Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 9824–9828, October 1995
Genetics

A maximum likelihood algorithm for genome mapping of cytogenetic loci from meiotic configuration data

MANUEL H. REYES-VALDÉS* AND DAVID M. STELLY*†

*Department of Soil and Crop Sciences, Texas A&M University, College Station, TX 77843-2474

Communicated by S. J. Peloquin, University of Wisconsin, Madison, WI, May 24, 1995

ABSTRACT **Frequencies of meiotic configurations in cytogenetic stocks are dependent on chiasma frequencies in segments defined by centromeres, breakpoints, and telomeres. The expectation maximization algorithm is proposed as a general method to perform maximum likelihood estimations of the chiasma frequencies in the intervals between such locations. The estimates can be translated via mapping functions into genetic maps of cytogenetic landmarks. One set of observational data was analyzed to exemplify application of these methods, results of which were largely concordant with other comparable data. The method was also tested by Monte Carlo simulation of frequencies of meiotic configurations from a monotelodisomic translocation heterozygote, assuming six different sample sizes. The estimate averages were always close to the values given initially to the parameters. The maximum likelihood estimation procedures can be extended readily to other kinds of cytogenetic stocks and allow the pooling of diverse cytogenetic data to collectively estimate lengths of segments, arms, and chromosomes.**

The length of a chromosomal segment separating two markers can be measured according to several criteria. One of them is physical distance, recorded in arbitrary or standard units such as microns. Another method is molecular in nature, in number of base pairs, which can be transformed to physical units to the extent that the packing ratio is known and constant. A different kind of criterion is the occurrence of crossing-over. The average number of crossovers per chromatid between two markers is the number of morgans, and, if multiplied by 100, the number of centimorgans (cM). The resulting recombination-related distances are strongly associated to physical lengths, but there is not a one-to-one functional relationship between them (1).

Historically, the first means used to measure the recombinational distance between markers was linkage analysis, using segregation of markers among progeny as the means of inference. Another means of constructing recombination-related maps is cytological—i.e., chiasmata analysis, using meiotic configurations as the means of inference. It may some day be possible to map by counting recombination nodules between molecular cytogenetic loci in latter stages of prophase I, but technical limitations currently preclude this approach.

Approaches for mapping that are cytological in nature are underdeveloped and largely underexploited. In this report we initiate an approach based on frequencies of meiotic configurations in metaphase I cells. In conjunction with the use of different cytogenetic stocks, this approach permits one to estimate the frequency of a specific segment being chiasmate—i.e., bearing at least one chiasma. The segments are defined by cytogenetic landmarks—namely, telomeres, centromeres, and breakpoints. Normal cytogenetic stocks, monotelodisomics, translocation heterozygotes, monotelodisomic translocations, tertiary monosomics, and inversions are some examples of the cytotypes that can be used.

A generalized statistical method for analyzing meiotic configurations to make inferences about chiasma frequencies and map distances can be helpful in several instances. The mapped centromeres, telomeres, and breakpoints can be used as cytogenetic landmarks for gene mapping. Other potential applications include the following: studies of genomic affinity between related varieties or species, based upon differences in chiasma frequencies; estimation of overall map lengths of chromosomes and complete genomes, to predict the number of polymorphic molecular markers needed for a saturated map; analysis of chiasma interference; analysis of the effects of different cytotypes and genotypes on the frequencies of chiasmata; evaluation of the effects of environmental factors on chiasma frequencies; and analysis of relationships between physical and genetic maps. For some types of investigation, cytologically based analyses of recombination can be much more efficient than recombination analysis based on marker segregation, since observations can be made directly from meiotic cells—i.e., obviating need for creating and growing segregating populations.

