Genet. Res., Camb. (1997), **69**, *pp.* 35–43 *With* 4 *text-figures*

Neighbour mapping as a method for ordering genetic markers

T. H. N. ELLIS
Department of Applied Genetics, John Innes Centre, Colney Lane, Norwich NR4 7UH, UK

(*Received 8 July 1996 and in revised form 12 November 1996 and 9 December 1996*)

Summary

A modification of the neighbour joining method of Saitou & Nei (1987) is shown to be applicable to the ordering of genetic markers. This neighbour mapping method is compared with some other procedures for ordering genetic markers using both real and test data sets. The limitations and likely errors associated with the use of neighbour mapping are discussed. The speed and simplicity of this method commend its application, as does its concurrence with other mapping methods.

1. Introduction

An important problem in the generation of linkage maps from segregation data is the determination of marker order. There is a well-developed theoretical framework which is concerned with this problem, and several methods have been described for the determination of the order of markers (Lathrop & Llouel, 1984; Lander & Green, 1987; Lander *et al*., 1987; Burr *et al*., 1988; Morton & Andrews, 1989; Stam, 1993; reviewed by Bryant, 1996).

A particular marker order is decided upon after searching among all possible orders for the one which, given the data, best fits the appropriate theory. An inherent problem in this approach is that any order could be compatible with the data, but some suggest much more recombination than others. For many purposes, then, the search among orders is for the one which proposes the least number of recombination events and/or of double recombination events. Any such searching procedure is computationally intensive, because, for n markers there are $n!/2$ possible linear orders. Methods for restricting searches to limited subsets of orders have been exploited in order to speed up calculation – for example the prior computation of three-point linkage tests in MAPMAKER (Lander *et al*., 1987) or sequential addition of markers and local reshuffling of orders in JoinMap (Stam, 1993).

Here I describe another approach which can be used to determine marker orders. This method is suited to the rapid determination of marker orders from large sets of data, and can be used in conjunction with other procedures, or mapping programs, to generate a linkage map. This method treats a linkage map in much the same way as a cluster analysis, and owes much to the neighbour joining method of Saitou & Nei (1987) from which it is derived. The method was developed from attempts to consider the consequences for linkage maps of errors in data sets, and an attempt to display potential errors in the presentation of linkage maps is also discussed.

2. Materials and methods

(i) *Segregation data*

Pea segregation data are from a recombinant inbred population derived from the cross JI281 × JI399 which has been described previously (Ellis *et al*., 1992, 1993).

(ii) *Test data sets*

Two data sets were generated for a hypothetical recombinant inbred population of 100 lines, one with ten markers and the other with 49 markers, as follows. A set of scores was created by generating a sequence of + or − 'scores', chosen at random. This simulates the scores for 100 RI lines for one marker. This set was copied with occasional changes to generate the second maker scores. For the larger test data set the frequency of these changes was selected using the product of three random variables. The first variable (v1) was uniformly distributed between 0 and 0·4, and the second two (v2 and v3) between 0 and 0·7. This procedure was repeated on the second set of 'scores' to generate the third, and so on to generate the test

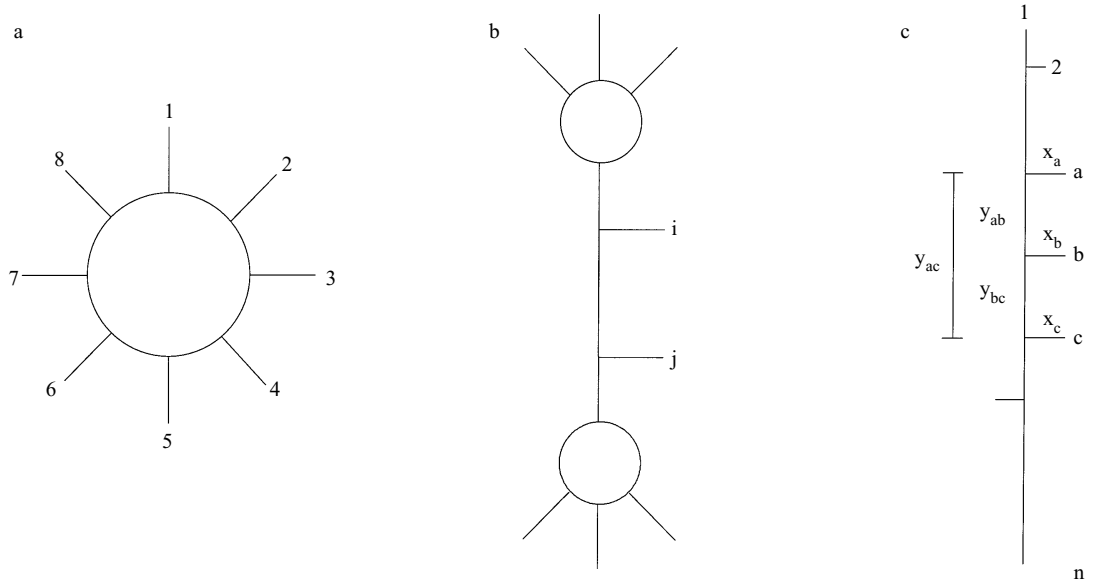

Fig. 1. Graphical illustrations of the methods used to determine marker orders and errors. (*a*) An unordered group. (*b*) A group where two adjacent markers have been identified. (*c*) A linkage map showing deviation from additivity (see text).

