

# Effects of genotyping errors, missing values and segregation distortion in molecular marker data on the construction of linkage maps

CA Hackett and LB Broadfoot

*Biomathematics & Statistics Scotland, Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, UK*

A simulation study was performed to investigate the effects of missing values, typing errors and distorted segregation ratios in molecular marker data on the construction of genetic linkage maps, and to compare the performance of three locus-ordering criteria (weighted least squares, maximum likelihood and minimum sum of adjacent recombination fractions criteria) in the presence of such effects. The study was based upon three linkage groups of 10 loci at 2, 6, and 10 cM spacings simulated from a doubled-haploid population of size 150. Criteria performance were assessed using the number of replicates with correctly estimated orders, the mean rank correlation between the estimated and the true order and the mean total map length. Bootstrap samples from replicates in the maximum likelihood analysis produced

a measure of confidence in the estimated locus order. The effects of missing values and/or typing errors in the data are to reduce the proportion of correctly ordered maps, and this problem worsens as the distances between loci decreases. The maximum likelihood criterion is most successful at ordering loci correctly, but gives estimated map lengths, which are substantially inflated when typing errors are present. The presence of missing values in the data produces shorter map lengths for more widely spaced markers, especially under the weighted least-squares criterion. Overall, the presence of segregation distortion has little effect on this population.

*Heredity* (2003) 90, 33–38. doi:10.1038/sj.hdy.6800173

**Keywords:** doubled haploid; genotyping errors; linkage mapping; locus order; missing values; segregation distortion

## Introduction

The ideal set of molecular marker data for linkage mapping has no missing values, no genotyping errors and the markers segregate in the expected ratio for that type of population. In practice, however, mapping data is complicated by all of these factors. The effect of genotyping errors, especially errors arising in data entry, was quickly recognised in the human genetics mapping literature (Shields *et al*, 1991). Simulation studies have shown that undetected typing errors at a rate of 1% can lead to incorrect map orders and inflation of map lengths, particularly as marker density increases (Buetow, 1991). The effects of missing values and distorted segregation ratios on map order and map length have had less attention.

This study was motivated by difficulties in placing several markers on a linkage map constructed from marker data scored on a population of 154 doubled-haploid lines of barley (*Hordeum vulgare* L.). The population had missing values with a mean proportion 4.0% of genotypes missing, and it was thought likely that typing errors could be present in the dataset. In common with many barley doubled-haploid populations, the population also showed some segregation distortion and it has been suggested that this is because of genes

affecting survival in anther or microspore culture (Zivy *et al*, 1992; Manninen, 2000). Bailey (1961) showed that the equation for the maximum likelihood estimate of the recombination frequency between two loci in a backcross population (and similarly a doubled-haploid population) is unchanged if one of the loci affects the viability. It is straightforward to show that the maximum likelihood estimate of the recombination frequency is likewise unchanged for two loci that do not affect the viability themselves, but are linked to a third locus affecting viability. In principle, distorted segregation ratios should not give difficulties in mapping unless there are two linked loci, both of which affect viability.

To investigate influences on linkage maps, we have simulated populations of a similar size and structure to the experimental one and have introduced combinations of typing errors, missing values, and distorted segregation ratios. We report on their analysis by different locus-ordering methods.

## Methods

Doubled-haploid populations were simulated using the software package GREGOR (Tinker and Mather, 1993). Two inbred homozygous parental lines A and B, with genotypes AA and BB at each locus, were simulated initially. These were crossed to give an F<sub>1</sub> population with genotype AB at each locus, and doubled-haploid individuals were simulated from the gametes produced by the F<sub>1</sub> generation. In all 30 markers were simulated on three chromosomes, each chromosome consisting of 10

Correspondence: CA Hackett, Biomathematics and Statistics Scotland, Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, UK. E-mail: christine@bioss.ac.uk

