Genet. Res., Camb. (2002), **79**, *pp.* 85–96. *With* 4 *figures.* © 2002 Cambridge University Press
DOI: 10.1017/S0016672301005389 Printed in the United Kingdom

Linkage mapping of sex-specific differences

RONGLING WU[1], CHANG-XING MA[1, 2], S. S. WU[1]* AND ZHAO-BANG ZENG[3]
[1] *Department of Statistics, University of Florida, Gainesville, FL 32611 USA*
[2] *Department of Statistics, Nankai University, Tianjin 300071, China*
[3] *Program in Statistical Genetics, Department of Statistics, North Carolina State University, Raleigh, NC 27695, USA*

(*Received 4 April 2001 and in revised form 7 August 2001*)

Summary

Most current linkage analyses assume identical fractions of meiotic recombination between homologous marker loci of the two sexes. This assumption is not realistic, because considerable sex-related differences have been observed in recombination fraction. In this paper, a general EM-based algorithm is presented to estimate sex-specific recombination fractions for a mixed set of molecular markers segregating differently in a full-sib family derived from two heterozygous parents. The asymptotic variances of the estimates of linkage specifically for each of the parents are evaluated using a numerical analysis based on information functions. This approach will have important implications for precise gene mapping based on sex-specific linkage maps.

1. Introduction

Genetic mapping for outcrossing species is primarily based on existing full-sib families derived from heterozygous parents (Grattapaglia & Sederoff, 1994). Because these existing families were originally generated for the purpose of exploiting possible heterosis, most of the parent pairs for crosses are genetically divergent. If the two parents of a full-sib family are selected from different species or different populations, across-parent heterogeneity probably exists in map distance between homologous marker loci. There has been much well-documented evidence for the variation in the rate of meiotic recombination among related species, among families and over environments (Korol *et al*., 1994). Under the action of natural selection, the difference in recombination rate might also happen between the two sexes. Morgan (1912) first described an extreme case in which recombination is absent in male *Drosophila*. In angiosperm plants, the females appear to have a higher genomic map length than the males (e.g. Burt *et al*., 1991; de Vicente & Tanksley, 1991; Graner *et al*., 1991). In gymnosperms, greater meiotic recombination is observed in male than in female gametes (Moran *et al*., 1983; Groover *et al*., 1995). However, the effect of sex on recombination rate is not uniform throughout the genome, being stronger in some intervals, absent in others and, in a few cases, opposite to the general tendency (de Vicente & Tanksley, 1991; Groover *et al*., 1995; Plomion & O'Malley, 1996).

The identification of chromosomal regions heterogeneous between two different parents in a full-sib family owing to either the sex effect or population divergence, or to both, requires appropriate statistical analyses of recombination data. Linkage analysis for a mixed set of markers with a variable number of alleles per locus and different segregation patterns has been carried out by Ritter *et al*. (1990), Arus *et al*. (1994), Ritter and Salamini (1996), Maliepaard *et al*. (1997) and Ridout *et al*. (1998). However, all of these studies are based on the assumption that the recombination rate of each pair of homologous marker loci is identical between the two parents. Generally, in a finite population, a downward bias frequently happens in the set of estimates for recombination rate with a significant test statistic (Maliepaard *et al*., 1997). Thus, when these previous methods are used to perform linkage analysis for two linkage-divergent parents, recombination rates would be underestimated more seriously for the parent with higher recombination rates. Overall, the use of a sex-averaged map would lead to higher type I errors when the two parents are more different in recombination rate. To

* Corresponding author. Tel: +1 352 392 3806. Fax: +1 352 392 8555. e-mail: rwu@stat.ufl.edu

consider the effect of possible differences of recombination rate between two parents, many analyses have been performed separately for the parents (e.g. Grattapaglia & Sederoff, 1994; Maliepaard *et al.*, 1998). Such analyses, however, do not make simultaneous use of segregating markers from both parents, thereby reducing their power to detect a linkage.

In this paper, we develop a general statistical algorithm for estimating sex-specific recombination rates between different types of markers by using an entire marker data set in a full-sib family. Simulation results suggest that simultaneous estimates for recombination rates that are different between the parents can increase the precision and power of linkage estimation.

2. Statistical methods

(i) *Likelihood function*

In a full-sib family derived from two parents of an outcrossing species, the number of different alleles (up to four non-null types *a*, *b*, *c* and *d*, and a null type *o*) can vary across marker loci. Considering all possible combinations of alleles between the two outbred parents, Ritter and Salamini (1996) and Maliepaard *et al.* (1997) specified seven different segregation patterns, each of which provides unique information for linkage analysis, as follows.

A Loci that are heterozygous in both parents and segregate in a 1:1:1:1 ratio, e.g. $ab \times cd$.