data set of 49 markers where the 'interval distances' were variable. For a smaller set of 10 'scores' a single random variable was used, so the marker interval distances were more even. In effect these procedures simulated the scores corresponding to recombination without interference, with variable inter-marker distances where the correct order is known.

(iii) *The computational method*

(1) This procedure is for determining the order of markers within a group. It follows that the markers to be included within the group must first be selected. This can be done in a variety of ways, such as two-point linkage tests with some threshold for inclusion. The group may comprise those markers which show an association with a LOD score above some value (usually 3), or for which there is less than a certain frequency of recombinant types.

(2) Having established the group of markers, a table of all two-point distances is calculated. These values could be map distances, frequencies of recombinant types, or some other function of the raw data. As the method is based on this table it can be applied to any population structure for which two-point data can be calculated, but the procedure is limited to those cases where the full table can be calculated.

(3) Following Saitou & Nei (1987) we have a starting group such as that shown in Fig. 1(*a*). This group is unstructured and the markers can be considered independent branches; the sum of all branch lengths is given by

$$S = \left(\frac{1}{n-1}\right) \sum_{x<y}^{n} D_{x,y} \qquad (1)$$

(see Saitou & Nei, 1987, equation 1), where S is the sum of all the branch lengths, n is the number of markers and D_{xy} is the two-point distance between the markers x and y.

If a pair of markers (i and j) are abstracted from this group, as in Fig. 1(*b*), then we can determine the sum of all remaining branch lengths, together with the distance between i and j. If these two markers are genuinely adjacent, then the sum of all branch lengths for the group will be minimized. The determination of minimum branch lengths by this procedure is discussed by Saitou & Nei (1987). This sum (S_{ij}) is given as

$$S_{i,j} = \left[\left(\frac{1}{n-3}\right) \sum_{\substack{x<y \\ x\neq i,j \\ y\neq i,j}}^{n} D_{x,y} \right] + D_{i,j}, \qquad (2)$$

where S_{ij} is the sum of all branch lengths where i and j are assumed to be adjacent. The sum given in (2) is a modification of equation 4 of Saitou & Nei (1987), and takes account of the difference in the shape of a neighbour joining tree and a linkage map.

Note that in the neighbour joining method the adjacent pair of OTUs is external to the remaining group, and can be replaced by a node which is the point of connection to the remaining unstructured group. This allows a recalculation of the S_{ij} table replacing i and j by a single node. In the present method the marker pair remains embedded within the linear group, so this recalculation is not appropriate. However, a neighbour joining tree, in the normal sense, can be calculated from the data, and this is also informative.

(4) Given a table of two-point linkage distances D_{ij}, the corresponding S_{ij} values can be tabulated and ordered with increasing S_{ij}, noting the corresponding i and j.

https://doi.org/10.1017/S0016672397002632 Published online by Cambridge University Press