Received 12 September 2001; accepted 3 August 2002

**Table 1** Simulated marker data sets

Marker configuration	Missing values (%)	Typing errors (%)	Segregation ratio distortion
A	0	0	No
B	10	0	No
C	20	0	No
D	0	1	No
E	0	3	No
F	20	3	No
G	0	0	Yes
H	10	0	Yes
I	20	0	Yes
J	0	1	Yes
K	0	3	Yes
L	20	3	Yes

markers placed at equal intervals. The interval lengths for the three chromosomes were 2, 6 and 10 cM, respectively. No interference was simulated, and so Haldane's mapping function was used throughout in the subsequent analyses. The number of individuals in each doubled-haploid population was 150, and 100 populations were simulated for each configuration of marker data.

A total of 12 configurations of marker data were simulated to determine the effect of missing values, typing errors, distorted segregation ratios and combinations of these on linkage mapping. Table 1 shows the combinations used. To generate the populations with distorted segregation ratios, a larger population of doubled-haploid individuals was simulated as described above. The sixth marker on each chromosome was chosen to affect viability, with reduced viability for individuals with the *AA* genotype for the chromosomes with 2 or 6 cM spacing between markers, or with the *BB* genotype for the chromosome with 10 cM markers. Individuals with any of these genotypes were removed from the population randomly, with a probability of removal of 0.5. This left a population with distorted segregation ratios of approximately 1:2 at the sixth marker of each chromosome. The degree of segregation distortion at linked markers decreases with distance from the sixth marker.

Missing values and typing errors were randomly generated in the nondistorted and distorted populations to give populations with 10% missing values, 20% missing values, 1% typing errors, 3% typing errors and a combination of both 20% missing values and 3% typing errors.

There were 3 criteria were used to compare locus orders: (a) weighted least squares (Stam, 1993) weighted by the LOD scores, (b) maximum likelihood (Lander and Green, 1987) and (c) minimum sum of adjacent recombination fractions (SARF) (Falk, 1989). The minimum SARF criterion is motivated by the consideration that if the true order of three loci is  $L_1-L_2-L_3$ , then the recombination frequency between  $L_1$  and  $L_3$  is greater than between  $L_1$  and  $L_2$ , or between  $L_2$  and  $L_3$ , and so the sum of adjacent recombination frequencies is smallest for the true order. Olson and Boehnke (1990) have shown in a study of eight methods (not including maximum likelihood) that minimum SARF was among the two best overall locus-

**Table 2** Bootstrap results for analysis of one simulation of set D with 10 loci at 2 cM intervals

Genome position	Locus number									
	1	2	3	4	5	6	7	8	9	10
1	74	19	0	4	0	0	3	0	0	0
2	20	72	5	1	1	1	0	0	0	0
3	4	6	86	2	2	0	0	0	0	0
4	0	2	6	43	41	4	4	0	0	0
5	2	0	3	40	41	6	4	3	0	1
6	0	1	0	4	3	61	23	2	5	1
7	0	0	0	0	9	22	58	2	3	6
8	0	0	0	0	0	1	3	80	13	3
9	0	0	0	0	3	5	0	8	53	31
10	0	0	0	6	0	0	5	5	26	58

The frequencies were obtained from 100 bootstrap replications.

ordering criteria. Locus-ordering criteria (a)–(c) can be found in linkage analysis software. For example JoinMap (Stam and Van Ooijen, 1995) uses weighted least squares and PGRI (Lu and Liu, 1995) has several criteria for ordering loci, including minimum SARF and maximum likelihood.

A method is required to search among the locus orders for the optimum value of the criterion. JoinMap uses a stepwise algorithm similar to the seriation method (Buetow and Chakravarti, 1987). PGRI has a choice of search methods: we chose a simulated annealing algorithm (Kirkpatrick *et al.*, 1983) with the maximum likelihood criterion and a combination of simulated annealing and branch-and-bound algorithms (Thompson, 1987) with the minimum SARF criterion. Linkage analyses on the 12 configurations of marker data were performed using each locus-ordering method. The linkage maps for the maximum likelihood and minimum SARF criteria were almost identical to each other and so only the results for the maximum likelihood are presented.