This cytological means of mapping centromeres, telomeres, and breakpoints may be preferred to the use of molecular markers in linkage analysis in several instances. For example, suitable gel-based markers or mapping populations may not be available for targeting accurately cytogenetic loci of interest; sometimes the only available polymorphic markers are from interspecific crosses, recombinational behavior of which may not reflect the intraspecific situation. Perhaps even more important is that linkage analysis with molecular markers is also usually more time-consuming.

The use of meiotic configurations for mapping is based on the following rationale. Telomeres, centromeres, and breakpoints in a given group of chromosomes define a set of segments. From this set, a given combination of chiasmate segments gives a specific configuration(s). By counting configurations, information about chiasma frequencies is obtained. However, the task is complicated by the fact that the relationship between chiasmate segments and meiotic configurations is not always one-to-one. There are configurations that can originate by more than one combination of chiasmate and achiasmate segments.

Let us define two terms we used throughout the paper. For a given cytogenetical condition ("cytotype") a configuration is "nonambiguous" if and only if it arises from a unique combination of chiasmate and achiasmate segments. In contrast, a configuration is "ambiguous" if it can arise from more than one combination of chiasmate and achiasmate segments—i.e., nonuniquely. Ambiguous configurations are at the root of the absence of an established statistical method to perform estimations using all of the information provided by meiotic configurations. As a matter of practice, there are situations in

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: cM, centimorgans; EM, expectation maximization; RFLP, restriction fragment length polymorphism.

†To whom reprint requests should be addressed.

which reliable classification of a meiotic configuration cannot be made beyond a limited set of alternatives, leading to "binning" of visually indistinguishable configurations. Special statistical techniques are required if the frequency data of ambiguous or "binned" configurations are to be used to estimate the overall frequencies of chiasmata.

A statistical approach to the use of meiotic configuration data has been discussed by Sybenga (2–4). In those works, the expected frequencies of the different configurations were equated to the observed frequencies, assuming no chiasma interference among segments. A description of a generalized and systematic algorithm to perform the calculations was not included, but the approach can be considered as moments estimation (5). However, there are typically more equations than unknowns, and an exact solution usually does not exist.

Menzel *et al.* (6) constructed a chromosome translocation breakpoint map in disomic tetraploid cotton (*Gossypium hirsutum* L., $2n = 4x = 52$) from meiotic configuration data. In samples that involved configurations of ambiguous origin, they calculated the minimum and the maximum frequencies of chiasmata possible for the segment under consideration by, respectively, attributing all or none of the ambiguous configurations to the presence (absence) of a chiasma in that segment. Also, they stated that the recombination map lengths of the six regions were slightly underestimated because the total number of cells was used as the divisor to determine frequencies. On the basis of empirical deduction, Stelly (7) inferred that the overall map length reported by Menzel *et al.* (6) must be an underestimate. The recently published recombination map of restriction fragment length polymorphism (RFLP) loci (8) confirmed that inference and indicated that the genome of cotton is at least 60% longer than Menzel *et al.* (6) estimated.

Here we present an algorithm that gives maximum likelihood estimates of the probabilities of there being at least one chiasma in a segment, for each of the different segments. Description of the algorithm can be facilitated by considering a monotelodisomic translocation heterozygote (TeNT)—e.g., a cytogenetic stock heterozygous for a reciprocal translocation involving two chromosomes and a related telosome in the same arm as the respective breakpoint (Fig. 1). The segments, as defined by the centromeres, breakpoints, and telomeres, are denoted by A, B, C, D, E, and F. Relative to the breakpoints, A and C are the opposing arms, B and D are distal segments, and E and F are the interstitial segments. For the TeNT, there are $2^5 = 32$ possible combinations of segments with or without chiasmata; segment (arm) A is hemizygous due to the telosomy and undergoes no homologous reciprocal exchange, so only five regions (B–F) can undergo crossing-over. For a euploid translocation heterozygote, there are twice as many combinations—i.e., 64—to consider.

