2.6-INS, which is identical to that observed in others' analyses of these data (Fearon et al. 1984; Kittur et al. 1985). Again, using the five-locus uniform and nonuniform simulation results, one would conclude that, given a sample of 69 equivalent meioses, this order has a 46%–80% chance of being correct. If one allows one inversion, the probability of being correct increases 73%–95%. This result suggested that although the observed order had a good chance of being correct, there was still a very high probability (15%–27%) that the order could be incorrect by a single inversion. For this data set the mapping-parameter estimate was $p = 1.0$.

DISCUSSION

From the above analysis we conclude that the seriation methodology is an accurate and efficient means of obtaining preliminary multilocus gene orders and map-distance estimates from pairwise linkage data. From the Monte Carlo simulation results it was observed that when an incorrect order was obtained, it was most likely incorrect only by a single inversion between the most closely linked locus pair. The origin of these incorrect orders was not due to failure of the algorithm per se but was due instead to the method's attempt to reach an optimal configuration on the basis of a poor sample. These simulations demonstrate that, in order to obtain accurate orders of large linkage groups, sample sizes >100 meioses will be required. Even larger sample sizes (i.e., $\geq$500 meioses) will be required to obtain a precise estimate of the mapping parameter. However, with more reasonable sample sizes (~200–300 meioses), the estimated mapping parameter will lie in a relatively narrow range surrounding the true value. Thus, some qualitative information on interference (low, medium, or high) may be obtained.

Analysis of the GAW III data showed that the correct order and map distance could be recovered from a collection of nuclear families and small pedigrees. The results of the current analysis compared very favorably with those obtained by investigators who applied other multipoint methodologies. In our analysis, the mapping parameter could be accurately estimated and suggested that interference is best estimated on the basis of longer maps. The chromosome 1 data showed that the assumption of $p = 0.351$ did not provide the best overall fit to the observations. In our study the best estimate of the mapping parameter was $p = 0.58$.

The chromosomes 1 and 11p linkage data showed that the seriation methodology could recover maps from a variety of types of data. For chromosome 1, published lod-score data recovered, very quickly and efficiently, maps comparable with those obtained by other multipoint techniques. Furthermore, on the basis of the 11p analysis we demonstrated that the seriation and least-squares techniques provided results comparable with those obtained from maximum-likelihood estimation procedures. Starting with pedigree data, seriation pro-

duced the same order and approximately the same map-distance estimates as those obtained by other methods that had been used to analyze the Johns Hopkins Hospital 11p data set at the Fourth Genetic Analysis Workshop (GAW IV).

The studies suggest that a major limitation in multipoint mapping is presented by the data itself. This is best demonstrated by the analysis of the chromosome 11p data at GAW IV. Regardless of the methodology used, analysis of the Johns Hopkins' data set generated the previously presented order, which differed from the consensus order by an INS-RAS inversion (Grzeschik and Kazazian 1985). This suggests that the difference was due to the data set rather than to the analysis technique—and that it indeed represented the best order. Given the simulation results, such inversions between closely linked loci will be common and will require large volumes of data to be resolved. The seriation technique facilitates the accumulation of such data.

Seriation can produce accurate multipoint maps that are directly comparable with those obtained by other techniques. Although maximum-likelihood techniques are preferable, owing to their superior statistical properties, practical considerations in human gene mapping make seriation an attractive method. On the basis of the results of the simulated and empirical data analysis, we suggest the following multipoint mapping strategy:

1. Perform pairwise linkage analysis using the lod-score method to establish linkage groups and to test for heterogeneity.
2. Construct a matrix of the recombination values between all locus pairs and perform seriation to obtain locus order.
3. Estimate map distances and the mapping parameter by using the least-squares method.
4. If raw pedigree data are available on a subset of loci for which more detailed information is of interest, use a maximum likelihood based on simultaneous estimation procedure (e.g., the LINKAGE programs [Lathrop et al. 1984, 1985]) to obtain the maximum-likelihood estimates.