Locus order is subject to sampling variation and small changes in the marker data can affect the estimated order. PGRI enables a measure of the confidence in the locus order to be calculated by bootstrapping. From each set of marker data and the corresponding locus order derived. A frequency matrix was constructed for the bootstrap samples with genome position as the row indicator and locus number as the column indicator, as shown in Table 2. The higher the frequencies of entries there are on the leading diagonal, the more consistency there is among the bootstrap samples and the more confidence that can be had in the estimated locus order. Table 2 can be summarised by the mean frequency of correct locus placings  $\lambda$ , equal to the sum of the leading diagonal divided by the number of loci.

## Results

Table 3 presents the number of correct orders out of 100 replicate simulations for each configuration of marker data, the mean rank correlation between the estimated and the true order and the mean total map length for maps calculated using weighted least squares

**Table 3** Linkage analysis using weighted least-squares criterion

	Original	10% missing	20% missing	1% error	3% error	3% error, 20% missing
Segregation ratio 1:1						
Locus distance (cM)	A	B	C	D	E	F
(a) Number of maps with correct order						
2	79	52	40	29	16	6
6	97	83	66	86	61	31
10	100	91	90	92	87	63
(b) Mean rank correlations (s.d.)						
2	0.99 (0.02)	0.98 (0.07)	0.96 (0.11)	0.95 (0.10)	0.94 (0.11)	0.88 (0.15)
6	1.00 (0.00)	0.99 (0.02)	0.99 (0.04)	1.00 (0.01)	0.99 (0.02)	0.98 (0.04)
10	1.00 (0.00)	1.00 (0.01)	1.00 (0.01)	1.00 (0.01)	1.00 (0.01)	0.98 (0.06)
(c) Mean map length (s.d.)						
2	18.1 (3.6)	18.9 (3.6)	18.5 (3.8)	23.1 (4.0)	33.2 (4.6)	32.5 (5.1)
6	53.5 (6.9)	50.4 (6.1)	49.9 (6.5)	57.1 (6.0)	68.9 (7.6)	66.0 (8.0)
10	89.7 (9.7)	88.5 (8.6)	86.4 (9.1)	97.8 (9.2)	114.3 (11.2)	109.3 (13.2)
Segregation distortion present						
	G	H	I	J	K	L
(a) Number of maps with correct order						
2	70	49	32	30	5	6
6	100	98	84	92	63	28
10	100	98	90	99	82	60
(b) Mean rank correlations (s.d.)						
2	0.99 (0.02)	0.99 (0.02)	0.94 (0.14)	0.94 (0.12)	0.93 (0.08)	0.87 (0.15)
6	1.00 (0.00)	1.00 (0.00)	0.99 (0.03)	1.00 (0.00)	0.99 (0.03)	0.97 (0.07)
10	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.01)	0.99 (0.01)
(c) Mean map length (s.d.)						
2	17.0 (3.3)	16.9 (4.1)	16.4 (4.5)	21.5 (4.8)	31.5 (5.0)	31.2 (5.3)
6	53.4 (6.9)	49.4 (8.4)	48.0 (8.7)	56.3 (9.2)	68.3 (11.4)	64.9 (11.7)
10	89.9 (8.8)	84.3 (10.9)	82.9 (12.8)	93.1 (12.7)	108.9 (13.7)	103.5 (14.6)

The mean of 100 replicate simulations are given, with standard deviations in brackets. Figures for rank correlations are rounded to 2 decimal places.

in JoinMap. Table 4 shows the same results using the maximum likelihood criteria in PGRI. Table 5 gives the mean percentage of correct locus placings,  $\lambda$ , obtained from the bootstrap samples.