Table 1. *Linkage order from S_{ij} values*

S_{ij}	i	j	Group	S_{ij}	i	j	Group
8·79403	1	2	1:2	8·95331	41	45	*x*
8·79416	3	4	3:4	8·95759	12	13	10:11:12:13:14
8·80883	2	3	1:2:3:4	8·95871	7	8	1:2:3:4:5:6:7:8:9
8·82899	2	5	*x*	…			
8·83194	2	4	*x*	8·97966	4	7	*x*
8·84547	4	5	1:2:3:4:5	8·98378	18	19	18:19
8·84698	42	43	42:43	…			
8·85829	1	5	*y*	8·98812	6	8	*x*
8·86285	1	4	*x*	8·98913	36	37	36:37
8·86418	3	5	*x*	…			
8·86518	1	3	*x*	8·99632	10	12	*x*
8·87401	41	42	41:42:43	8·99727	16	17	16:17
8·87433	38	39	38:39	…			
8·87507	2	6	*x*	8·99864	6	9	*x*
8·87991	10	11	10:11	8·99966	31	32	31:32
8·8818	46	47	46:47	9·00762	15	16	15:16:17
8·88678	5	6	1:2:3:4:5:6	…			
8·88746	42	47	*x*	9·01537	40	46	*x*
8·88946	1	6	*y*	9·01567	17	18	15:16:17:18:19
8·89358	42	45	*x*	…			
8·89497	47	48	46:47:48	9·02212	11	14	*x*
8·90251	6	7	1:2:3:4:5:6:7	9·02348	35	36	35:36:37
8·9043	41	43	*y*	…			
8·90586	8	9	8:9	9·03876	9	10	1:2:3:4:5:6:7:8:9:10:11:12:13:14
8·90665	13	14	13:14	9·04229	21	22	21:22
8·90688	45	46	45:46:47:48	9·04543	20	21	20:21:22
8·90857	44	45	44:45:46:47:48	…			
8·90874	43	45	*x*	9·05575	3	8	*x*
8·9117	42	46	*x*	9·05711	19	20	15:16:17:18:19:20:21:22
8·9177	43	47	*x*	…			
8·91803	42	44	*x*	9·06014	17	21	*x*
8·92086	4	6	*x*	9·06034	30	31	30:31:32
8·92348	40	41	40:41:42:43	…			
8·9274	48	49	44:45:46:47:48:49	9·06234	16	19	*x*
8·92747	40	42	*x*	9·0679	28	29	28:29
8·92758	41	47	*x*	9·06808	22	23	15:16:17:18:19:20:21:22:23
8·9316	43	44	40:41:42:43:44:45:46:47:48:49	…			
8·93694	5	9	*x*	9·07022	20	22	*x*
8·93768	42	48	*x*	9·07073	37	38	35:36:37:38:39
8·93853	11	12	10:11:12	9·07113	17	23	*x*
…				9·0712	33	34	33:34
9·07311	14	15	1:2:3:4:5:6:7:8:9:10:11:12:13:14:15:16:17:18:19:20:21:22:23				
…							
9·21866	26	27	1:2:3:4:5:6:7:8:9:10:11:12:13:14:15:16:17:18:19:20:21:22:23:24:25:26:27:28:29:30:31:32:33:34:35:36:37:38:39:40:41:42:43:44:45:46:47:48:49				

x, at least one of the markers is already connected to two others.
y, joining these two markers would form a ring.

(5) From the table of S_{ij}, i, and j, an order of markers can be determined as follows; the procedure is illustrated in Table 1 as applied to the pea linkage maps shown later.

(a) The lowest value of S_{ij} corresponds to the first pair of markers considered to be adjacent.
(b) The successively higher values of S_{ij} correspond to successive candidates for adjacent markers.
(c) No marker is considered to be adjacent to more than two other markers.
(d) No two markers are adjacent if they are already connected through a sequence of other markers. In the absence of this restriction rings of markers could be formed.

Simple computer programs which generate and/or read tables of D_{ij} and S_{ij} are available on request, and see Gelfand (1971) for a similar formulation. From Table 1 it can be seen that there is a possibility for ambiguity when two values of S_{ij} are the same. This is a problem only in those cases where the particular S_{ij} values imply alternative decisions; this can be discovered either by inspection of a table such as Table 1, by a computer search of S_{ij} values, or by repeating the ordering procedure from a different starting

https://doi.org/10.1017/S0016672397002632 Published online by Cambridge University Press

sequence. Note that the starting sequence is not relevant if all the S_{ij} values are different as in Table 1. For convenience the marker numbers 1 to 49 correspond to the marker order for the pea linkage maps shown later.

3. Results

(i) *Tests of the procedure*

A trial of the neighbour mapping method was performed on a test data set generated as described above. The scored data were then jumbled and the procedure applied. The same data set was used to test MAPMAKER and the method of computing the minimum number of recombinants for all possible marker orders (Ellis, 1994). For a set of 10 markers these three methods derived the same, correct, marker order (not shown).

This test was repeated for the data set with 49 markers, and the comparison made with MAPMAKER alone. This comparison is shown in Fig. 2. Four maps are drawn. The first (T) denotes the interval distances which were set in the generation of the data set. These were derived from the product of the three variables discussed above which simulated variability in map distances. The second map (D) is the map derived from the data when the markers are in the correct order, i.e. this is the best possible map which could be derived from the data. Map D is a sample of the expected data from T. The third map (N) shows the map derived from the neighbour mapping method. This map is indistinguishable from D, with the exception that unresolved markers cannot be ordered. The MAPMAKER map (M) is very similar except that three markers were not included in the first trial. The intervals to which these were assigned are marked with an asterisk, and these were appropriate positions.

From these tests it is clear that the simple procedure described here is capable of determining the 'correct' marker order. For real data sets the 'correct' order is not known, and is inferred from the data on the basis of some assumptions, such as presence or absence of interference in otherwise random events, additivity of the mapping function, and a lack of scoring and sampling errors. If all these hold then the table of D_{ij} values will behave so that the present procedure generates an appropriate marker order. The interesting question about any marker-ordering method is how the procedure behaves with non-ideal data, so the method has been applied to some experimental data from pea.