B Loci that are heterozygous in both parents and segregate in a 1:2:1 ratio. This is divided into three groups, as follows

- B_1 Three alleles form a non-symmetrical cross type between the two parents, $ab \times ao$.
- B_2 The reciprocal of B_1, $ao \times ab$.
- B_3 Two alleles form a symmetrical type between the two parents, $ab \times ab$.

C Loci that are heterozygous in both parents and segregate in a 3:1 ratio (symmetrical, $ao \times ao$).

D Loci that are in the test-cross configuration (Grattapaglia & Sederoff, 1994) between the parents and segregate in a 1:1 ratio. This is divided into two groups, as follows

- D_1 Heterozygous in one parent and homozygous in the other, e.g. $ab \times aa$.
- D_2 The reciprocal of D_1, $aa \times ab$.

We denote one of the parents P and the other parent Q, and the two chromosomes in each parent are P_1, P_2, Q_1 and Q_2, respectively. For a particular marker, regardless of its marker type, there are four possible parental chromosome pairings (PCPs) in the progeny: $P_1 \| Q_1$, $P_1 \| Q_2$, $P_2 \| Q_1$ and $P_2 \| Q_2$, where $\|$ stands for two homologous chromosomes at the left and right. Depending on the segregating pattern of a marker, these four PCPs will produce different phenotypes observed as bands on a gel.

Consider two adjacent markers in a linkage group of m markers with a known order. We denote the recombination fraction between the two markers as r_P in parent P and r_Q in parent Q. The estimates of the recombination fractions rely upon the linkage phase of the parents at the two markers; that is, the relative assignment of the two alleles from each marker in the two single chromosomes of each parent. Obviously, for each pair of markers, there are four possible combinations of linkage phases between the two

k	$\mathbf{H}^k$	$\mathbf{D}_P^k$	$\mathbf{D}_Q^k$	
1	$\begin{pmatrix} (1-r_P)(1-r_Q) & (1-r_P)r_Q & r_P(1-r_Q) & r_Pr_Q \\ (1-r_P)r_Q & (1-r_p)(1-r_Q) & r_Pr_Q & r_P(1-r_Q) \\ r_P(1-r_Q) & r_Pr_Q & (1-r_P)(1-r_Q) & (1-r_P)r_Q \\ r_Pr_Q & r_P(1-r_Q) & (1-r_P)r_Q & (1-r_P)(1-r_Q) \end{pmatrix}$	$\begin{pmatrix} 0&0&1&1 \\ 0&0&1&1 \\ 1&1&0&0 \\ 1&1&0&0 \end{pmatrix}$	$\begin{pmatrix} 0&1&0&1 \\ 1&0&1&0 \\ 0&1&0&1 \\ 1&0&1&0 \end{pmatrix}$	(1)
2	$\begin{pmatrix} (1-r_P)r_Q & (1-r_P)(1-r_Q) & r_Pr_Q & r_P(1-r_Q) \\ (1-r_P)(1-r_Q) & (1-r_P)r_Q & r_P(1-r_Q) & r_Pr_Q \\ r_Pr_Q & r_P(1-r_Q) & (1-r_P)r_Q & (1-r_P)(1-r_Q) \\ r_P(1-r_Q) & r_Pr_Q & (1-r_P)(1-r_Q) & (1-r_P)r_Q \end{pmatrix}$	$\begin{pmatrix} 0&0&1&1 \\ 0&0&1&1 \\ 1&1&0&0 \\ 1&1&0&0 \end{pmatrix}$	$\begin{pmatrix} 1&0&1&0 \\ 0&1&0&1 \\ 1&0&1&0 \\ 0&1&0&1 \end{pmatrix}$	(2)
3	$\begin{pmatrix} r_P(1-r_Q) & r_Pr_Q & (1-r_P)(1-r_Q) & (1-r_P)r_Q \\ r_Pr_Q & r_P(1-r_Q) & (1-r_P)r_Q & (1-r_P)(1-r_Q) \\ (1-r_P)(1-r_Q) & (1-r_P)r_Q & r_P(1-r_Q) & r_Pr_Q \\ (1-r_P)r_Q & (1-r_P)(1-r_Q) & r_Pr_Q & r_P(1-r_Q) \end{pmatrix}$	$\begin{pmatrix} 1&1&0&0 \\ 1&1&0&0 \\ 0&0&1&1 \\ 0&0&1&1 \end{pmatrix}$	$\begin{pmatrix} 0&1&0&1 \\ 1&0&1&0 \\ 0&1&0&1 \\ 1&0&1&0 \end{pmatrix}$	(3)
4	$\begin{pmatrix} r_Pr_Q & r_P(1-r_Q) & (1-r_P)r_Q & (1-r_P)(1-r_Q) \\ r_P(1-r_Q) & r_Pr_Q & (1-r_P)(1-r_Q) & (1-r_P)r_Q \\ (1-r_P)r_Q & (1-r_P)(1-r_Q) & r_Pr_Q & r_P(1-r_Q) \\ (1-r_P)(1-r_Q) & (1-r_P)r_Q & r_P(1-r_Q) & r_Pr_Q \end{pmatrix}$	$\begin{pmatrix} 1&1&0&0 \\ 1&1&0&0 \\ 0&0&1&1 \\ 0&0&1&1 \end{pmatrix}$	$\begin{pmatrix} 1&0&1&0 \\ 0&1&0&1 \\ 1&0&1&0 \\ 0&1&0&1 \end{pmatrix}$	(4)