A nonambiguous configuration can be exemplified with the chain of four chromosomes in a TeNT meiocyte, since it results only if segments B–D are chiasmate (one or more chiasmata each) and segments E and F are achiasmate (see configuration M_{17} in Table 1). An ambiguous configuration can be exemplified by a rod bivalent plus two univalents (a telosome and a normal chromosome), in that it results when one or more chiasmata occur in just one of the segments C, D, or F (see configuration M_{16} in Table 1). Binning can be illustrated by pooling configurations M_{14} and M_{16}, respective occurrences of which are not readily distinguished if unrelated chromosomes of similar size and morphology also form rod and ring bivalents. The maximum likelihood algorithm that is presented here permits use of information provided by ambiguous and binned configurations.

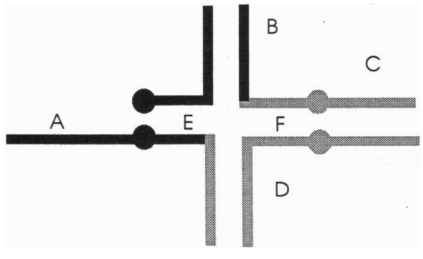

FIG. 1. Pachytene representation of a monotelodisomic translocation heterozygote (TeNT), where each line represents two chromatids, and the telosome is in the same arm as the breakpoint.

For the monotelodisomic depicted in Fig. 1, configurations M_1, M_2, M_4, M_6, M_8, M_9, and M_{16} are ambiguous. Statistically, one can opt to ignore or to use frequency data from ambiguous and binned configurations. The maximum likelihood method that is presented here permits use of these data, avoiding the possibility of bias due to the use of only a subset of the available data and reducing the variance of the estimations.

Likelihood Function

We are going to assume independence among the different segments for the occurrence of chiasma—i.e., no interference. Let $n_1, n_2, \ldots, n_t$ be the observed numbers of cells bearing the configurations $M_1, M_2, \ldots, M_t$, respectively. Let C_{ik} be the kth set (combination) of chiasmate segments giving rise to the ith meiotic configuration—i.e., the kth pattern of crossovers leading to the ith configuration. Let us state a function that indicates the presence or absence of chiasma in a given segment as follows: the indicator function of C_{ik}, denoted by $I_{Cik}(\cdot)$, has domain equal to the complete set of segments $\mathbf{S} = \{S_1, S_2, \ldots, S_s\}$, where s is the total number of distinguishable segments that can form chiasmata, and counterdomain equal

Table 1. Meiotic configurations in a TeNT, their associated combinations of segments with chiasmata, and observed and expected frequencies

Code	Configuration	Chiasmate segments	Frequency	
			Observed	Expected
M_1		BCDEF, BCEF, CDEF	0	0.17
M_2		BDEF, BEF, DEF	0	0.01
M_3		CEF	1	0.01
M_4		EF, CE, BD	2	4.14
M_5		BCDE	0	0.57
M_6		BCE, CDE	1	1.73
M_7		BDE	0	0.03
M_8		BE, DE, BF	1	0.14
M_9		B, E	0	0.49
M_{10}		BCDF	1	4.95
M_{11}		BCF	0	0.61
M_{12}		BDF	0	0.29
M_{13}		CDF	12	14.56
M_{14}		CF	5	1.79
M_{15}		DF	4	0.84
M_{16}		C, D, F	37	36.20
M_{17}		BCD	78	67.89
M_{18}		BC	6	8.32
M_{19}		CD	196	199.84
M_{20}		none	0	1.42

The telosome is in the same arm as the breakpoint.

to the set consisting of the two real numbers 0 and 1 and defined by

$$I_{C_{ik}}(S_j) = \begin{cases} 1 \text{ if } S_j \in C_{ik} \\ 0 \text{ if } S_j \notin C_{ik} \end{cases}.$$

