Multilocus Analysis for Gene-Centromere Mapping Using First Polar Bodies and Secondary Oocytes

Yang Da, Vickie L. Jarrell, Tianlin Wang, Rohan L. Fernando, Matthew B. Wheeler and Harris A. Lewin

Department of Animal Sciences, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

Manuscript received June 2, 1994
Accepted for publication October 25, 1994

ABSTRACT

Polar body and oocyte typing is a new technique for gene-centromere mapping and for generating female linkage maps. A maximum likelihood approach is presented for ordering multiple markers relative to the centromere and for estimating recombination frequencies between markers and between the centromere and marker loci. Three marker-centromere orders are possible for each pair of markers: two orders when the centromere flanks the two markers and one order when the centromere is flanked by the two markers. For each possible order, the likelihood was expressed as a function of recombination frequencies for two adjacent intervals. LOD score for recombination frequency between markers or between the centromere and a marker locus was derived based on the likelihood for each gene-centromere order. The methods developed herein provide a general solution to the problem of multilocus gene-centromere mapping that involves all theoretical crossover possibilities, including four-strand double crossovers.

GENETIC linkage maps are used for mapping genes controlling monogenic and polygenic traits (LANDER and BOTSTEIN 1989). Detailed linkage maps are also a prerequisite for positional cloning strategies aimed at isolating genes of physiological, medical and agricultural interest (SCHOOK *et al.* 1991). Sex-differences in genetic maps may complicate the problem of positional cloning because linkage maps are based on recombination rate, which varies across the genome for males and females (DUNN and BENNET *et al.* 1967). Powerful new methods for meiotic mapping, such as sperm typing (CUI *et al.* 1989) and gene-centromere mapping employing first polar bodies (PB1) and secondary oocytes (SO; CUI *et al.* 1992), make it feasible to construct highly accurate ordered gene maps for any mammalian species. In addition to assisting positional cloning strategies, accurate sex-specific maps can be useful for predicting disease transmission, prediction of genotypes at quantitative trait loci linked to genetic markers and for elucidating the mechanisms of recombination in males and females.

Gene-centromere maps can be developed from the analysis of families that contain autosomal trisomies (CHARKRAVARTI *et al.* 1989), the analysis of non-disjunction in ovarian teratomas (OTT *et al.* 1976; EPPIG and EICHER 1983) and from linkage analysis in families when a centromeric marker is available (MATHER 1938). However, the biological material required to perform gene-centromere mapping using teratomas and trisomies is relatively rare and few polymorphisms at the centromere have been described. Recently, CUI and coworkers (1992) described a broadly applicable method for gene-centromere mapping that utilizes the polymerase chain reaction (PCR) to analyze the products of meiosis I in individual SO. Although statistical methods for analysis of single-locus gene-centromere mapping data have been described (MATHER 1938; CÔTÉ and EDWARDS 1975; OTT *et al.* 1976; SNOW 1979; MORTON 1982; CHAKRAVARTI and SLAUGENHAUPT 1987), a method for multilocus linkage analysis, including ordering with respect to the centromere, is not available for analysis of PB1 and SO typing data.

The ability to locate the position of the centromere by PB1 and SO typing is a distinct advantage of gene-centromere mapping compared with sperm typing or standard linkage mapping in families. Furthermore, multilocus PB1 and SO genotypes contain genetic information on all products of a single meiosis, analogous to tetrad analysis used to study recombination in *Neurospora*. A method for multilocus analysis of PB1 and SO typing data would permit simultaneous mapping of multiple genes relative to the centromere and the identification and confirmation of rare recombination events, such as four strand double crossover (FSDC; ROTHWELL 1988) that cannot be identified with male gametes. The purpose of this study was to develop a maximum likelihood approach for ordering multiple markers on the same or different chromosome arms relative to the centromere and to estimate recombination frequencies using PB1 and SO typing data. The method developed provides a general solution to

Corresponding author: Harris A. Lewin, 206 Plant and Animal Biotechnology Laboratory, 1201 W. Gregory Dr., University of Illinois, Urbana, IL 61801.

multilocus gene-centromere mapping and a powerful tool for developing genetic maps in species with underdeveloped genetic maps (LEWIN *et al.* 1990).

THEORY AND METHODS

Observations for two genetic markers when the centromere flanks the markers: After the completion of meiosis I, a primary oocyte (PO) divides into a PB1 and SO, each of which contains two sister chromatids joined at the centromere. For genotyping, the PB1 is separated from the SO by micromanipulation, and locus-specific typing using the PCR then is performed, resulting in observations on pairs of alleles for each locus (CUI *et al.* 1992). To order two marker loci relative to the centromere, the female must be heterozygous at both marker loci. For two marker loci *A* and *B*, with alleles *A*, *a* and *B*, *b*, nine genotypes of the PB1 and SO are possible: *AABB, aabb, aaBB, AAbb, AABb, aaBb, AaBB, Aabb* and *AaBb*. After normal meiotic division, a PB1 and its matched SO have complementary genotypes. When the centromere flanks the two markers, assuming normal chromosome disjunction during meiosis, the nine possible genotypes of the PB1 and its matched SO can be grouped into five categories. They are

$$T_1 = AABB \text{ or } aabb$$

$$T_2 = AAbb \text{ or } aaBB$$

$$T_3 = AABb \text{ or } aaBb$$

$$T_4 = AaBB \text{ or } Aabb$$

$$T_5 = \text{AaBb}$$

Assuming the order *centromere-A-B* or *centromere-B-A*, T_1 is the parental type, *i.e.*, no recombination between the two markers or between the markers and the centromere, T_2 is a result of FSDC, T_3 is a result of recombination between marker loci *A* and *B* (Figure 1), T_4 is a result of recombination between marker loci *A* and *B* given recombination between the centromere and marker locus *A* and T_5 can be a result of recombination or nonrecombination between marker loci *A* and *B* (Figure 2).