https://doi.org/10.1017/S0016672301005389 Published online by Cambridge University Press

parents for the parent cross $P \times Q$: (1) coupling $\times$ coupling; (2) coupling $\times$ repulsion; (3) repulsion $\times$ coupling; and (4) repulsion $\times$ repulsion. Under each linkage phase combination or assignment, the conditional probability of the progeny PCP genotype of marker 2 given marker 1 can be derived, which represents the transition probability of recombination events between the two markers at the PCP genotype level. The transition probability is represented by a (4×4) matrix $\mathbf{H}$, whose elements are given in Eqns 1–4 for any assignment k ($k = 1 \ldots 4$). The two (4×4) matrices at the right of $\mathbf{H}^k$ are those of the number of recombination events between the two markers for parents P ($\mathbf{D}_P^k$) and Q ($\mathbf{D}_Q^k$), respectively (see equations 1–4).

Because the PCP genotypes and phenotypes might not be correspondent (except for marker type A), depending on the segregation patterns of markers, they should be connected using a $(4 \times p_i)$ incident matrix (see equations 5),

$$\mathbf{I}_{Ap_i} = \begin{cases} \begin{pmatrix} 1&0&0&0\\ 0&1&0&0\\ 0&0&1&0\\ 0&0&0&1 \end{pmatrix} & \text{when the marker type} = \text{A} \\ \begin{pmatrix} 1&0&0\\ 1&0&0\\ 0&1&0\\ 0&0&1 \end{pmatrix} & \text{when the marker type} = \text{B}_1 \\ \begin{pmatrix} 1&0&0\\ 0&1&0\\ 1&0&0\\ 0&0&1 \end{pmatrix} & \text{when the marker type} = \text{B}_2 \\ \begin{pmatrix} 1&0&0\\ 0&1&0\\ 0&1&0\\ 0&0&1 \end{pmatrix} & \text{when the marker type} = \text{B}_3 \\ \begin{pmatrix} 1&0\\ 1&0\\ 1&0\\ 0&1 \end{pmatrix} & \text{when the marker type} = \text{C} \\ \begin{pmatrix} 1&0\\ 1&0\\ 0&1\\ 0&1\\ 1& \end{pmatrix} & \text{when the marker type} = \text{D}_1 \\ \begin{pmatrix} 0&1\\ 0&1\\ 1&0\\ 1&0 \end{pmatrix} & \text{when the marker type} = \text{D}_2 \end{cases} \quad (5)$$

where p_i is the number of PCP phenotypes observable as bands on a gel for marker i, which is 4, 3, 3, 3, 2, 2 or 2 for marker types A, B_1, B_2, B_3, C, D_1 or D_2, respectively. Thus, assuming that the marker data are independent among N offspring, the likelihood of the PCP phenotype ($\mathbf{U}$) for the two markers can be expressed as,

$$l(\mathbf{U}|r_P, r_Q, k) = \prod_{j=1}^{N} W_{jj} \quad (6a)$$

where W_{jj} is the jth diagonal element of symmetric matrix,

$$\mathbf{W} = [\mathbf{U}_{p_1}^{\mathrm{T}}(\mathbf{I}_{Ap_1}^{\mathrm{T}}\mathbf{H}^k\mathbf{I}_{Ap_2})\mathbf{U}_{p_2} \quad (6b)$$

$$= \mathbf{U}_{p_1}^{\mathrm{T}}\mathbf{P}_{p_1p_2}^k\mathbf{U}_{p_2}, \quad (6c)$$

where $\mathbf{U}_{pi}$ is a $(p_i \times N)$ 1/0 matrix indicating the phenotypes of marker i in which 1 denotes the phenotype observed for an individual and 0 denotes the phenotype not observed for the same individual. $\mathbf{P}_{p_1p_2}^k = \mathbf{I}_{Ap_1}^{\mathrm{T}}\mathbf{H}^k\mathbf{I}_{Ap_2}$ is a $(p_1 \times p_2)$ matrix of the transition probability from markers 1 to 2 at the phenotype level. $\mathbf{T}$ denotes the transpose of a vector or matrix.