In other words, $I_{Cik}(S_j)$ "indicates" binarily the presence (1) or absence (0) of chiasma in segment S_j for the given (ik) combination of possible chiasmate segments. On the other hand, the absence of chiasma in the given segment is indicated by

$$I_{\bar{C}_{ik}}(S_j) = 1 - I_{C_{ik}}.$$

Using the indicator function, the expected frequency of the ith configuration can be written as follows:

$$f_i = \sum_{k=1}^{m_i} \prod_{j=1}^{s} [p_j I_{C_{ik}}(S_j) + (1 - p_j) I_{\bar{C}_{ik}}(S_j)],$$

where m_i ($i = 1, 2, \ldots, t$) is the number of different possible sets (combinations) of chiasmate segments for the ith meiotic configuration, and p_j is the probability of at least one chiasma in the segment S_j. An $m_i = 1$ will define a nonambiguous configuration, whereas $m_i > 1$ will define an ambiguous configuration. For the TeNT example (Table 1), configuration M_3 is nonambiguous ($m_3 = 1$) and has an expected frequency

$$f_3 = p_C p_E p_F (1 - p_B)(1 - p_D),$$

whereas the configuration M_6 is ambiguous ($m_6 = 2$) and has an expected frequency

$$f_6 = p_B p_C p_E (1 - p_D)(1 - p_F) + p_C p_D p_E (1 - p_B)(1 - p_F).$$

For a set of N meiotic cells, the likelihood function of $\boldsymbol{p}$, the vector of probabilities of at least one chiasma, is defined by

$$L(\mathbf{p}) = \frac{N!}{n_1! n_2! \ldots n_t!} f_1^{n_1} f_2^{n_2} \ldots f_t^{n_t}.$$

The values of p_j ($j = 1, 2, \ldots, s$) in the interval (0, 1) that maximize $L(\boldsymbol{p})$ are the maximum likelihood estimates being sought.

An Iterative Algorithm

Analytic maximization of the mentioned likelihood function is quite complex in most cases, so we used an iterative numerical algorithm to solve this problem—namely, the expectation maximization (EM) algorithm described by Dempster *et al.* (9). This simple but elegant algorithm is suitable for incomplete data and has been shown to give maximum likelihood solutions. It has been used successfully in several analogous genetic situations (10–12). A good "seed value" in the EM algorithm can expedite convergence and help to avoid local maxima that do not represent the global maximum. Thus, we propose use of the following technique to obtain relatively good preliminary estimates.

We use data from specific pairs of nonambiguous configurations that differ in chiasmate condition for only a single segment. The frequency ratios of pairs of configurations can be used singly or multiply to estimate the frequency of a segment being chiasmate. For each chromosome segment S_j, define two sets of nonambiguous configurations, R_j and K_j, whose intersection is empty and with the following characteristics: (*i*) All the configurations in R_j indicate the presence of chiasma in S_j. (*ii*) For each configuration in R_j there is one and only one corresponding configuration in K_j. This element in K_j is originated by the same combination of segments with chiasmata as the related element of R_j, except for the segment S_j. There are no elements in K_j other than these. As an example for segment D in the TeNT case, the meiotic configuration M_{13} is chiasmate in segments C, D, and F and is an element of R_D, whereas M_{14} is chiasmate in segments C and F but not D and is the related element of K_D.

The numbers of cells observed to have meiotic configurations listed in sets R_j and K_j are then used to calculate the "seed value" for segment j:

$$\hat{p}_j = \frac{n_{R_j}}{n_{R_j} + n_{K_j}}.$$

This estimation states that the probability of segment j being chiasmate equals p_j, irrespective of the occurrence of chiasma in other segments. Other information about the segments in question can also be used to set up seeding information for the EM algorithm.