Each complementary PB1 and SO pair should be considered as one observation, because they are the result of the same meiotic event. When typing efficiency is not 100%, typing both PB1 and SO may increase the number of observations and provide mutual verification of the same meiotic event.

Likelihood analysis when the centromere flanks two markers: The likelihood for the order *o-A-B* or *o-B-A*, where *o* denotes the centromere, can be expressed as

$$L_1 = \prod_{i=1}^{5} (p_i^{n_i}) \qquad (1)$$

where n_i = number of observations for genotype categories T_i and p_i = probability of T_i, $i = 1, \ldots, 5$.

The probability of T*i* in Equation 1 can be determined based on the probability of observing T_i given the genotype of the marker proximal to the centromere (MPC) and the marginal probability of the MPC being a particular genotype, *i.e.*,

$$p_i = \text{Prob}(T_i = \sum_G \text{Prob}(T_i|\text{MPC} = G)\,\text{Prob}(\text{MPC} = G)$$

where *G* denotes the genotype of the MPC.

For each T_i, the MPC can be either homozygous or heterozygous. Thus, the above formula reduces to

$$p_i = \text{Prob}(T_i|\text{MPC} = G) \times \text{Prob}(\text{MPC} = G)$$

The probability of homozygous or heterozygous PB1 and SO, Prob(MPC = *G*), can be expressed using the formula to relate heterozygosity with recombination frequency in MORTON (1982) and CHAKRAVARTI and SLAUGENHAUPT (1987):

$$\text{Prob}(\text{MPC} = AA \text{ or } aa) = \lambda_{oa1} = 1 - {}^2/_3\,[1 - (1 - 2\theta_{oA})^{3/2}] \qquad (2)$$

$$\text{Prob}(\text{MPC} = Aa) = \lambda_{oa2} = {}^2/_3\,[1 - (1 - 2\theta_{oA})^{3/2}] \qquad (3)$$

where θ_{oA} = recombination frequency between the centromere and locus *A* (the MPC).

To express the likelihood functions as relatively simple functions of recombination frequencies, the relationship $\text{Prob}(T_i|\text{MPC} = G) = \Sigma k \text{Prob}(T_i|\text{MPC} = G, k)\text{Prob}(k)$ will be used. Then, the probability of T_i can be expressed as:

$$p_i = [\textstyle\sum_{k=0}^{\infty} \text{Prob}(T_i|\text{MPC} = G, k)\,\text{Prob}(k)] \times \text{Prob}(\text{MPC} = G) \qquad (4)$$

where *k* is the number of crossovers.

The conditional probability of a PB1 or SO genotype given the MPC genotype and the number of crossovers, $\text{Prob}(T_i|\text{MPC} = G, k)$, can be obtained by considering the consequences of crossovers during the first meiotic division.

The possible outcomes of up to two crossovers between two markers are described in Figures 1 and 2. Assuming the centromere flanks the two marker loci, *A* and *B*, then either T_1 or T_2 can be the parental genotype of the two markers. In this study, T_1 is assumed to be the parental genotype at all intervals, *i.e.*, no recombination between the centromere and any marker or between markers. Two cases regarding the MPC will be presented; whether the MPC of a PB1 or SO is homozygous (Figure 1) or heterozygous (Figure 2). For the order *o-A-B*, marker *A* is the MPC. When the MPC (locus *A*) is homozygous, the first crossover between marker loci *A* and *B* results in the recombinant genotype T_3. A second crossover between marker loci *A* and *B* for parental genotype T_1 (or one crossover for T_3) results in three possible PB1 or SO genotypes, T_1, T_3 and T_2, with proportions 1:2:1. Note that T_2 is a result of FSDC between homologous chromosomes and is the only observable product of two or more crossovers between two markers (Figure 1). When the MPC (locus *A*) is heterozygous, the first crossover between loci *A* and *B* results in two possible PB1 or SO genotypes with equal frequencies: T_4 and T_5. For a PO with a T_5 genotype, a second crossover results in the same array of genotypes in the PB1 and SO as is expected after one crossover, 1(T_4):1(T_5). However, after a second crossover, a PO with genotype T_4 is converted to a T_5 genotype that has three possible arrangements of alleles, 1(*AB*/*ab* and *AB*/*ab*):2(*AB*/*ab* or *Ab*/*aB*):1(*Ab*/*aB* and *Ab*/*aB*), with the latter having an unobservable FSDC between loci *A* and *B* (Figure 2). From Figures 1 and 2, probabilities of genotypes T_3 and T_5 in PB1 and SO given *k* crossovers (abbreviated as P_k) satisfy $P_k = 1 - {}^1/_2 P_{k-1}$, as in MATHER (1938) and CHAKRAVARTI and SLAUGENHAUPT (1987), with initial value of 0 for T_3 and initial value of 1 for T_5. The probability of genotype T_4 is simply 1 − Prob T_5. The probabilities of genotypes T_1 and T_2 differ only when the number of crossovers is 0, because only PO with genotype T_3 can yield PB1 and SO with genotypes T_1 and T_2 and does so with equal frequency. Primary oocytes with genotypes T_1 and T_2 convert to T_3 after one crossover between the two marker loci. Therefore, the following general formulations can be derived:

$$\text{Prob}(T_1|\text{MPC} = AA \text{ or } aa;\, k) = {}^1/_2\{1 - {}^2/_3[1 - (-{}^1/_2)^k]\} + \delta \qquad (5)$$

$$\text{Prob}(T_2|\text{MPC} = AA \text{ or } aa;\ k) = {}^1/_2[1 - {}^2/_3[1 - (-{}^1/_2)^k]] - \delta \quad (6)$$

$$\text{Prob}(T_3|\text{MPC} = AA \text{ or } aa;\ k) = {}^2/_3[1 - (-{}^1/_2)^k] \quad (7)$$

$$\text{Prob}(T_4|\text{MPC} = Aa;\ k) = 1 - {}^2/_3[1 - ({}^1/_2)^{k+1}] \quad (8)$$

$$\text{Prob}(T_5|\text{MPC} = Aa;\ k) = {}^2/_3[1 - (-{}^1/_2)^{k+1}] \quad (9)$$

where $\delta = {}^1/_2$ if $k = 0$; and $\delta = 0$ otherwise.

Following CHAKRAVARTI and SLAUGENHAUPT (1987), chiasma and chromatid interference are assumed absent, and the number of crossovers is assumed to follow a Poisson distribution:

$$\text{Prob}(k) = e^{-2x}\frac{(2x)^k}{k!} \quad (10)$$

where $x = -{}^1/_2 \ln(1 - 2\theta)$, which is the Haldane map distance (HALDANE 1919), and θ = recombination frequency.

Then, using Equations 5–10, and the relationship $\text{Prob}(T_i|\text{MPC}) = \Sigma_k \text{Prob}(T_i|\text{MPC}, k)\,\text{Prob}(k)$, the conditional probabilities of T_i given the MPC are:

$$\text{Prob}(T_1|\text{MPC} = AA \text{ or } aa) = \lambda_{ab1} = {}^2/_3 - \theta_{AB} + {}^1/_3(1 - 2\theta_{AB})^{3/2} \quad (11)$$

$$\text{Prob}(T_2|\text{MPC} = AA \text{ or } aa) = \lambda_{AB2} = -{}^1/_3 + \theta_{AB} + {}^1/_3(1 - 2\theta_{AB})^{3/2} \quad (12)$$

$$\text{Prob}(T_3|\text{MPC} = AA \text{ or } aa) = \lambda_{ab3} = {}^2/_3[1 - (1 - 2\theta_{AB})^{3/2}] \quad (13)$$

$$\text{Prob}(T_4|\text{MPC} = Aa) = \lambda_{ab4} = {}^1/_3[1 - (1 - 2\theta_{AB})^{3/2}] \quad (14)$$

$$\text{Prob}(T_5|\text{MPC} = Aa) = \lambda_{ab5} = {}^2/_3[1 + {}^1/_2(1 - 2\theta_{AB})^{3/2}] \quad (15)$$

where θ_{AB} = recombination frequency between marker loci *A* and *B*.

Substituting Equations 2 and 3 and 11–15 into Equation 4, the probabilities of T_i given the order *o-A-B* are as follows:

$$\text{Prob}(T_1) = p_1 = \lambda_{oa1}\lambda_{ab1} \quad (16)$$

$$\text{Prob}(T_2) = p_2 = \lambda_{oa1}\lambda_{ab2} \quad (17)$$

$$\text{Prob}(T_3) = p_3 = \lambda_{oa1}\lambda_{ab3} \quad (18)$$

$$\text{Prob}(T_4) = p_4 = \lambda_{oa2}\lambda_{ab4} \quad (19)$$

$$\text{Prob}(T_5) = p_5 = \lambda_{oa2}\lambda_{ab5} \quad (20)$$

Substituting the p_i values given by Equations 16–20 into Equation 1 yields the likelihood for observations of T_i in the sample for the order *o-A-B*. For the order *o-B-A*, the likelihood differs from that for the order *o-A-B* in the definition of the MPC, *i.e.*, locus *B* is now the MPC, instead of locus *A*. As a result of reversing the order of loci *A* and *B*, the genotypes T_3 and T_4 given the order *o-A-B* now must be redefined as T_4 and T_3; alternatively, n_3 and n_4 given the order *o-A-B* must be redefined as n_4 and n_3. Then, the probabilities of T_i are obtained by replacing λ_{oai} in Equations 11–15 with λ_{obi}, *i.e.*,

$$p_1 = \lambda_{ob1}\lambda_{ab1} \quad (21)$$

$$p_2 = \lambda_{ob1}\lambda_{ab2} \quad (22)$$

$$p_3 = \lambda_{ob2}\lambda_{ab3} \quad (23)$$

$$p_4 = \lambda_{ob1}\lambda_{ab4} \quad (24)$$

$$p_5 = \lambda_{ob2}\lambda_{ab5} \quad (25)$$

where λ_{obi} has the same expression as λ_{oai} in (2) and (3) except that subscript *a* is replaced by subscript *b*. Substituting Equations 21–25 yields the likelihood of observing all values of T_i in the sample given the order *o-B-A*.