REFERENCES

Buetow, K. H., and A. Chakravarti. 1987. Multipoint mapping using seriation. I. General methods. Am. J. Hum. Genet. **41**:180–188.

Donald, J., E. Stuart, G. Der, and K. McKuster. 1985. Genetic Analysis Workshop III: use of multidimensional scaling in chromosome mapping. Genet. Epidemiol. **2**:207–209.

Edwards, J. H. 1971. The analysis of X-linkage. Ann. Hum. Genet. **34**:229–249.

Fain, P. R., and D. E. Goldgar. 1985. Genetic Analysis Workshop III: a biological model. Genet. Epidemiol. **2**:215–216.

Fearon, E. R., S. E. Antonarakis, D. A. Meyers, and M. A. Levine. 1984. c-Ha-ras-1 Oncogene lies between beta-globin and insulin loci on human chromosome 11p. Am. J. Hum. Genet. **36**:329–337.

Grzeschik, K. H., and H. H. Kazazian. 1985. Report of the Committee on the Genetic Constitution of Chromosomes 10, 11, and 12: Eighth International Workshop on Human Gene Mapping. Cytogenet. Cell Genet. **40**:179–205.

Haldane, J. B. S. 1919. The combination of linkage values, and the calculation of distances between the loci of linked factors. J. Genet. **8**:299–309.

Hodge, S. E., C. E. Anderson, K. Neiswanger, R. S. Sparkes, and D. L. Rimoin. 1983. The search for heterogeneity in insulin-dependent diabetes mellitus (IDDM): linkage studies, two-locus models, and genetic heterogeneity. Am. J. Hum. Genet. **35**:1139–1155.

Kittur, S. D., J. W. M. Hoppener, S. E. Antonarakis, J. D. J. Daniels, D. A. Meyers, M. Jansen, R. G. Korneluk, B. D. Nelson, and H. H. Kazazian. 1985. Linkage map of the short arm of human chromosome 11: location of the genes for catalase, calcitonin, and insulin-like growth factor II: Eighth International Workshop on Human Gene Mapping. Cytogenet. Cell Genet. **40**:669.

Kosambi, D. D. 1944. The estimation of map distance from recombination values. Annals of Eugenics **12**:172–175.

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.

———. 1985. Multilocus linkage analysis in humans: detection of linkage and estimation of recombination. Am. J. Hum. Genet. **37**:482–498.

MacCluer, J. W., C. Falk, and D. K. Wagener. 1985. Genetic Analysis Workshop III: summary. Genet. Epidemiol. **2**:185–198.

Ott, J. 1974. Estimation of the recombination fraction in human pedigrees: efficient computation of the likelihood for human linkage studies. Am. J. Hum. Genet. **26**:588–597.

Ott, J. 1985. Genetic Analysis Workshop III: Combining two-point analysis under the constraints of a linear map and constant female/male ratio. Genet. Epidemiol. **2**:217–218.

Povey, S., N. E. Morton, and S. L. Sherman. 1985. Report of the Committee on the Genetic Constitution of Chromosomes 1 and 2: Eighth International Workshop on Human Gene Mapping. Cytogenet. Cell Genet. **40**:67–106.

Rao, D. C., N. E. Morton, J. Lindsten, M. Hultén, and S. Yee. 1977. A mapping function for man. Hum. Hered. **27**:99–104.

Risch, N. 1985. Genetic Analysis Workshop III: simple methods for accurate gene mapping with multipoint data. Genet. Epidemiol. **2**:221–222.

Robson, E. G., and J. King. 1984. Family data on sixteen chromosome 1 loci. Ann. Hum. Genet. **48**:347–358.

Sherman, S. L., J. King, E. B. Robson, et al. 1984. A revised map of chromosome 1. Ann. Hum. Genet. **48**:243–252.

Sherman, S. L., and N. E. Morton. 1985. Genetic Analysis Workshop III: construction of genetic maps using two-point lod tables. Genet. Epidemiol. **2**:223–224.

Wedd, N. 1984. A computer program for constructing a maximum-likelihood map from linkage data and its application to human chromosome 1. Ann. Hum. Genet. **48**:333–346.