Considering all markers in the linkage group, the likelihood of Eqns (6a–c) can be calculated via a hidden Markov chain process (Lathrop *et al.*, 1985; Lange, 1997; Jiang & Zeng, 1997; Ling, 2000), using Eqn 7,

$$l(\mathbf{U}|r_{P_{12}}, r_{Q_{12}}, \ldots, r_{P_{(m-1)m}}, r_{Q_{(m-1)m}}, k)$$

$$= \prod_{i=1}^{m-1} \mathrm{W}[\mathbf{U}_{p_1}^{\mathrm{T}}(\mathbf{I}_{Ap_1}^{\mathrm{T}}\mathbf{H}_{i(i+1)}^k\mathbf{I}_{Ap_{i+1}})\mathbf{U}_{p_{i+1}}] \quad (7a)$$

$$= \prod_{i=1}^{m-1} \mathrm{W}[\mathbf{U}_{p_i}^{\mathrm{T}}\mathbf{P}_{p_ip_{i+1}}^k\mathbf{U}_{p_{i+1}}] \quad (7b)$$

where i and $i+1$ are indices of two adjacent markers. For an assignment k of linkage phase between markers i and $i+1$, the recombination fractions of the markers in parents P and Q, can be estimated by maximizing the log-likelihood of Eqn 6 with respect to $r_{P_{i(i+1)}}$ and $r_{Q_{i(i+1)}}$, and solving the likelihood equations.

$$\frac{\partial}{\partial r_{P_{i(i+1)}}} \ln[l(\mathbf{U}|r_{P_{i(i+1)}}, r_{Q_{i(i+1)}}, k)] = 0 \quad (8a)$$

$$\frac{\partial}{\partial r_{Q_{i(i+1)}}} \ln[l(\mathbf{U}|r_{P_{i(i+1)}}, r_{Q_{i(i+1)}}, k)] = 0 \quad (8b)$$

To obtain the MLE for $r_{P_{i(i+1)}}$ and $r_{Q_{i(i+1)}}$, the expectation and maximization (EM) algorithm can be used for parameter estimation (Dempster *et al.*, 1977; Lander & Green, 1987). The general equations formulating the iteration of the $\{t+1\}$th EM step are given as follows.

1. *E step*: Calculate the expected numbers of recombination events between markers i and $i+1$ for offspring j under assignment k,

$$c_{Pl_il_{i+1}j}^{k\{t+1\}} = \frac{\mathbf{u}_{l_ij}^T[\mathbf{I}_{Ap_i}^{\mathrm{T}}\mathbf{D}_{P_{i(i+1)}}^k]o[\mathbf{H}_{i(i+1)}^k\mathbf{I}_{Ap_{i+1}}\mathbf{u}_{l_{i+1}j}]}{\mathbf{u}_{l_ij}^{\mathrm{T}}\mathbf{P}_{p_ip_{i+1}}^{k\{t\}}\mathbf{u}_{l_{i+1}j}} \quad (9a)$$

https://doi.org/10.1017/S0016672301005389 Published online by Cambridge University Press

for parent P and

$$c^{k\{t+1\}}_{Ql_il_{i+1}j} = \frac{\mathbf{u}^{\mathrm{T}}_{l_ij}[\mathbf{I}^{\mathrm{T}}_{Ap_i}\mathbf{D}^{k}_{Q_{i(i+1)}}]o[\mathbf{H}^{k}_{i(i+1)}\mathbf{I}_{Ap_{i+1}}\mathbf{u}_{l_{i+1}j}]}{\mathbf{u}^{\mathrm{T}}_{l_ij}\mathbf{P}^{k\{t\}}_{p_ip_{i+1}}\mathbf{u}_{l_{i+1}j}} \tag{9b}$$

for parent Q, where $\mathbf{u}_{l_ij}$ is the jth column of $\mathbf{U}_{p_i}$ describing the PCP phenotype l_i of individual j at marker i and $\circ$ denotes an elementwise product of two matrices.

2. *M step*: Calculate under assignment k using the equations,

$$r^{k\{t+1\}}_{P_{i(i+1)}} = \frac{1}{2N}\sum_{j=1}^{N}\sum_{l_i}^{p_i}\sum_{l_{i+1}}^{p_{i+1}} c^{k\{t+1\}}_{Pl_il_{i+1}j} \tag{10a}$$

$$r^{k\{t+1\}}_{Q_{i(i+1)}} = \frac{1}{2N}\sum_{j=1}^{N}\sum_{l_i}^{p_i}\sum_{l_{i+1}}^{p_{i+1}} c^{k\{t+1\}}_{Ql_il_{i+1}j} \tag{10b}$$

These iterative procedures are repeated between Eqns 9 and 10 until the values converge to stable values. These stable values represent the maximum likelihood estimates (MLEs) of the recombination fraction between markers i and $i+1$ for assignment k.