Likelihood analysis when the centromere is flanked by two markers: For the order *A-o-B*, all crossovers between marker loci *A* and *B* are the result of crossovers between the centromere and the marker loci. For this order, FSDC between marker loci *A* and *B* can be distinguished only when both loci are homozygous and the FSDC occurs between one locus and the centromere. Given normal disjunction during meiosis, the expected genotypes of PB1 or SO fall into six categories: S_1, both loci are homozygous but FSDC between the two loci can not be discerned; S_2, both loci are homozygous and a FSDC occurs between locus *A* and the centromere; S_3, both loci are homozygous and a FSDC occurs between locus *B* and the centromere; S_4, locus *A* is homozygous and locus *B* is heterozygous; S_5, locus *A* is heterozygous and locus *B* is homozygous; and S_6, both loci are heterozygous. Then, the probabilities of PB1 or SO genotypes for the order *A-o-B* are as follows:

$$\text{Prob}(S_1) = q_1 = \lambda_{a1}\lambda_{b1} + \lambda_{a2}\lambda_{b2} \quad (26)$$

$$\text{Prob}(S_2) = q_2 = \lambda_{a2}\lambda_{b1} \quad (27)$$

$$\text{Prob}(S_3) = q_3 = \lambda_{a1}\lambda_{b2} \quad (28)$$

$$\text{Prob}(S_4) = q_4 = (\lambda_{a1} + \lambda_{a2})\lambda_{b3} \quad (29)$$

$$\text{Prob}(S_5) = q_5 = \lambda_{a3}(\lambda_{b1} + \lambda_{b2}) \quad (30)$$

$$\text{Prob}(S_6) = q_6 = \lambda_{a3}\lambda_{b3} \quad (31)$$

where

$$\lambda_{a1} = {}^2/_3 - \theta_{oA} + {}^1/_3(1 - 2\theta_{oA})^{3/2}$$

$$\lambda_{a2} = -{}^1/_3 + \theta_{oA} + {}^1/_3(1 - 2\theta_{oA})^{3/2}$$

$$\lambda_{a3} = {}^2/_3[1 - (1 - 2\theta_{oA})^{3/2}]$$

and λ_{bi} is expressed in the same way as λ_{ai} except that subscript *a* in λ_{ai} is replaced by *b* and θ_{oA} is replaced by θ_{oB}.

Then, the likelihood for observed PB1 or SO genotypes when the centromere is inside of two markers is

$$L_2 = \prod_{i=1}^{6}(q_i^{n_i}) \quad (32)$$

where n_i = number of observations consistent with genotype category S_i.

For the order *A-o-B*, the location of FSDC cannot be obtained directly from the typing data. Theoretically, the exact location of a FSDC could be determined by comparing likelihoods for all possible locations and selecting the location with the largest likelihood as the most probable location of occurrence. For example, if the data are consistent with two FSDCs between loci *A* and *B*, three locations of the crossovers are possible: between the centromere and locus *A* only, between the centromere and locus *B* only, or one FSDC between the centromere and locus *A* and one FSDC between the centromere and locus *B*. The site(s) with the largest likelihood should be considered the most likely site(s) of recombination, if the true order is *A-o-B*. However, as will be shown by numerical examples, finding the exact location of the FSDCs for the order *A-o-B* may have only limited success.

Ordering marker loci relative to the centromere: The most probable order of markers relative to the centromere can be determined by comparing likelihoods for all possible orders using likelihoods (1) and (32). A numerical method must be used to maximize likelihoods (1) and (32) because they do not have analytical solutions. The order with the largest likelihood is considered as the most likely order. To order

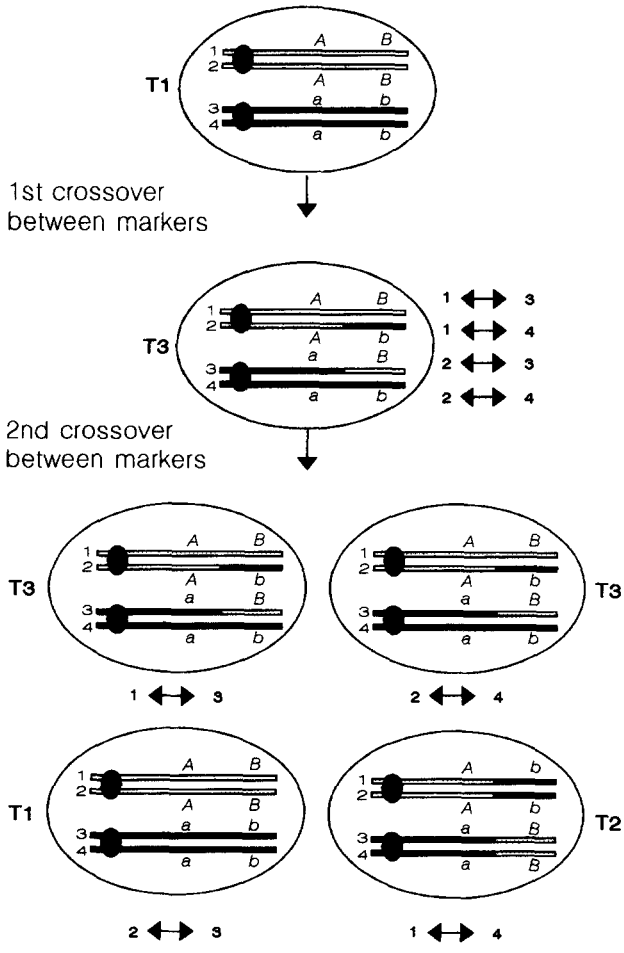

FIGURE 1.—Outcomes of two crossovers between marker loci *A* and *B* when the centromere flanks the two markers, and the MPC (locus *A*) is homozygous. Each circle indicates a PO undergoing meiosis I. T1, T2 and T3 indicate the genotype class of the PO and the resulting PB1 and SO at the completion of meiosis I. The two-way arrow (↔) indicates crossover between the two numbered strands. A single crossover may involve one chromatid of each homologous chromosome. The resulting PB1 or SO genotype is *AABb* or *aaBb*, and 50% of the chromatids are recombinant for the two loci *A* and *B*. A second crossover may occur between any two strands of the homologues and results in one of three possible PB1 or SO genotypes, depending on which two chromatids are involved. The three PB1 or SO genotypes are 1 (*AABB* or *aabb*) :2 (*AABb* or *aaBb*) :1 (*AAbb* or *aaBB*). The category *AAbb* or *aaBB* represents a FSDC between loci *A* and *B*. The expected frequency of recombinant strands after two crossovers between loci *A* and *B* is $0(^1/_4) + (^1/_2)\ ^1/_2 + (1)\ ^1/_4 = ^1/_2$.