A comparison between the present method and MAPMAKER is shown in Fig. 3 for pea linkage group III. This group has the same number of markers as the test data set. The time taken to compute the marker orders by the present method was just under 3 min. MAPMAKER, which recognized this as a single linkage group with the default thresholds of LoD 4 and max distance of 20 centimorgans, took a little over 3 min to compute three-point linkage data, and then about 9 min to derive the larger partial order shown in Fig. 3. The time taken to generate the smaller partial group was not noted. The two methods used different computing systems; MAPMAKER was run using a UNIX operating system on a DEC workstation, while the present method used simple, compiled BASIC programs run on a VAX. Both machines were used in a network, so it is likely that faster times could be achieved. Nevertheless, it seems clear that the method presented here requires less computation.

It can be seen from Fig. 3 that these two methods are not exactly equivalent, although the marker orders are in broad agreement. MAPMAKER has failed to place some of the markers, but suggested an interval where each marker might lie, as was the case for the test data set. The linkage group was not assembled as a single contiguous linkage map at the first attempt. The 14 markers at the top in Fig. 3 were not ordered in the first trial, although they were assigned to the upper exterior of the large linkage segment. A separate attempt with these 14 markers alone generated the smaller linkage map which is included in the figure. In addition MAPMAKER did not suggest a location for the two markers (C2/1− and G12/4+) where the arrows are drawn in Fig. 3.

The discrepancies between the methods are informative. There are several local inversions of the order of closely linked markers – for example the segment between cDNA40/7 and A7/11+. Presumably this reflects the fact that this is a tight cluster of markers flanked by two large distances devoid of markers. Arrangements of this type have little informative data on the orientation of the internal cluster with respect to the flanking pair.

Two of the local inversions between the two maps flank the break in the contiguous group and are presumably the reason why MAPMAKER did not assemble a single map. This probably also accounts for the failure to place the markers designated cDNA331 and C2/1+ (a cDNA RFLP and an AFLP respectively). A similar type of argument could explain the loss of the marker G12/4+.

Neither of the maps shown in Fig. 3 is claimed to be 'correct', and clearly the MAPMAKER analysis is preliminary. However, as a first approximation to a map the present method is fast, simple, and in broad agreement with the more complex MAPMAKER analysis. For these reasons the method would seem to be of some general interest.

(ii) *Errors on maps*

The method described here coalesces a group of markers around pairs which have a strong tendency to co-segregate. The reliability of this procedure depends

https://doi.org/10.1017/S0016672397002632 Published online by Cambridge University Press