The performance of both locus-ordering criteria improves as the distances between marker loci increase. The number of maps with correctly estimated orders is lower for all configurations when the distance between marker loci is 2 cM compared to the 6 and 10 cM distances. The percentage of correct orders at the 2 cM distance is between 70 and 83% in sets A and G (with neither missing values nor typing errors) but this is reduced by the presence of missing values or typing errors, especially in sets E, F, K and L, which all have 3% typing errors. However, at 10 cM intervals, locus order is relatively robust to missing values and typing errors, and only when a combination of 3% typing errors and 20% missing values is present (sets F, L) does the number of correctly ordered maps drop dramatically. At 6 cM distances, there is a marked decrease in the percentage of correct orders when the data has 20% missing values and/or 3% typing errors, especially under the weighted least-squares criterion. A comparison between sets A and G, B and H, etc shows that the presence of distorted

segregation ratios has little effect on the number of correctly ordered maps.

The number of correct maps when the data are analysed using the weighted least-squares criterion (JoinMap) is consistently lower than for the maximum likelihood criterion ( PGRI ). This is particularly noticeable when missing values or typing errors are present. However, the mean rank correlations for maps constructed with loci at 6 and 10 cM intervals are very similar for the two criteria, signifying that although the weighted least-squares criterion gets more orders wrong, bigger differences from the true locus order result using the maximum likelihood criterion. Overall, the mean rank correlation between estimated and true order is very high for both criteria and only drops below 0.9 for sets F and L when the loci are spaced at 2 cM intervals.

The results from sets D, E, F, J, K and L (with 1 or 3% typing errors) show substantial increases in the map lengths compared to the original sets A and G. This effect worsens as marker separation decreases and is most noticeable in those simulations where the markers are at 2 and 6 cM intervals. Using the weighted least-squares criterion, the 2 cM distance groups in sets E, F, K and L have average map lengths of 31.2–33.2 cM, almost twice

**Table 4** Linkage analysis using maximum likelihood criterion

	<i>Original</i>	<i>10% missing</i>	<i>20% missing</i>	<i>1% error</i>	<i>3% error</i>	<i>3% error, 20% missing</i>
Segregation ratio 1:1						
Marker loci distance (cM)	A	B	C	D	E	F
(a) Number of maps with correct order						
2	83	67	45	56	29	11
6	99	92	82	97	80	52
10	100	100	97	100	97	79
(b) Mean rank correlations (s.d.)						
2	1.00 (0.01)	0.99 (0.01)	0.98 (0.08)	0.98 (0.03)	0.95 (0.10)	0.88 (0.15)
6	1.00 (0.00)	1.00 (0.00)	1.00 (0.01)	1.00 (0.01)	0.99 (0.01)	0.97 (0.11)
10	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.01)
(c) Mean map length (s.d.)						
2	18.3 (3.3)	19.2 (3.5)	19.1 (3.6)	31.6 (4.5)	58.6 (6.7)	58.5 (7.7)
6	54.6 (6.1)	53.4 (6.1)	54.1 (7.1)	65.8 (6.3)	92.7 (8.3)	93.2 (12.5)
10	92.2 (9.5)	92.2 (8.9)	91.8 (9.4)	105.1 (8.6)	133.1 (11.1)	133.7 (11.0)
Segregation distortion present						
	G	H	I	J	K	L
(a) Number of maps with correct order						
2	75	61	35	43	20	8
6	99	99	85	96	85	51
10	100	100	94	100	94	64
(b) Mean rank correlations (s.d.)						
2	1.00 (0.01)	0.99 (0.01)	0.97 (0.06)	0.98 (0.04)	0.92 (0.13)	0.86 (0.17)
6	1.00 (0.00)	1.00 (0.00)	0.99 (0.05)	1.00 (0.00)	1.00 (0.02)	0.96 (0.13)
10	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.01)	0.98 (0.07)
(c) Mean map length (s.d.)						
2	17.2 (3.3)	16.8 (4.3)	17.4 (4.9)	29.9 (5.6)	57.4 (7.2)	57.0 (8.2)
6	54.3 (6.8)	52.1 (6.6)	52.1 (7.6)	65.2 (7.8)	91.2 (9.9)	91.5 (11.3)
10	92.1 (8.6)	88.7 (10.2)	89.8 (12.2)	101.8 (11.7)	129.3 (11.7)	130.4 (14.9)