more than two markers, markers can be ordered two at a time using methods presented above, and then the order of all marker loci relative to the centromere can be determined using a three-locus strategy (BOEHNKE *et al.* 1989) or the trisection strategy (GORADIA and LANGE 1990).

Estimation of recombination frequencies: For a single locus analysis, genotypes of PB1 or SO for a marker locus (denoted locus *A*) have two categories, homozygous and heterozygous. Let m_1 = number of homozygotes and m_2 = number of heterozygotes, the likelihood of observations for one locus is

$$L_3 = \lambda_{oa1}^{m_1} \lambda_{oa2}^{m_2} \qquad (33)$$

The maximum likelihood estimate for the recombination frequency between a marker locus and the centromere (θ_{oA}) obtained by maximizing likelihood (33) is

$$\theta_{oA} = {}^1/_2[1 - (1 - {}^3/_2 y)^{2/3}] \qquad (34)$$

where y = observed frequency of heterozygous PB1 or SO for the marker locus. It should be noted that Equation 34 is the inverse function of Equation 3, which originates from MORTON (1982) and CHAKRAVARTI and SLAUGENHAUPT (1987).

The LOD score for recombination frequency between a marker locus and the centromere (Z_{oA}) is

$$Z_{oA} = m_1 \log (3\lambda_{oa1}) + m_2 \log ({}^3/_2\ \lambda_{oa2}) \qquad (35)$$

where log is the common logarithm (base 10).

Equations 34 and 35 also apply to the estimation of θ_{oA} for the MPC based on likelihood (1), because the marginal likelihood for the MPC in (1) is the same as (33). When the estimation of θ_{oA} is based on likelihood (1), $m_1 = n_1 + n_2 + n_3$ and $m_2 = n_4 + n_5$, where n_i is defined the same as for likelihood (1). However, Equations 34 and 35 do not apply to the estimation of θ_{oA} using likelihood (32) for the order *A-o-B*. For this case, a numerical method must be used to obtain the estimates of θ_{oA} and θ_{oB}. The marginal LOD score for each locus is unavailable from likelihood (32) but a generalized LOD score (Z_G) can be defined for both flanking loci (*A* and *B*) jointly:

$$Z_G = \log L_2 + n_1 \log (18) + (n_2 + n_3) \log (36) - (n_4 + n_5) \log ({}^2/_9) - n_6 \log ({}^4/_9)$$

For two marker loci *A* and *B*, the maximum likelihood estimate of recombination frequency is obtained by maximizing likelihood (1) with respect to θ_{AB}, and a numerical method must be used. The LOD score for θ_{AB} can be defined using the conditional likelihood for marker loci *A* and *B* in likelihood (1), *i.e.*,

$$Z_{AB} = (n_1 + n_2) \log (6) + n_4 \log (3) - (n_3 + n_5) \log ({}^2/_3) + \sum_{i=1}^{5} n_i \log (\lambda_{abi})$$

NUMERICAL EXAMPLES

Example 1: The purpose of this example is to show the likelihood analysis for identifying the order *A-o-B*. Recombination frequencies were assumed as $\theta_{oA} = 0.125$, and $\theta_{oB} = 0.220$. A total of 200 phase-known genotype determinations on PB1 or SO was assumed. Then, the number of observations for S_i, $i = 1, \ldots, 6$, was generated using (32) and the generated number was rounded to its nearest integer. The generated PB1 or SO genotypes were $89S_1 + 1S_2 + 4S_3 + 59S_4 + 29S_5 + 18S_6$. Likelihoods (32) and (1) were applied to estimate recombination frequencies and to calculate the log-likelihood (base 10). For the order *A-o-B*, six analyses were performed to identify the probable locations of FSDCs (Table 1). The evidence for location was not strong; the ratio of two adjacent likelihoods was less than 10:1. This is not surprising, considering the fact that the data did not have information for locating FSDCs, and the change in the number of observations for the FSDCs was small, *i.e.*, an increase of one observation for one interval and one less observation for the