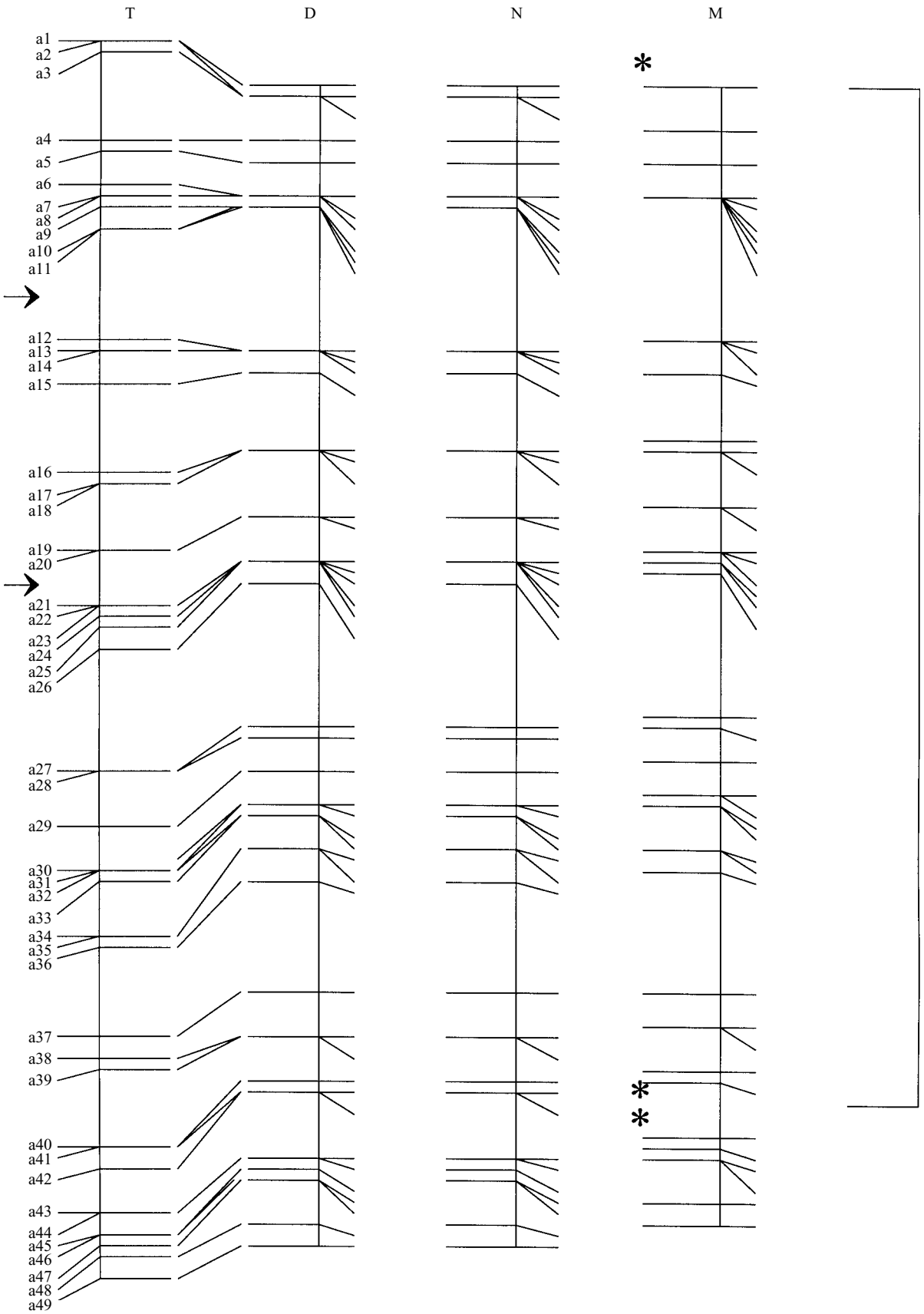

Fig. 2. Comparison of linkage maps from a test data set determined by different methods. T, the Theoretical model from which the test data set was generated; D, the Data set derived from the model; N, the New method described here; M, the MAPMAKER-derived map. Scale bar represents 50 centimorgans. * Suggested intervals for missing markers.

https://doi.org/10.1017/S0016672397002632 Published online by Cambridge University Press