The first step of the EM algorithm is the E step. The seeded p_j estimates are used to calculate the expected number $E[n_{C_{ik}}]$ of each combination of chiasmate segments in ambiguous (or binned) configurations. The number of cells having the configuration is multiplied by the conditional probability of each combination of chiasmate segments given that configuration. This conditional probability is estimated as the expected frequency of that combination over the sum of the expected frequencies of all combinations giving rise to the common meiotic configuration:

$$E[n_{C_{ik}}] = n_i \frac{\prod_{j=1}^{s} [p_j I_{C_{ik}}(S_j) + (1 - p_j) I_{\bar{C}_{ik}}(S_j)]}{\sum_{z=1}^{m_i} \prod_{j=1}^{s} [p_j I_{C_{iz}}(S_j) + (1 - p_j) I_{\bar{C}_{iz}}(S_j)]}.$$

In the M step, the maximization step, the probability of a segment being chiasmate is reestimated for each segment. For calculation, the expected numbers of combinations giving rise to ambiguous configurations are temporarily taken as if they were true observations. The probability of at least one chiasma in segment S_j is estimated by adding the observed and estimated numbers of cells in which the segment was chiasmate and dividing by the total number of observations in the experiment. With this new set of estimated p_j values, the E step is started again; the cycling will stop at convergence.

Our maximum likelihood algorithm ends with the estimations of the probabilities of at least one chiasma. Recombinational map distances can then be calculated by transformation, but the final result will depend on the particular function being used and it will be considered, by functional invariance, the maximum likelihood estimate of that function. Note, however, that mapping functions may be inaccurate when applied to long distances (4) and no global mapping function can be correct everywhere (13).

Map distances can be estimated from the relationship $p_i = 1 - e^{-2\alpha}$, as $\alpha = -(1/2)\ln(1 - p)$, where p is the probability of at least one chiasma in a given interval and α is its length in morgans (2). This is the equivalent to use of Haldane's mapping function (14): $2r = 1 - e^{-2\alpha}$, where r, the recombination frequency between two markers, is equated to half the probability of chiasmate condition.

Numerical Example

To illustrate the method, consider observations on 344 cotton (*Gossypium hirsutum* L.) metaphase I meiotic cells (Table 1) from a TeNT involving the telosome for the short arm of chromosome 4 and a related reciprocal translocation involving chromosomes 4 and 5 (TT04-05, line IV2).

The seeded values used to start the EM algorithm, where the subscript of p denotes the segment, were:

$$\hat{p}_B = \frac{n_{10} + n_{11} + n_{12} + n_{17}}{(n_{10} + n_{11} + n_{12} + n_{17}) + (n_{13} + n_{14} + n_{15} + n_{19})} = 0.27,$$

$$\hat{p}_C = \frac{n_5 + n_{10} + n_{13}}{(n_5 + n_{10} + n_{13}) + (n_7 + n_{12} + n_{15})} = 0.76,$$

$$\hat{p}_D = \frac{n_{13} + n_{17}}{(n_{13} + n_{17}) + (n_{14} + n_{18})} = 0.89,$$

$$\hat{p}_E = \frac{n_3 + n_5}{(n_3 + n_5) + (n_{14} + n_{17})} = 0.01,$$

$$\hat{p}_F = \frac{n_{10} + n_{11} + n_{13}}{(n_{10} + n_{11} + n_{13}) + (n_{17} + n_{18} + n_{19})} = 0.04.$$

For the E step one must calculate the expected numbers of cells having the specific chiasma combinations that generate the ambiguous configurations. Such calculations are shown for the ambiguous configuration M_{16} (see Table 1); the subscript of n represents a combination of segments with chiasmata.