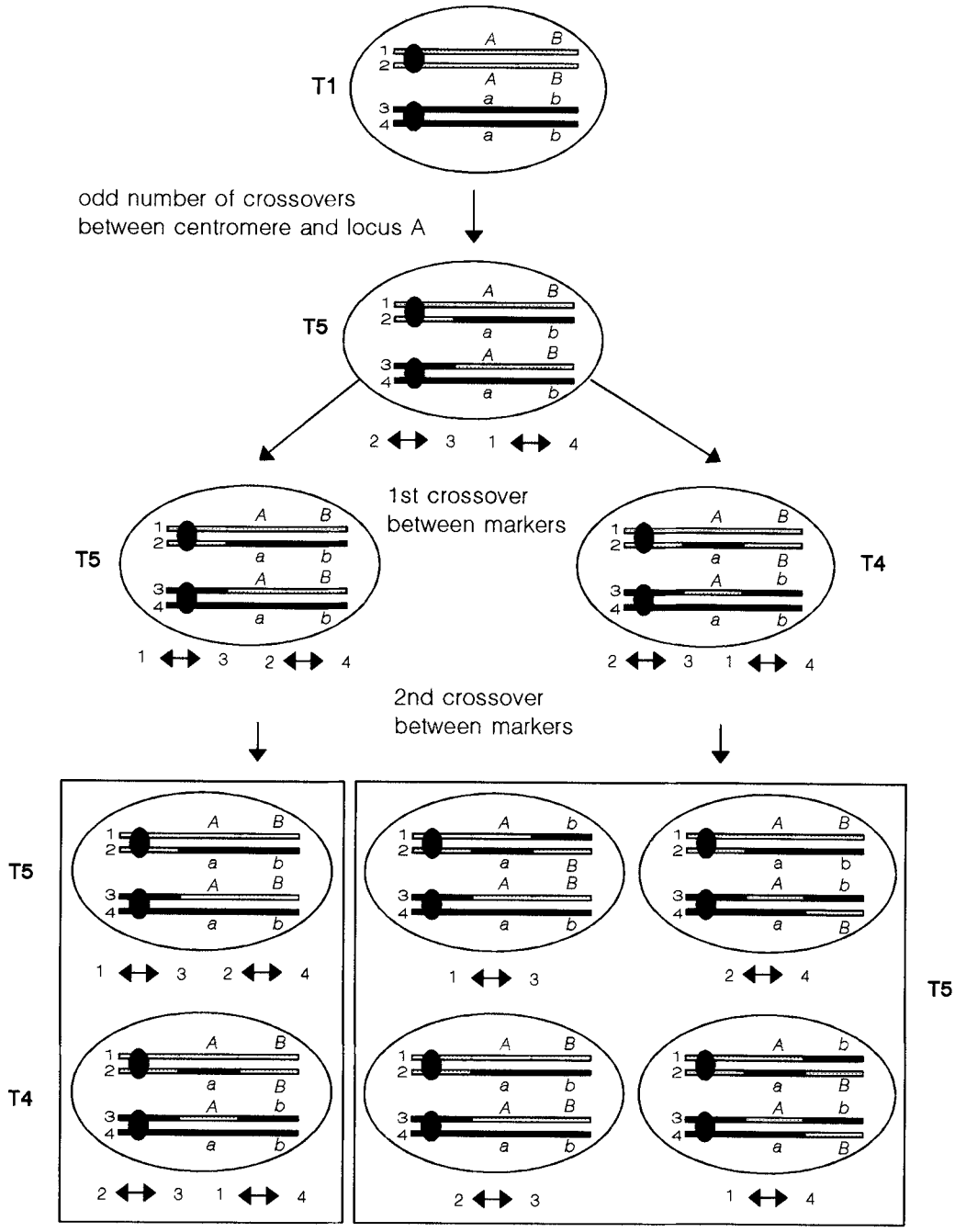

FIGURE 2.—Outcomes of two crossovers between marker loci *A* and *B* when the centromere flanks the two markers, and the MPC (locus *A*) is heterozygous. Each circle indicates a PO undergoing meiosis I. T1, T2 and T3 indicate the genotype class of the PO and the resulting PB1 and SO at the completion of meiosis I. The two-way arrow (↔) indicates crossover between the two numbered strands. After one crossover between loci *A* and *B*, two PB1 or SO genotypes are possible, 1 (*AaBb* or *AaBb*) :1 (*AaBB* or *Aabb*). Two crossovers between loci *A* and *B* for PO genotype *AaBb* or *AaBb* produces the same array of PB1 or SO types as one crossover, whereas two crossovers for PO genotype *AaBB* or *Aabb* produces three possible PB1 or SO genotypes, 1 (*AaBb* or *AaBb*) :2 (*AabB* or *AaBb*) :1 (*AabB* or *AabB*), the latter being an unobservable FSDC between loci *A* and *B*. The frequency of recombinant strands with respect to loci *A* and *B* for genotype *AaBB* or *Aabb* is 0.50, irrespective of whether they resulted from one or two crossovers.

other interval. However, the likelihood analyses correctly indicated that most of the FSDCs occurred in the *o-B* interval. A minor problem with the likelihood method was that zero FSDC events in the interval *o-A* and five FSDC events in *o-B* had slightly higher likelihood than that for the true locations, one in the *o-A* interval and four in *o-B* (Table 1). A solution to this problem would be to examine the estimated recombination frequency. If the recombination frequency is not close to 0, then the (0, 5) locations probably should not be considered as the correct locations. The likelihood method identified the correct order *A-o-B* with a like-

TABLE 1

Estimates of recombination frequency and the log-likelihood (Log(L))

Order	Recombination frequency		Log (L)
A-*o*-B			
0[a], 5[b]	$\theta_{oA} = 0.1208$	$\theta_{oB} = 0.2239$	−114.1770
1, 4	$\theta_{oA} = 0.1259$	$\theta_{oB} = 0.2188$	−114.8179
2, 3	$\theta_{oA} = 0.1310$	$\theta_{oB} = 0.2138$	−115.3900
3, 2	$\theta_{oA} = 0.1361$	$\theta_{oB} = 0.2088$	−115.8941
4, 1	$\theta_{oA} = 0.1412$	$\theta_{oB} = 0.2038$	−116.3309
5, 0	$\theta_{oA} = 0.1463$	$\theta_{oB} = 0.1988$	−116.3309
o-A-B	$\theta_{oA} = 0.1258$	$\theta_{oB} = 0.2635$	−121.5876
o-B-A	$\theta_{oA} = 0.2185$	$\theta_{oB} = 0.2635$	−141.1468

Simulation 1 is represented here. Estimates are for the order A-*o*-B, when $\theta_{oA} = 0.125$, $\theta_{oB} = 0.220$.