The mean of 100 replicate simulations are given, with standard deviations in brackets. Figures for rank correlations are rounded to 2 decimal places.

that of the respective original data sets, A and G (17.0–18.1 cM). The average map lengths for E, F, K and L calculated using the maximum likelihood criterion were 57.4–58.6 cM, showing that this criterion is affected more by typing errors. The effect of missing values in the data is to shorten map lengths for more widely spaced markers, especially when there is also segregation distortion. However, this shortening effect is less obvious under the maximum likelihood criterion.

Bootstrapping was used with the maximum likelihood criterion to obtain a measure of confidence in the locus order. Table 2 shows an example of a single simulation of set D with the 10 loci spaced at 2 cM intervals. Of the 100 bootstrap replications, 74 placed locus 1 in position 1, 20 placed it in position 2, 4 in position 3 and 2 in position 5. The mean percentage of correct locus placings  $\lambda = 62.6\%$  for Table 2. Table 5 shows the mean  $\lambda$  values for each configuration and marker separation distance. The entries in this table may be interpreted as measures of confidence in the estimated locus orders. Confidence in the estimated orders is highest for sets A and G and lowest for sets F and L. It is consistently lowest for those maps constructed from locus markers at 2 cM intervals: the difference between these and 6 cM distances is

substantial. Confidence remains high at above 90% for 10 cM distances in all sets except sets F and L, the nondistorted and distorted populations with a combination of 20% missing values and 3% typing errors. At 6 cM distances, confidence in the loci order falls below 90% for sets C, E, H and K (with 20% missing values or 3% typing errors) and falls still further to below 70% for sets F and L. There are no substantial differences between non-distorted and distorted populations.

## Discussion

This study was motivated by difficulties in mapping in a population of 154 doubled-haploid lines of barley. The effects of the factors considered here will depend on the population size: the smaller the population, the more severe the effects are likely to be. Our simulation study has shown that when markers are separated by distances of about 10 cM, a map based on 150 individuals is quite robust to the presence of typing errors, missing values and segregation distortion. However, finer scale genetic maps are of increasing interest, and in this case the map order and length are increasingly affected. The bootstrap analysis reveals the extent of variation in locus orders as

**Table 5** Percentage of correct locus placings  $\lambda$ , from the analysis of 100 bootstrap samples

	Original	10% missing	20% missing	1% error	3% error	3% error, 20% missing
Segregation ratio 1:1						
Marker loci distance (cM)	A	B	C	D	E	F
2	84.6 (7.6)	78.4 (7.7)	65.5 (12.0)	72.7 (10.6)	55.7 (13.8)	39.8 (11.9)
6	98.3 (2.1)	93.0 (5.2)	86.7 (7.0)	94.1 (6.9)	86.0 (8.8)	67.0 (11.3)
10	99.3 (1.2)	97.5 (2.9)	91.5 (9.8)	98.5 (1.8)	94.3 (5.3)	79.8 (11.1)
Segregation distortion present						
	G	H	I	J	K	L
2	82.7 (7.6)	73.8 (12.8)	61.0 (12.3)	68.6 (14.1)	51.7 (14.0)	38.6 (10.5)
6	98.5 (2.2)	94.9 (4.4)	87.0 (8.0)	94.8 (4.7)	85.8 (9.1)	64.9 (13.7)
10	99.3 (1.6)	96.8 (2.7)	91.1 (5.8)	97.6 (2.5)	92.0 (6.3)	77.3 (13.3)

The means over 100 replicates for each configuration are given, with standard deviations in brackets.

a result of sampling. There was found to be less than 85% agreement between bootstrap samples without missing values or typing errors for a 2cM separation, and introducing either of these introduced more variation in the ordering.