Let us define α, β, and γ as:

$$\alpha = (1 - \hat{p}_B)\hat{p}_C(1 - \hat{p}_D)(1 - \hat{p}_E)(1 - \hat{p}_F),$$

$$\beta = (1 - \hat{p}_B)(1 - \hat{p}_C)\hat{p}_D(1 - \hat{p}_E)(1 - \hat{p}_F),$$

$$\gamma = (1 - \hat{p}_B)(1 - \hat{p}_C)(1 - \hat{p}_D)(1 - \hat{p}_E)\hat{p}_F.$$

Therefore the expectations are:

$$E[n_C] = n_{16}\frac{\alpha}{\alpha + \beta + \gamma} = 10.37,$$

$$E[n_D] = n_{16}\frac{\beta}{\alpha + \beta + \gamma} = 26.49,$$

$$E[n_F] = n_{16}\frac{\gamma}{\alpha + \beta + \gamma} = 0.14.$$

Note that $E[n_C] + E[n_D] + E[n_F] = 37$.

The M step reestimates the frequency of the chiasmate condition for each segment. For each segment the summed number of observations in which that segment was chiasmate, including the expected numbers for combinations within ambiguous configurations, is divided by the total number of cells. For example, to reestimate p_F, we have

$$\hat{p}_F = (1/N)\begin{bmatrix} n_1 + n_2 + n_3 + E[n_{BF}] + n_{10} + \\ n_{11} + n_{12} + n_{13} + n_{14} + n_{15} + E[n_F] \end{bmatrix} = 0.07.$$

The E step is reexecuted with the new estimates. The results for five cycles are summarized in Table 2 including the log-likelihood in each cycle (without the logarithm of the multinomial constant).

After 20 cycles, the estimation exhibited convergence, at values of 0.254 for p_B, 0.945 for p_C, 0.891 for p_D, 0.008 for p_E, and 0.068 for p_F. The vector of first derivatives of the log-likelihood evaluated on our estimates was (0, 0, 0, 0, 0) with 12-digit accuracy, which is a good indication that the point is indeed a maximum. The inverse of the information matrix was used to estimate standard errors according to the maximum likelihood theory. The procedure will be presented elsewhere; meanwhile the reader is referred to basic texts (15, 16). The estimates were 0.024 for S_{p_B}, 0.019 for S_{p_C}, 0.023 for S_{p_D}, 0.005 for S_{p_E}, and 0.014 for S_{p_F}.

Using the transformation proposed by Sybenga (2), which assumes a Poisson distribution of chiasmata, we estimated map lengths of 14.6 cM for region B, 145.2 cM for region C, 110.7 cM for region D, 0.4 cM for region E, and 3.5 cM for region F. Our estimates of the lengths of chromosome 5 segments C and D are more than twice the lengths reported graphically by Menzel *et al.* (6); however, there is agreement when compared to our nontransformed estimates. The difference is that they did not use any mapping function. Our results and those of Menzel *et al.* (6) concordantly indicated that interstitial segments E (chromosome 4) and F (chromosome 5) are recombinationally very short. A significant disparity seems to exist for the short arm of chromosome 4 (segments B and E), which Menzel *et al.* (6) estimated to be 50 cM, whereas we estimated it to be 15 cM. The discordance could reflect effects of the telosomy.

There is a very good agreement between our estimations and the recently published RFLP map of cotton by Reinisch *et al.* (8). From our results the length of chromosome 5 can be estimated as the sum of the lengths for segments C, D, and F—i.e., 259.4 cM, which is close to the reported map length of 244.3 cM. Due to the use of the telosome, the overall length of chromosome 4 cannot be estimated from our Te4sh NT4-5 cytogenetic data alone, and the RFLP loci have not yet been placed relative to the centromeres.

In this example we tested the algorithm with 10 random guesses and the estimations always converged to the same point, so it seems that good seeding may be dispensable in some instances—e.g., when we have only a few ambiguous types.

Expected frequencies of the different configurations calculated with our estimations are presented in Table 1. Appreciable excesses of M_{14} and M_{17} configurations were found. In addition, the observed frequency of M_{10} was less than expected. Such departures do not collectively indicate failure of the assumption of no interference among segments, because configuration M_{14} originates by crossing-over in adjacent segments, whereas configuration M_{17} originates by crossing-over in nonadjacent segments. Configuration M_{10} is originated by crossing-over in both, adjacent and nonadjacent segments. A goodness-of-fit test was done with the likelihood ratio (5), and a significant departure ($P < 0.01$) from expectation was found. Misclassifications could have led to these results.