(ii) *Characterization of linkage phases*

The statistical analyses described above are performed separately for different linkage phase combinations in parents. Thus, with two markers, we will have four different estimates for a single recombination fraction in each parent, with each estimate corresponding to a particular linkage phase combination or allelic assignment in parents. Here, Bayes' theorem is used to characterize the most likely linkage phase combination in parents (and therefore the best estimate for recombination fraction). When considering recombination fractions between all possible adjacent markers, the most likely parental linkage phase combination over the m markers is characterized by calculating the posterior probability for each of a total 4^{m-1} parental assignment combinations,

$$\begin{aligned} &P(k_1k_2 \ldots k_{m-1}|\mathbf{U}) \\ &= \frac{P(\mathbf{U}|k_1k_2 \ldots k_{m-1})P(k_1k_2 \ldots k_{m-1})}{\sum^4_{k_1}\sum^4_{k_1}\ldots\sum^4_{k_{m-1}} P(\mathbf{U}|k_1k_2 \ldots k_{m-1})P(k_1k_2 \ldots k_{m-1})} \\ &= \frac{P(\mathbf{U}|k_1k_2 \ldots k_{m-1})}{\sum^4_{k_1}\sum^4_{k_2}\ldots\sum^4_{k_{m-1}} P(\mathbf{U}|k_1k_2 \ldots k_{m-1})} \end{aligned} \tag{11}$$

where: k_1 to k_{m-1} are the k_1th assignment of the alleles between markers 1 and 2 to the k_{m-1}th assignment of the alleles between markers $m-1$ and m; $P(k_1k_2 \ldots k_{m-1})$ is the prior probability of a particular assignment combination and assumed to be uniform; and $P(\mathbf{U}|k_1k_2 \ldots k_{m-1})$ is the likelihood of the data $\mathbf{U}$ given the assignment combinations for the m markers, which can be calculated from Eqn 7. The largest posterior probability corresponds to the most likely assignment combination among the m markers.

(iii) *Hypothesis tests*

Given any two adjacent markers i and $i+1$, their linkage under an optimal linkage phase combination for the two parents can be tested by formulating the hypotheses H_0 (free recombination; $r_{P_{i(i+1)}} = r_{Q_{i(i+1)}} = 0{\cdot}5$) and H_1 (linkage; at least one of them is not equal to 0·5). The significance of linkage is tested based on the likelihood ratio test statistic,

$$LR_{1i(i+1)} = -2\ln\left[\frac{l(\mathbf{U}|r_{P_{i(i+1)}} = r_{Q_{i(i+1)}} = 0.5, k)}{l(\mathbf{U}|\hat{r}_{P_{i(i+1)}}, \hat{r}_{Q_{i(i+1)}}, k)}\right], \tag{12}$$

where the numerator and denominator are the likelihood of the data for assignment k of markers i and $i+1$ under absence of linkage and under linkage, respectively. The value of $LR_{1i(i+1)}$ approximately follows a χ-square distribution with two degrees of freedom under the null hypothesis. The sex-specific difference of the recombination fraction under an optimal linkage phase configuration can also be tested by formulating $H_0{:}r_{P_{i(i+1)}} = r_{Q_{i(i+1)}}$ and $H_1{:}r_{P_{i(i+1)}} \neq r_{Q_{i(i+1)}}$. The hypotheses are tested by calculating a similar test statistic,

$$LR_{2i(i+1)} = -2\ln\left[\frac{l(\mathbf{U}|\tilde{r}_{P_{i(i+1)}} = \tilde{r}_{Q_{i(i+1)}}, k)}{l(\mathbf{U}|\hat{r}_{P_{i(i+1)}}, \hat{r}_{Q_{i(i+1)}}, k)}\right], \tag{13}$$

where $\sim$ stands for the MLE of recombination fraction under the null hypothesis. The value of $LR_{2i(i+1)}$ is distributed as a χ-square variable with one degree of freedom. Several other statistics can also be used to test linkage, such as Mather's linkage test χ^2_L (Mather, 1951) or the contingency test for independence (Garcia-Dorado & Gallego, 1992).

3. Precision analysis

The precision of the MLE can be assessed by its sampling variance, which is approximately equal to the inverse of Fisher's information (i.e. the expectation of minus the second derivative of the log-likelihood function). Under the assumption of identical recombination fraction between two parents, the information functions were compared between different marker combinations (Mather, 1951; Ritter *et al.*, 1990; Ritter & Salamini, 1996; Maliepaard *et al.*, 1997). In this study, we use the information functions to examine the effect of sex-specific difference on the precision of the estimate of recombination fraction.

https://doi.org/10.1017/S0016672301005389 Published online by Cambridge University Press