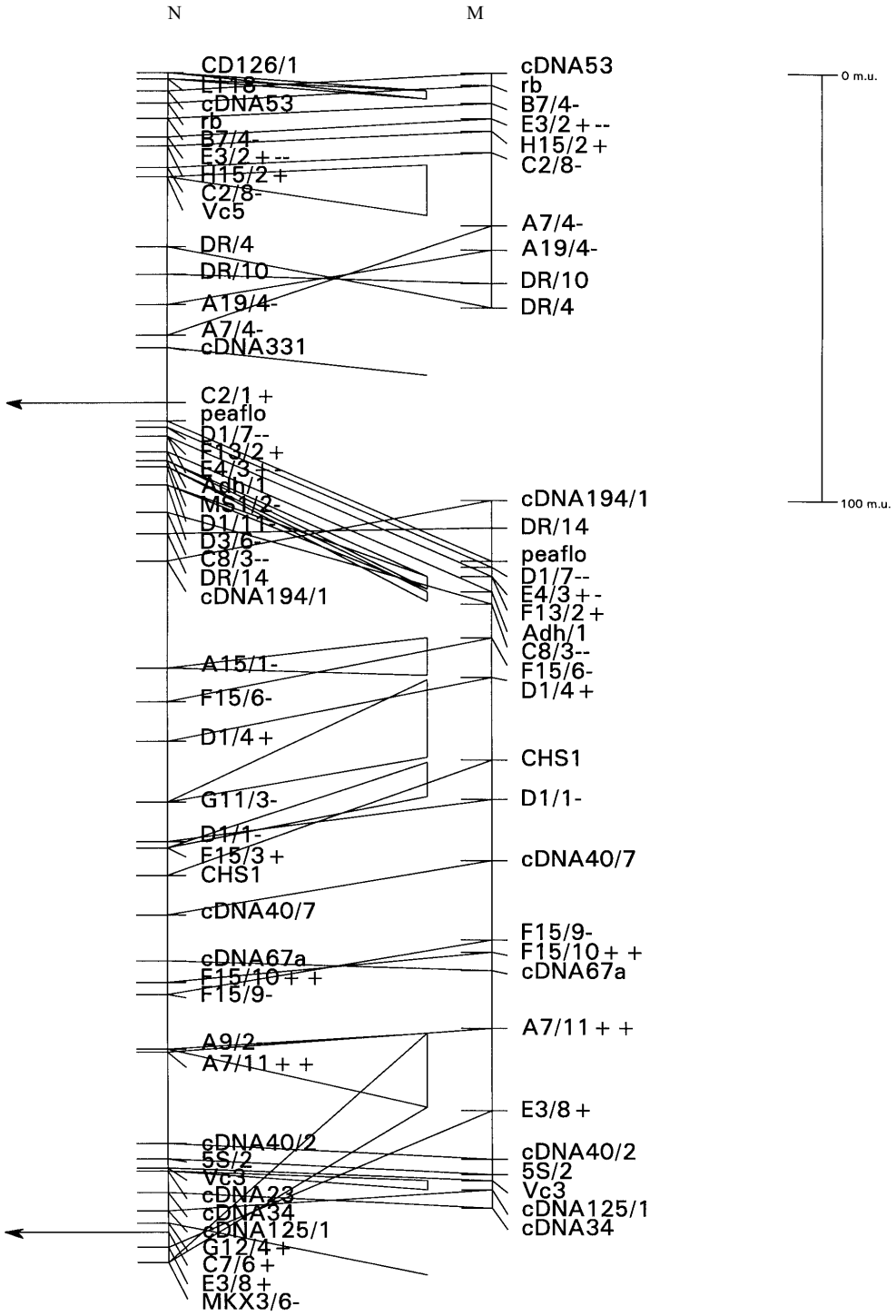

Fig. 3. Comparison between a MAPMAKER-derived map and the new method for pea linkage group III. Corresponding marker loci are connected by diagonal lines; suggested intervals according to MAPMAKER are mapped as open triangles. The two diagonal lines which do not connect to the MAPMAKER map were suggested to lie external to the MAPMAKER map at the end indicated. The two arrows indicate markers excluded from the MAPMAKER map. Scale bar represents 100 centimorgans.

on the markers having been scored accurately. If a marker is scored with frequent error, then it is possible that any procedure will mis-place the marker. However, the present mapping method gives most weight to the most closely linked marker pairs (by identifyinging these first), which should minimize the influence of error-prone markers for the map as a whole. Closely-linked markers could be considered independent tests for (non-)recombination in a larger interval.

The neighbour mapping method was developed from attempts to understand the consequences of data errors; this was what suggested an approach related to cluster analysis. A simple graphical method for the

https://doi.org/10.1017/S0016672397002632 Published online by Cambridge University Press

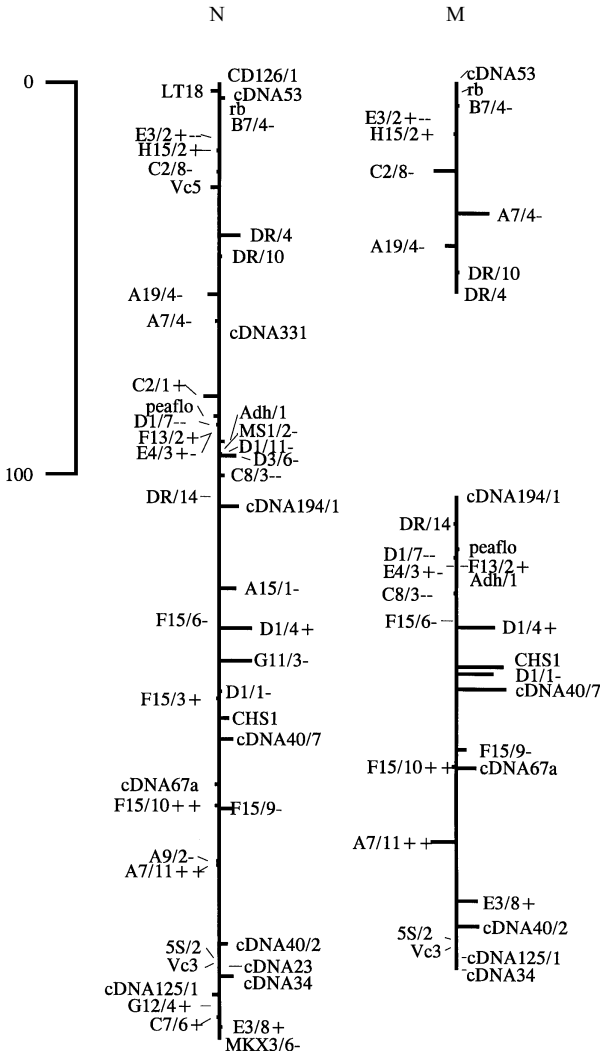

Fig. 4. The maps of Fig. 3 redrawn to indicate local non-additivity.

representation of error is shown in Fig. 4, corresponding to the pea maps already presented. Note that this diagram is intended to present potential problems with the map, not the data. Different maps from the same data will appear different. However, highlighting potential problems with the map directs attention to problematic markers.