We also applied the algorithm to some data from rye, *Secale cereale*, that involved binning, and allowed comparison to the method of Sybenga (2). No good seeding values were available, so we used 60 different sets of random "seeds" to start the EM algorithm. In every instance, convergence occurred at one of two different local maxima. Although the maximum likelihood estimates differed from the moments estimates reported by Sybenga (2), his estimates corresponded to those placed in the lower local maximum of the likelihood function. We did not find significant differences between the observed frequencies of meiotic configurations and the expected ones generated by our maximum likelihood method.

Simulation Results

To test the method further, we generated sets of observations by Monte Carlo simulation for a TeNT with the telosome in the same arm as the breakpoint. For each set, we estimated the frequency of segments being chiasmate. A total of 300 simulations were made for each of six sample sizes—namely, 50, 100, 150, 200, 250, and 300 cells. The true frequencies for the segments were fixed at 0.3 for p_B, 0.7 for p_C, 0.9 for p_D, 0.2 for p_E, and 0.1 for p_F. For each sample size we calculated the average estimations and their standard deviations (Table 3). Given the considerable number of simulations, these standard deviations provided good estimations of standard errors.

The average estimations were always very close to the true values. In one case the two-sided t test (5) revealed a statistically significant departure of the mean from the true param-

Table 2. Iterative estimation of chiasma frequencies with the EM algorithm

Cycle	Segment					Log-likelihood
	B	C	D	E	F	
0	0.270	0.760	0.890	0.010	0.040	−496.840
1	0.253	0.902	0.934	0.008	0.068	−465.673
2	0.253	0.915	0.922	0.008	0.068	−464.481
3	0.253	0.923	0.913	0.008	0.068	−463.875
4	0.254	0.930	0.906	0.008	0.068	−463.559
5	0.254	0.934	0.902	0.008	0.068	−463.393

Results for five cycles are given.

eter. Nevertheless, the deviation was not big from a practical standpoint and, since only 1 case in 36 exhibited a significant departure, this could have occurred by chance. Therefore, one cannot conclude that the method gives biased estimations.

Extensions and Limitations of the Algorithm

In situations when binning is used, a new pooled class is created by the union of two or more possible meiotic configurations, usually ones that are visually difficult to distinguish. To avoid potential bias, no meiotic configuration that belongs to a pooled class can belong to any other class. Otherwise, differential representation of component configurations could result, which would lead to erroneous estimation in the expectation step—i.e., $E[n_{C_{ik}}]$.

Since the assumption in the model is that the frequencies of chiasmata conditions among segments are statistically independent, large departures from independence could give rise to considerable error in estimations. Such errors would be especially likely if there were a high frequency of ambiguous or binned configurations. Errors could also arise if the cytogenetic condition affects chiasma frequencies. However, the method described here can be used to detect such an effect. Finally, as in recombination mapping, additional assumptions are entailed by whatever mapping function is used to transform the algorithm-derived estimates of p_j (frequencies of the chiasmate conditions) into the average number of chiasmata per segment and centimorgans.

Table 3. Statistics from 300 replications of simulations for a monotelodisomic translocation, with different sample sizes (N) and fixed frequencies of at least one chiasma of 0.3 for segment B, 0.7 for C, 0.9 for D, 0.2 for E, and 0.1 for F

Parameter	Segment				
	B	C	D	E	F
N = 50					
Mean	0.294	0.709	0.882*	0.206	0.104
SD	0.074	0.087	0.080	0.063	0.042
N = 100					
Mean	0.299	0.701	0.902	0.201	0.101
SD	0.049	0.056	0.046	0.045	0.029
N = 150					
Mean	0.298	0.701	0.898	0.198	0.099
SD	0.041	0.043	0.038	0.036	0.023
N = 200					
Mean	0.299	0.701	0.901	0.201	0.101
SD	0.034	0.038	0.031	0.032	0.022
N = 250					
Mean	0.302	0.699	0.898	0.199	0.101
SD	0.028	0.034	0.029	0.026	0.020
N = 300					
Mean	0.300	0.701	0.900	0.201	0.102
SD	0.028	0.030	0.025	0.025	0.018

*, $P < 0.01$.