[a] Number of observations consistent with four-strand-double crossovers in A-*o* interval.

[b] Number of observations consistent with four-strand-double crossovers in *o*-B interval.

TABLE 2

Estimates of recombination frequency and the log-likelihood (Log(L))

Order	Recombination frequency		Log (L)
*o*A-B	$\theta_{oA} = 0.1258$	$\theta_{AB} = 0.2188$	−108.8185
A-*o*-B			
0[a], 4[b]	$\theta_{oA} = 0.1210$	$\theta_{oB} = 0.2679$	−115.5853
1, 3	$\theta_{oA} = 0.1261$	$\theta_{oB} = 0.2631$	−116.4525
2, 2	$\theta_{oA} = 0.1312$	$\theta_{oB} = 0.2583$	−117.2543
3, 1	$\theta_{oA} = 0.1363$	$\theta_{oB} = 0.2535$	−117.9919
4, 0	$\theta_{oA} = 0.1414$	$\theta_{oB} = 0.2486$	−118.6662
o-B-A	$\theta_{oB} = 0.2898$	$\theta_{AB} = 0.2188$	−136.6746

Simulation 2 is represented here. Estimates are for the order *o*-A-B, when $\theta_{oA} = 0.125$, $\theta_{AB} = 0.220$.

[a] Number of observations consistent with four-strand-double crossovers in A-*o* interval.

[b] Number of observations consistent with four-strand-double crossovers in *o*-B interval.

lihood ratio (most likely/second most likely) of 5,884,370:1. Normally, a likelihood ratio of 100:1 (WEEKS 1991) or 1000:1 (MATISE *et al.* 1994) can be considered as strong evidence for multilocus ordering.

Example 2: The purpose of this example is to show the likelihood analysis for identifying the correct order *o-A-B*. Recombination frequencies are assumed as $\theta_{oA} = 0.125$, and $\theta_{AB} = 0.220$. A total of 200 phase-known genotype determinations on PB1 or SO was assumed. Then, the number of observations for T_i, $i = 1, \ldots, 5$, was generated using (1), and the generated number of PB1 or SO genotypes was rounded to its nearest integer. The numbers of genotypes generated were $90T_1 + 4T_2 + 59T_3 + 9T_4 + 38T_5$. Likelihoods (1) and (32) were applied to estimate recombination frequencies and to calculate the log-likelihood (base 10). For the order *A-o-B*, five likelihood analyses were performed to identify the largest likelihood given the order *A-o-B*. The order *o-A-B* (the correct order) is favored over the next most likely order *A-o-B* (a wrong order) by a likelihood ratio of 5,845,208:1. Therefore, the likelihood method identified the correct order with strong statistical confidence.

From Tables 1 and 2, it is interesting to note that one FSDC corresponded to a change in θ about $^1/_{200} = 0.005$. That is, if the percentage of FSDC is increased or decreased by p, the estimate of recombination frequency is changed by the same amount. This phenomenon also was observed for other simulations. The theoretical explanation is that FSDCs result in 100% recombinant strands (Figure 1).

DISCUSSION

Advantages and limitations: In comparison with linkage analysis using family and sperm typing data, the multilocus gene-centromere method using polar body and oocyte data has two advantages: the method can order markers relative to the centromere and the method provides a new and possibly more powerful approach to construct female linkage maps. Using reproductive techniques, such as superovulation and micromanipulation, large numbers of polar bodies and oocytes could be collected for most mammals, providing sufficient meiosis to construct female maps. The statistical method described in this paper provides a general solution to estimate recombination frequencies for every possible gene-centromere order and to order multiple loci relative to the centromere. Therefore the method presented in this paper is an important step in making genetic analysis of PB1 and SO data generally applicable.

The method presented in this paper has certain limitations: the paper does not consider chiasma and chromatid interference and genotyping efficiencies. Chiasma interference for three loci, or for the centromere and two loci, could be estimated from coefficient of coincidence or interference. However, a large number of observations is needed to have reliable statistical inference about interference (OTT 1991). Chromatid interference can increase or decrease the frequency of crossovers, but it is much more difficult to detect than chiasma interference (MATHER 1938). In general, PCR-based genotyping of gametes is not completely efficient, *i.e.,* the genotyping method may fail to detect one or both alleles of a locus. Consequently, PB1 or SO phenotype may not reflect the true genotype, and this uncertainty may affect the linkage analysis. Maximum likelihood methods have been proposed to account for amplification efficiencies (CUI *et al.* 1989; CUI *et al.* 1992) but the results have not deviated appreciably from those without considering efficiencies because of high genotyping efficiencies (CUI *et al.* 1989, 1992; VAN EIJK *et al.* 1993). The maximum likelihood approach

to account for genotyping efficiency (CUI *et al.* 1989) could be combined with the maximum likelihood analysis presented in this study to account for the problem of PCR efficiency.

Phase determination: Correct determination of linkage phase is required to implement the analysis of PB1 and SO typing data. If linkage phase is unknown, the parental type could be either T_1 (*AABB* or *aabb*) or T_2 (*AAbb* or *aaBB*). If both T_1 and T_2 are observed in PB1 and SO obtained from the same female, then the more frequent type should be considered as T_1 (the parental type), and the less frequent type should be considered as T_2, which would be caused by FSDC, on the grounds that parental type occurs at least as frequently as the recombinant type.