If we consider three markers which have been placed on a linkage map in the order … a, b, c, …, we have three two-point map distances d_{ab}, d_{bc} and d_{ac}. If the mapping function is appropriate, then we expect $d_{ab}+d_{bc}=d_{ac}$. If the scores for b were error-prone we might expect $d_{ab}+d_{bc}>d_{ac}$; i.e. an excess of double recombinants flanking marker b appears. If there are fewer double recombinants than anticipated by the mapping function, then $d_{ab}+d_{bc}>d_{ac}$. The deviation of d_{ac} from its expected value is simple to calculate, and can be used to calculate an offset of the position of the marker from the map. This offset, x_b, is given as $x_b=\frac{1}{2}(d_{ab}+d_{bc}-d_{ac})$ and represents the deviation of the map from local additivity. The interval distances d_{ab} and d_{bc} can be replaced by y_{ab} and y_{bc}, respectively, where $y_{ab}=d_{ab}-x_b$ and $y_{bc}=d_{bc}-x_b$ respectively (Fig. 1(*c*)). Thus the total map length assumes an absolute interference in adjacent intervals, with all close doubles treated as error. For linkage group III this modified linear map has been reduced in overall length by about 12%.

(iii) *Errors in data*

Problems arising from inaccurate data can be estimated by considering the frequency distribution of lengths of linkage segments bounded by double exchange. For an individual, a linkage map can be considered to be a sequence of two types of intervals: those with an exchange (E), and those without an exchange (W). An isolated marker flanked by an exchange on either side can be considered to be two adjacent intervals carrying an exchange (EE). Two adjacent markers of the same allelic type flanked by exchanges can be considered to be a run of three intervals: exchange, non-exchange, exchange (EWE). In general this type of distribution has been described by Mood (1940), and an approximation to this has been used in the study of repetitive DNAs by Slack (1974) and Southern (1975). If the probability of an interval having an exchange is p_e and the probability of an interval lacking an exchange is $(1-p_e)$, then a run of z intervals without an exchange, bounded by intervals with an exchange (EW_zE) will have an expected frequency (F) of $p_e^2(1-p_e)^z$. From this relationship p_e can be determined by plotting $\log(F)$ against z. Note that these estimates of p_e should approximate to, but may be different from, the fraction of intervals with an exchange. The estimates of p_e from this regression can be used to estimate the value of p_e when there is no error contributing to the $z=0$ class.

Following this procedure for the data used to generate the linkage maps of Figs. 3 and 4 shows that there is an excess of the classes where $z=0$ or 1. If these classes are eliminated from analysis p_e can be determined by a regression analysis from the longer exchange segments. This treatment also suggests that the length of the linkage map of group III shown in Fig. 3 is probably exaggerated by about 10%. This treatment suggests that the errors presented by the map in Fig. 4 are consistent with the expected data errors, rather than errors in the map *per se*.

4. Discussion

A simple method for deducing the orders of molecular markers in linkage maps has been presented. This method is based on the neighbour joining method of Saitou & Nei (1987), and has two main advantages: it is fast because the amount of calculation required is fairly small for a given data set, and it always returns a map which places all the markers for a given group.

https://doi.org/10.1017/S0016672397002632 Published online by Cambridge University Press

The second advantage is also an inherent problem with the method, because the inclusion of an inappropriate marker will be tolerated whereas MAPMAKER, for example, would exclude such a marker. However, there are two likely outcomes from such an erroneous inclusion. The first is that the marker will be placed at the end of the map, and at a considerable distance. The second is that the marker will be placed at an internal position in the map, and will be flanked by large intervals. These could break the map if it is drawn in such a way as to forbid intervals greater than some predetermined size. Large intervals signal a problem with the map and warrant further investigation.

The neighbour mapping method has some similarities with 'seriation' as described by Gelfand (1971) and Buetow & Chakravarti (1987). The rules given for the derivation of marker order are essentially those described as 'method I' by Gelfand (1971). The significant difference between the two procedures is that seriation operates on the D_{ij} matrix while neighbour mapping, like neighbour joining (Saitou & Nei, 1987) operates on the S_{ij} matrix. When the D_{ij} matrix is not purely additive (or a Robinson matrix as defined by Gelfand, 1971) then these two procedures are not necessarily equivalent, the neighbour mapping method finds those close pairs of markers which, when joined, cause the other markers to be least disturbed, and this is the essential difference between the procedures.

The suggestion for the presentation of linkage maps is an attempt to highlight such problems in a graphical way. The difference in the errors associated with the markers C2/8− and A7/4− in the two orders presented in Fig. 4 are clearly associated with the different marker orders proposed by these two maps. The single contiguous map minimizes these particular errors in the present case, but this simply says that the map is compatible with the data, not that it is correct. The maps in Fig. 2 show clearly that different ordering methods produce similar results, but that the mere fact of dealing with a finite population size has a large effect.