The algorithm can be extended to cases where ambiguities arise from uncertainty during cytogenetic classification. In fact, the EM algorithm will help most when the frequency of ambiguous or binned configurations is high. The method can be applied coherently across multiple cytogenetic types affecting the same chromosome segments. For example, it would be feasible to pool observations from a translocation heterozygote and a related tertiary monosomic or duplication-deficient, to generate a collective estimation of chiasma frequencies in each segment.

In situ hybridization to meiotic chromatin enhances the amount of information provided by metaphase I configurations (17). Moreover, applicability of the algorithm described here is empowered by extension to molecular meiotic data, as we will discuss elsewhere (unpublished).

The following persons contributed valuable suggestions to the present work: Charles F. Crane and Jeremy F. Taylor from the Departments of Soil and Crop Sciences, and Animal Science, respectively, of Texas A&M University. This work was supported by funds provided by the Mexican Council of Science and Technology (CONACYT), U.S. Department of Agriculture Plant Genome Grant 92-37300-7655, and the Texas Agricultural Experimental Station. The programing language used for calculations and simulations was MATHEMATICA (student version) from Wolfram Research, Inc.

1. Wagner, R. P., Maguire, M. P. & Stallings, R. L. (1993) *Chromosomes: A Synthesis* (Wiley-Liss, New York), pp. 349–374.
2. Sybenga, J. (1966) *Genetica* **37,** 199–206.
3. Sybenga, J. (1975) *Meiotic Configurations: A Source of Information for Estimating Genetic Parameters* (Springer, New York), pp. 93–100.
4. Sybenga, J. (1992) *Cytogenetics in Plant Breeding* (Springer, New York), pp. 239–243.
5. Larsen, R. J. & Marx, M. L. (1986) *An Introduction to Mathematical Statistics and Its Applications* (Prentice-Hall, Englewood Cliffs, NJ), pp. 267, 268, 305–308, 347–350.
6. Menzel, M. Y., Richmond, K. L. & Dougherty, B. J. (1985) *J. Hered.* **76,** 406–414.
7. Stelly, D. M. (1993) *Proc. Beltwide Cotton Conf.*, 1545–1550.
8. Reinisch, A. J., Dong, J., Brubaker, C. L., Stelly, D. M., Wendel, J. F. & Paterson, A. H. (1994) *Genetics* **138,** 829–847.
9. Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977) *J. R. Stat. Soc. B.* **39,** 1–38.
10. Donis-Keller, H., Green, P., Helms, C., Cartinhour, S., Weiffenabach, B., *et al.* (1987) *Cell* **51,** 319–337.
11. Lander, E. S. & Green, P. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 2363–2367.
12. Luo, Z. W. & Woolliams, J. A. (1993) *Heredity* **70,** 245–253.
13. Crow, J. F. (1990) *Genetics* **125,** 669–671.
14. Haldane, J. B. S. (1919) *J. Genet.* **8,** 299–309.
15. Crow, J. F. & Kimura, M. (1970) *An Introduction to Population Genetics Theory* (Harper & Row, New York), pp. 509–514.
16. Edwards, A. W. F. (1992) *Likelihood* (Johns Hopkins Univ. Press, Baltimore), pp. 144–160.
17. Crane, C. F., Price, H. J., Stelly, D. M. & Czeschin, D. G. (1993) *Genome* **36,** 1015–1021.