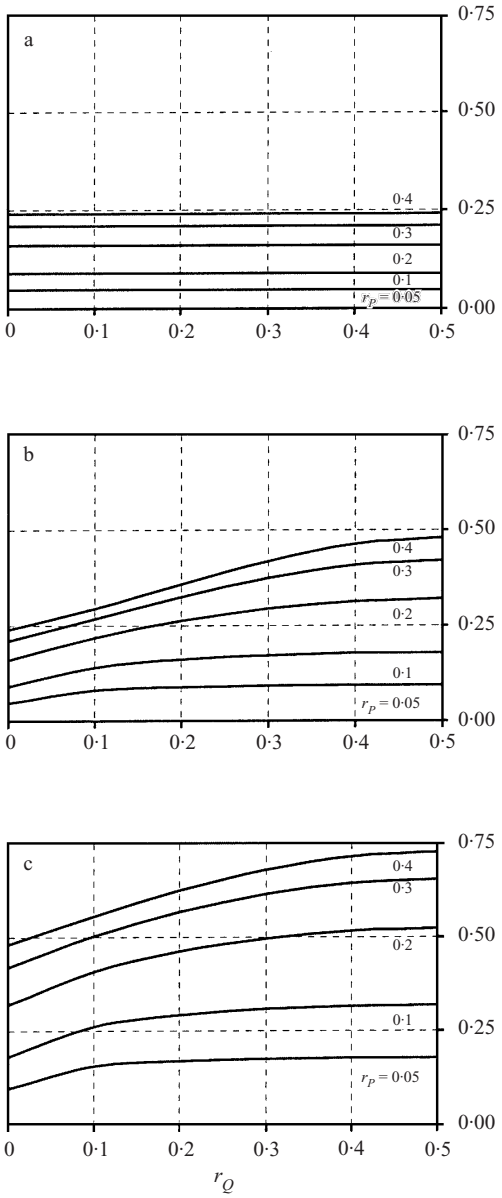

Fig. 1. Asymptotic sampling variance of the estimated recombination fraction between a pair of markers at least one of which is from type A (1:1:1:1), as affected by the degree of linkage in alternative parents. (*a*) A–A pair, (*b*) A–B_3 pair and (*c*) A–C pair.

The information function measuring the quality of the MLE $r_{P_{i(i+1)}}$ is given by

$$I_{r_{P_{i(i+1)}}} = -E\left(\frac{\partial^2}{\partial r^2_{P_{i(i+1)}}} ln[l(\mathbf{U}|r_{P_{i(i+1)}}, r_{Q_{i(i+1)}}, k)]\right)$$

$$= N\sum_{l_i=1}^{p_i}\sum_{l_{i+1}=1}^{p_{i+1}}\frac{1}{P^k_{l_i l_{i+1}}}\left(\frac{\partial P^k_{l_i l_{i+1}}}{\partial r_{P_{i(i+1)}}}\right)^2. \quad (14)$$

A similar equation can also be written for the MLE of the recombination fraction for parent Q.

We derive the information functions of the estimated recombination fractions for all possible combinations of different types of markers when the recombination fractions are considered to be different between two parents (Table 1). The amount of information about the recombination fraction is lower when a sex-specific difference is assumed than when no sex-specific difference is assumed, with the rate of reduction depending on marker combination. For example, the information for two 1:1:1:1-segregating markers is reduced by a half if the recombination fraction is considered to be different between the two parents, whereas such a reduction can be less than a half or even less than a third for two markers, one from type A and the other from type B_1 or B_2. The information about the recombination fraction is unchanged for any two paired markers one of which from type D, because one parent homozygous at marker type D does not provide information for segregation analysis.

The information functions of recombination fraction might have different forms for the same marker pair when allelic configurations are not symmetrical between the two parents (Table 1). For example, one null allele in one parent for marker B_1 or B_2 make its resulting allelic configurations asymmetrical, leading to different forms of information function between the two parents when marker B_1 or B_2 is paired with any other marker (except for the case in which these two markers are paired with each other). Although a type C dominant marker is symmetrical, an asymmetrical allelic configuration will occur when two dominant alleles are in a coupling phase in one parent and in a repulsion phase in the other, and therefore have different forms of the information function for the two parents. The forms of the information function can be different when different allelic configurations are assumed for marker pairs B_1–B_2 and C or C and C (Table 1).