As discussed by several earlier authors, even a low frequency of typing errors can have a substantial impact on the order and length of a linkage map. The most likely effect of a typing error is to introduce a double recombination, so that an individual's phenotype at three neighbouring loci might change from a true phenotype of AAA to ABA. This is increasingly the case as the marker density increases and the proportion of true recombinations between neighbouring markers falls. We found that map inflation was more extreme using the maximum likelihood criterion than using weighted least squares. This was also reported by Shields *et al* (1991). The advantage of the weighted least-squares approach is that the distances between markers are calculated from the map distances between all pairs of markers on a chromosome, and so the impact of typing errors on the distance between adjacent markers is less severe.

In the human genetics literature, several authors have proposed methods to test for, eliminate or model typing errors. Morton and Collins (1990) adjusted their map of human chromosome 10 by subtracting a multiple of the error rate (determined by double typing a subset of marker data) from the length of each interval. Ott (1991) expressed the apparent rate of recombination  $p$  in terms of the rate of misclassification recombinants as non-recombinants and *vice versa*,  $s$ , and the true recombination frequency,  $\theta_0$ , as

$$p = \theta_0(1 - s) + (1 - \theta_0)s$$

Shields *et al* (1991) incorporated this formulation of the apparent rate of recombination into an iterative weighted least-squares analysis and found it to give more realistic chromosome lengths for simulated data, human chromosome 10, and *Drosophila*. Lincoln and Lander (1992) worked with the idea of incomplete penetrance between the true marker genotype, and the observed marker phenotype, which may be subject to error. They included a penetrance function in their maximum likelihood analysis of simulated F2 data, and estimated a LOD score for error for each phenotype, equal to the logarithm

of the odds ratio of the likelihood of the data given a correct phenotype to the likelihood of the data given an incorrect phenotype. The phenotypes with a high LOD score for error can then be rechecked and corrected if necessary, and a final map estimated with a smaller error rate.

In plants, genetic researchers typically estimate a linkage map assuming no errors present and then look for improbable genotypes, such as those originating from double recombinations, after the map has been calculated. The module JMCHK of JoinMap (Stam and Van Ooijen, 1995) calculates, for each locus, the probability of each genotype, conditional on the genotypes at the two flanking loci and on the distances between them. Improbable genotypes can be checked against the original autoradiograms and corrected where necessary. After correction, the map should be recalculated and rechecked. An alternative approach would be to incorporate a probability of mistyping into the likelihood or weighted least-squares approach, as has been used for human populations.

We found that the maximum likelihood criterion coupled with the simulated annealing search method gave the correct order more often than using the weighted least-squares criterion with the stepwise algorithm from JoinMap. Analysis using a combination of the simulated annealing and branch-and-bound algorithms to minimise the SARF criterion gave results almost identical to the maximum likelihood criterion. This result differs from that of Shields *et al* (1991), who found weighted least squares (with or without an error probability) to recover the correct locus order more often. It seems likely that the stepwise method may not always find the optimal value of the criterion. Combination of the simulated annealing and the branch-and-bound algorithms, which do guarantee to optimise a criterion, with weighted least squares, seems a promising technique.

Missing values at the proportion investigated in this study had less effect than typing errors, but they reduced the number of correctly ordered maps, especially for a marker separation of 2cM. The variability among the order in bootstrap replicates also increases when missing values are present. There was also a tendency for missing values to lead to slightly shorter map lengths for more widely separated markers, especially in the presence of

distorted segregation ratios and/or when using weighted least squares. The presence of missing values in the marker data means that information about the number of true recombinations that have taken place along the chromosome is lost. In our simulation data, missing values were dispersed randomly throughout the data, but in experimental data the distribution is likely to be nonrandom, with some markers being difficult to score and so having more missing values than others. This may have more effect on the estimates of map order and length.