One objective of the present study was to test the possibility that close double exchanges (or errors which have this appearance) may be responsible for the discrepancy between cytogenetic data and linkage maps (Nilsson *et al*., 1993; and see Sherman & Stack, 1995). For pea linkage group III as illustrated in Figs. 3 and 4 the total map length has been reduced from 278 to 244 centimorgans (using Haldane's function: Haldane, 1919). This corresponds to a reduction in the expected average number of chiasmata from 5·6 to 4·9 for the chromosome corresponding to linkage group III. Other estimates of probable exaggeration of the length of the linkage map, based on exchange segment length, suggest a similar reduction. Average chiasma counts for pea are in the range 10–20 per meiosis, (see Nilsson *et al*., 1993) and the frequency of crossing-over in both arms suggests that the number of crossovers does not differ greatly between chromosomes (these are expected to be in the range two or three per chromosome per meiosis). The map remains excessively long compared with the cytogenetic data.

In the examples discussed here, the method which has been described generated convincing maps directly, but it would be unwise to assume that this is always the case, especially when there are large gaps and clusters of tightly linked markers to be included in the map. Removal of the data corresponding to the markers between the segments arrowed in Fig. 2 led to a reordering of the top segment of the map with respect to the lower portion. A similar, but different reorganization occurred when MAPMAKER was used on this restricted data set. Segmental inversions are a likely error associated with the neighbour mapping method, and large gaps need to be treated with caution. Other ordering methods generate a different spectrum of likely errors, suggesting that the method described here could be used profitably in conjunction with other mapping procedures – for example by seeding a starting order in MAPMAKER or in suggesting orders for the addition of markers to JoinMap.

I thank R. Hellens for many useful discussions, R. Casey, J. Hofer, G. Moore and J. Snape for their comments on this manuscript, and one anonymous referee for attracting my attention to the similarities between this method and seriation.

References

Bryant, S. P. (1996). Software for genetic linkage analysis. *Molecular Biotechnology* **5**, 49–61.

Buetow, K. H. & Chakravarti A. (1987). Multipoint gene mapping usng seriation I. General methods. *American Journal of Human Genetics* **41**, 180–188.

Burr, B., Burr, F., Thompson, K. H., Albertson, M. C. & Stuber, C. W. (1988). Gene mapping with recombinant inbreds in maize. *Genetics* **118**, 519–526.

Ellis, T. H. N. (1994). Approaches to the genetic mapping of pea. *Modern Methods in Plant Analysis* **16**, 117–159.

Ellis, T. H. N., Turner, L., Hellens, R., Lee, D., Harker, C. L., Enard, C., Domoney, C. & Davies, D. R. (1992). Linkage maps in pea. *Genetics* **130**, 649–663.

Ellis, T. H. N., Hellens, R. P., Turner, L., Lee, D., Domoney, C. & Welham, T. (1993). On the pea linkage map. *Pisum Genetics* **25**, 5–12.

Gelfand, A. E. (1971). Seriation. In *Mathematics in the Archaeological and Historical Sciences*, ed. F. R. Hodson, D. G. Kendall & P. Tăutu, pp. 186–201. Edinburgh: Edinburgh University Press.

Haldane, J. S. B. (1919). The combination of linkage values, and the calculation of distance between loci of linked factors. *Journal of Genetics* **8**, 299–309.

Lander, E. S. & Green, P. (1987). Construction of multilocus linkage maps in humans. *Proceedings of the National Academy of Sciences*, *USA* **84**, 2363–2367.

Lander, E. S., Green, P., Abrahamson, J., Barlow, A., Daly, M. J., Lincoln, S. E. & Newburg, L. (1987). MAPMAKER: an interactive computer package for constructing genetic linkage maps of experimental and natural populations. *Genomics* **1**, 174–181.

Lathrop, G. M. & Llouel, J. M. (1984). Easy calculation of Lod scores and genetic risks on small computers. *American Journal of Genetics* **36**, 460–465.

Mood, A. M. (1940). The distribution theory of runs. *Annals of Mathematical Statistics* **11**, 367–392.

Morton, N. E. & Andrews, V. (1989). MAP, an expert system for multiple pairwise linkage analysis. *Annals of Human Genetics* **53**, 263–269.

Nilsson, N.-O., Sall, T. & Bengtsson, B. O. (1993). Chiasma and recombination data in plants: are they compatible? *Trends in Genetics* **9**, 344–348.

Saitou, N. & Nei, N. (1987). The neighbour joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**, 406–425.

Sherman, J. D. & Stack, S. M. (1995). Two dimensional spreads of synaptonemal complexes from solanaceous plants. VI. High resolution recombination nodule map for tomato (*Lycopersicon esculentum*). *Genetics* **141**, 683–708.

Slack, J. M. W. (1974). The interpretation of oligonucleotide maps: a theoretical study of nucleic acid digests with special reference to repeated diverged sequences. *Biopolymers* **13**, 224–264.

Southern, E. M. (1975). Long range periodicities in mouse satellite DNA. *Journal of Molecular Biology* **94**, 51–69.

Stam, P. (1993). Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. *The Plant Journal* **3**, 739–744.