Because two recombination fractions are estimated in this study, the information function for each cannot be used precisely to assess the sampling errors of the estimates when the two estimates are correlated. Such a correlation does not occur between marker pairs of type A, B_1 and B_2, and cannot be estimated between marker pairs with one marker from type D. The correlation between two estimates of sex-specific recombination fractions should be considered to estimate the asymptotic sampling variances of the estimates. We compare the sampling variances of the estimates as a function of recombination fraction by pairing the most informative marker A, the moderately informative marker B and the least informative marker C. As expected, the asymptotic sampling variance of the estimated recombination fraction between two type A markers is the lowest of all possible marker pairs (Figs 1–4). Also, the sampling variance of the estimated recombination fraction between two type A markers from one parent is only affected by the degree of linkage in this parent and is unaffected by the

https://doi.org/10.1017/S0016672301005389 Published online by Cambridge University Press

Table 1. *The formulae for calculating the information of the MLE of recombination fraction for different combinations of markers when the recombination fraction is different between two parents P & Q*

[a]Locus		There is a sex-specific difference		There is no sex-specific difference
1	2	Information function	Co-information function	Information function
A	A	$\frac{1}{r_P(1-r_P)}$	0	$\frac{2}{r(1-r)}$
A	D_1/D_2	$I_{r_h}=\frac{1}{r_h(1-r_h)}$	–	$\frac{1}{r(1-r)}$
$B_1(B_2)$	$D_1(D_2)$	$I_{r_h}=\frac{1}{r_h(1-r_h)}$	–	$\frac{1}{r(1-r)}$
$D_1(D_2)$	$D_1(D_2)$	$I_{r_h}=\frac{1}{r_h(1-r_h)}$	–	$\frac{1}{r(1-r)}$
A	B_1/B_2	$\begin{cases} I_{r_n}=\frac{1}{r_n(1-r_n)} \\ I_{r_o}=\frac{1}{2r_o(1-r_o)} \end{cases}$	0	$\frac{3-4r+4r^2}{2r(1-r)}$
$B_1(B_2)$	$B_1(B_2)$	$\begin{cases} I_{r_n}=\frac{1}{r_n(1-r_n)} \\ I_{r_o}=\frac{1-r_n}{2r_o(1-r_o)} \end{cases}$	0	$\frac{3-r}{2r(1-r)}$
B_1	B_2	$\frac{1}{2r_P(1-r_P)}$	0	$\frac{1}{r(1-r)}$
$B_1(B_2)$	$D_2(D_1)$	$\frac{1}{2r_h(1-r_h)}$	–	$\frac{1}{2r(1-r)}$
B_3	D_1/D_2	$\frac{1}{2r_h(1-r_h)}$	–	$\frac{1}{2r(1-r)}$
C	D_1/D_2	$\frac{1+4r_h-4r_h^2}{r_h(1-r_h)(1+r_h)(2-r_h)}$	–	$\frac{1+4r-4r^2}{r(1-r)(1+r)(2-r)}$
A	B_3	$\frac{1}{2}\left[\frac{1}{r_P(1-r_P)}+\frac{(1-2r_Q)^2}{\rho(1-\rho)}\right]$	$\frac{(1-2r_p)(1-2r_Q)}{2\rho(1-\rho)}$	$\frac{2(1-3r+3r^2)}{r(1-r)(1-2r+2r^2)}$