Segregation distortion at the level introduced in this study had very little effect on marker order or length. For doubled-haploid populations, segregation distortion does not affect the equation for the maximum likelihood estimate of the recombination fraction when a single locus on a chromosome causes differential viability, although problems arise in the form of biased recombination fraction estimates when more than one locus is involved (Lorieux *et al.*, 1995). Cheng *et al.* (1998) have proposed a method for inferring the presence of and mapping a locus causing distortion, which they refer to as a partial lethal-factor locus, and have applied this to the barley Steptoe–Morex population. There is scope for further research on checking the goodness of fit of a single locus affecting viability.

## Acknowledgements

This research was supported by a research grant from the UK Biotechnology and Biological Sciences Research Council and by the Scottish Executive Environment and Rural Affairs Department.

## References

- Bailey NTJ (1961). *Introduction to the Mathematical Theory of Genetic Linkage*. Clarendon Press: Oxford.
- Buetow KH (1991). Influence of aberrant observations on high-resolution linkage analysis outcomes. *Am J Hum Genet* **49**: 985–994.
- Buetow KH, Chakravarti A (1987). Multipoint gene mapping using seriation. I. General methods. *Am J Hum Genet* **41**: 180–188.
- Cheng R, Kleinhofs A, Ukai Y (1998). Method for mapping a partial lethal-factor locus on a molecular-marker linkage map of a backcross and doubled-haploid population. *Theor Appl Genet* **97**: 293–298.
- Falk CT (1989). A simple scheme for preliminary ordering of multiple loci: application to 45 CF families. In: Elston RC, Spence MA, Hodge SE, MacCluer JW (eds). *Multipoint Mapping and Linkage based upon Affected Pedigree Members*. Genetic Workshop 6, Liss: New York, pp 17–22.
- Kirkpatrick S, Gelatt CD, Vecchi MP (1983). Optimisation by simulated annealing. *Science* **220**: 671–680.
- Lander ES, Green P (1987). Construction of multilocus genetic linkage maps in human. *Proc Natl Acad Sci* **84**: 2363–2367.
- Lincoln SE, Lander ES (1992). Systematic detection of errors in genetic linkage data. *Genomics* **14**: 604–610.
- Lorieux M, Goffinet B, Perrier X, De Leon DG, Lanaud C (1995). Maximum-likelihood models for mapping genetic-markers showing segregation distortion. 1. Backcross populations. *Theor Appl Genet* **90**: 73–80.
- Lu YY, Liu BH (1995). PGRI, a Software for Plant Genome Research. Plant Genome III: San Diego, CA.
- Manninen OM (2000). Associations between anther-culture response and molecular markers on chromosomes 2H, 3H and 4H of barley (*Hordeum vulgare* L.). *Theor Appl Genet* **100**, 57–62.
- Morton NE, Collins A (1990). Standard maps of chromosome 10. *Ann Hum Genet* **54**: 235–251.
- Olson JM, Boehnke M (1990). Monte Carlo comparison of preliminary methods for ordering multiple genetic loci. *Am J Hum Genet* **47**: 470–482.
- Ott J (1991). Molecular and statistical approaches to the detection and correction of errors in genotype databases. *Am J Hum Genet* **53**: 1137–1145.
- Shields DC, Collins AK, Buetow H, Morton NE (1991). Error filtration, interference, and the human linkage map. *Proc Natl Acad Sci* **88**: 6501–6505.
- Stam P (1993). Construction of integrated genetic linkage maps by means of a new computer package: Joinmap. *Plant J* **3**: 739–744.
- Stam P, Van Ooijen JW (1995). JOINMAP. *Software for the Calculation of Genetic Linkage Maps*. Version 2.0. CPRO-DLO: Wageningen.
- Thompson EA (1987). Crossover counts and likelihood in multipoint linkage analysis. *IMA J Math Appl Med* **4**: 93–108.
- Tinker NA, Mather DE (1993). GREGOR software for genetic simulation. *J Hered* **84**: 237.
- Zivy M, Devaux P, Blaisonneaux J, Jean R, Thiellement H (1992). Segregation distortion and linkage studies in microspore-derived double haploid lines of *Hordeum vulgare* L. *Theor Appl Genet* **83**: 919–924.