Study of recombination hot spots: When the MPC is closely linked to the centromere, there should be multiple observations that provide direct evidence of two or more crossovers between two marker loci. Therefore, a MPC that is closely linked to the centromere could be used to study crossover behavior (such as a recombination "hot spot") between two marker loci. Indeed, our recent multilocus gene-centromere mapping experiment involving three loci on bovine chromosome *23* supports the presence of such a hot spot where frequent FSDC may be occurring (JARRELL, *et al.* 1995). Recent molecular genetic data demonstrating that specific DNA sequences spontaneously form four-strand complexes (GAILLARD and STRAUSS 1994) provides a possible mechanism by which such crossovers can be initiated. The methods developed herein provide a general solution to the problem of multilocus gene-centromere mapping which involves all theoretical crossover possibilities, including FSDC.

This study was supported in part by grants from the United States Department of Agriculture, grant IS-1939-91R from the Binational Agricultural Research and Development Fund and grant 91-37205-6335 under the National Research Initiative.

LITERATURE CITED

BOEHNKE, M., N. ARNHEIM, H. LI and F. S. COLLINS, 1989 Fine structure mapping of human chromosomes using the polymerase chain reaction on single sperm: experimental design consideration. Am. J. Hum. Genet. **45:** 21–32.

CHAKRAVARTI, A., P. P. MAJUMBER, S. A. SLAUGENHAUPT, R. DEKA, A. C. WARREN *et al.*, 1989 Gene-centromere mapping and the study of nondisjunction in autosomal trisomies and ovarian teratomas, pp. 45–79 in *Molecular and Cytogenetic Studies of Non-disjunction*, edited by T. J. HASSOLD and C. J. EPSTEIN. A. R. Liss Inc., New York.

CHAKRAVARTI, A., and S. A. SLAUGENHAUPT, 1987 Methods for studying recombination on chromosomes that undergo nondisjunction. Genomics **1:** 35–42.

CÔTÉ, G. B., and J. H. EDWARDS, 1975 Centromeric linkage in autosomal trisomies. Ann. Hum. Genet. **39:** 51–59.

CUI, X., J. GERWIN, W. NAVIDI, H. LI, M. KUEHN *et al.*, 1992 Gene-centromere linkage mapping by PCR analysis of individual oocytes. Genomics **13:** 713–717.

CUI, X., H. LI, T. M. GORADIA, K. LANGE, H. H. KAZAZIAN, D. GALAS and N. ARNHEIM, 1989 Single-sperm typing: determination of genetic distance between the $^{G}\gamma$-globin and parathyroid hormone loci by using the polymerase chain reaction and allele-specific oligomers. Proc. Natl. Acad. Sci. USA **86:** 9389–9393.

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

EPPIG, J. T., and E. M. EICHER, 1983 Application of the ovarian teratoma mapping method in the mouse. Genetics **103:** 797–812.

GAILLARD, C., and F. STRAUSS, 1994 Association of poly(CA).poly(TG) DNA fragments into four-stranded complexes bound by HMG1 and 2. Science **264:** 433–436.

GORADIA, T. M., and K. LANGE, 1990 Multilocus ordering strategies based on sperm typing. Ann. Hum. Genet. **54:** 49–77.

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

JARRELL, V. L., H. A. LEWIN, Y. DA and M. B. WHEELER, 1995 Gene-centromere mapping of bovine *DYA, DRB3* and *PRL* using secondary oocytes and first polar bodies: evidence for four-strand double crossovers between *DYA* and *DRB3*. Genomics. (in press).

LANDER, E. S., and D. BOTSTEIN, 1989 Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics **121:** 185–199.

LEWIN, H. A., M. B. WHEELER and L. B. SHOOK, 1989 UFO-PCR: a strategy for linkage mapping in species with underdeveloped genetic maps, pp. 95–102 in *Mapping the Genomes of Agriculturally Important Animals*, edited by J. E. WOMACK. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

MATHER, K, 1938 Crossing over. Biol. Rev. **13:** 252–292.

MATISE, T. C., M. PERLIN and A. CHAKRAVARTI, 1994 Automated construction of genetic linkage maps using an expert system (MultiMap): a human genome linkage map. Nature Genetics **6:** 384–390.

MORTON, N. E, 1982 *Outline of Genetic Epidemiology,* Karger, New York.

OTT, J., 1991 *Analysis of Human Genetic Linkage.* The Johns Hopkins University Press, Baltimore, MD.

OTT, J., D. LINDER, B. K. MCCAW, E. W. LOVRIEN and F. HECHT, 1976 Estimating distances from the centromere by means of benign ovarian teratomas in man. Ann. Hum. Genet. **40:** 191–196.

ROTHWELL, N. V., 1988 *Understanding Genetics.* 4th Ed. Oxford University Press, New York.

SCHOOK, L. B., H. A. LEWIN and D. G. MCLAREN (Editors), 1991 *Gene-Mapping Techniques and Applications.* Marcel Dekker, Inc., New York.

SNOW, R., 1979 Maximum likelihood estimation of linkage and interference from tetrad data. Genetics **92:** 231–245.

VAN EIJK, M. J. T., I. RUSS and H. A. LEWIN, 1993 Order of bovine DRB3, DYA, and PRL determined by sperm typing. Mamm. Genome **4:** 113–118.

WEEKS, D. E., 1991 Human linkage analysis: strategies for locus ordering, pp. 297–330 in *Advanced Techniques in Chromosome Research,* edited by K. W. ADOLPH. Marcel Dekker, New York.

Communicating editor: T. F. C. MACKAY