A	C	$\frac{1}{4}\left[\frac{1}{r_P(1-r_P)}+\frac{(2-r_Q)r_Q^2}{(1-r_Pr_Q)(1-r_Q+r_Pr_Q)}+\frac{(1+r_Q)(1-r_Q)^2}{v(1-r_P+r_Pr_Q)}\right]$	$\frac{1}{4}\left[\frac{r_Pr_Q}{1-r_Pr_Q}+\frac{(1-r_P)(1-r_Q)}{r_P+r_Q-r_Pr_Q}-\frac{r_P(1-r_Q)}{1-r_P+r_Pr_Q}-\frac{(1-r_P)r_Q}{1-r_P+r_Pr_Q}\right]$	$\frac{1+4r-3r^2-2r^3+r^4}{r(1-r^2)(2-r)}+\frac{(1-2r)^2}{1-r+r^2}$
B_1/B_2	B_3	$I_{r_n}=\frac{1}{4}\left[\frac{2}{r_n(1-r_n)}+\frac{(1-2r_o)^2}{(1-\rho)\rho}\right]$ $I_{r_o}=\frac{1}{4}\left[\frac{2}{r_o(1-r)_o}+\frac{(1-2r_n)^2}{(1-\rho)\rho}\right]$	$\frac{(1-2r_P)(1-2r_Q)}{4\rho(1-\rho)}$	$\frac{1+8r-8r^2}{4r(1-r)}+\frac{(1-2r)^2(1-3r+3r^2)}{r(1-r)(1-2r+2r^2)}$
B_1/B_2	C	$I^C_{r_n}=\frac{1}{4}\left[\frac{2-r_n^2}{r_n(1-r_n)(2-r_n)}+\frac{r_o^2}{1-r_o+r_nr_o}+\frac{(1-r_o)^2}{r_n+r_o-r_nr_o}\right]$ $I^C_{r_o}=\frac{1}{4}\left[\frac{1-r_n}{r_o(1-r_o)}+\frac{(1+r_n)(1-r_n)^2}{(1-r_o+r_nr_o)(r_n+r_o-r_nr_o)}\right]$	$\frac{(1-r_n)(1-r_n-r_o+r_nr_o-r_o^2)}{4(1-r_o+r_nr_o)(r_n+r_o-r_nr_o)}$	$\frac{1}{4}\left[\frac{4(1+6r-5r^2+2r^3)}{r(2-r)(2-r^2)}+\frac{(1-2r)^2(2-2r+r^2)}{(1-r)(1-r+r^2)}\right]$
		$I^R_{r_n}=\frac{1}{4}\left[\frac{1+2r_n-r_n^2}{r_n(1-r_n)(1+r_n)}+\frac{r_o^2}{1-r_nr_o}+\frac{(1-r_o)^2}{1-r_n+r_nr_o}\right]$ $I^R_{r_o}=\frac{1}{4}\left[\frac{r_n}{r_o(1-r_o)}+\frac{(2-r_n)r_n^2}{(1-r_nr_o)(1-r_n+r_nr_o)}\right]$	$-\frac{r_n(1-2r_o)}{4(1-r_nr_o)(1-r_n+r_nr_o)}$	$\frac{1}{4}\left[\frac{6}{(1+r)(1-r)}+\frac{(1-2)^2}{r(1-r)(1-r+r^2)}\right]$
B_3	B_3	$\frac{1}{2}\left[\frac{\rho}{r_P(1-r_P)}+\frac{(1-2r_Q)^2(2-\rho)}{\rho(1-\rho)}\right]$	$\frac{1}{2}\left[2+\frac{(1-2r_P)(1-2r_Q)(2-\rho)}{\rho(1-\rho)}\right]$	$\frac{2(1-3r+3r^2)}{r(1-r)(1-2r+2r^2)}$
B_3	C	$\frac{1}{4}\left[\frac{r_Q}{r_P(1-r_Pr_Q)}+\frac{2(1-2r_Q)^2}{\rho(2-p)}+\frac{(1+r_Pr_Q)(1-r_Q)^2}{v(1-p)}\right]$	$\frac{1}{4}\left[\frac{1}{1-r_Pr_Q}+\frac{2(1-2r_P)(1-2r_Q)}{\rho(2-\rho)}+\frac{(1+r_Pr_Q)(1-r_P)(1-r_Q)}{v(1-p)}\right]$	$\frac{1}{2}\left[\frac{2(1+2r-2r^2)}{r(2-r)(1-r^2)}+\frac{(1-2r)^2}{r(1-r)(1-r+r)^2}\right]$
C	C	$I^{CC}_{rp}=\frac{1}{4}\left[\frac{1-r_Q}{1-r_P}+\frac{(6-v)(1-r_Q)^2}{v(3-v)}\right]$	$\frac{1}{4}\left[\frac{(1-r_P)(1-r_Q)(6-5v)}{v(3-v)}+\frac{(1-r_P)(1-r_Q)}{1-v}\right]$	$\frac{2(2-4r+2r^2)}{r(2-r)(3-2r+r^2)}$
		$I^{CR}_{r_C}=\frac{1}{4}\left[\frac{1-r_R}{1-r_C}+\frac{(5+r_R-r_Cr_R)r_R^2}{(1-r_R+r_Cr_R)(2+r_R-r_Cr_R)}\right]$ $I^{CR}_{r_R}=\frac{1}{4}\left[\frac{1-r_C}{1-r_R}+\frac{(5+r_R-r_Pr_Q)(1-r_C)^2}{(1-r_R+r_Cr_R)(2+r_R-r_Cr_R)}\right]$	$\frac{1}{4}\left[1-\frac{(1-r_C)r_R(5+r_C-r_Cr_R)}{(2+r_R-r_Cr_R)(1-r_R+r_Cr_R)}\right]$	$1+\frac{5(1-2r)^2}{4(1-2r+2r^2)(2+r-r^2)}$
		$I^{RR}_{r_P}=\frac{r_Q(1+2r_Pr_Q)}{2r_P(1-r_Pr_Q)(2+r_Pr_Q)}\Big]$	$\frac{1+2r_Pr_Q}{2(1-r_Pr_Q)(2+r_Pr_Q)}\Big]$	$\frac{2(1+2r^2)}{(2+r^2)(1-r^2)}$

[a]Different marker combinations. The marker types for marker 1 given in parentheses are only compared with those for marker 2 in parentheses.
[b]r_n denotes the recombination fraction for the parent that contains no null allele at marker B_1/B_2; r_o is associated with the parent containing a null allele at the same marker; r_n is associated with the parent that is heterozygous at a D_1/D_2 marker. *C* or *R* mens a coupling or repulsion phase, respectively, for two non-null alleles, each from a marker in a parent. When there is the same form of the information function for the two parents, only one formula for the information is given. $\rho = r_P+r_Q-2r_Pr_Q$, $v = r_P+r_Q-r_Pr_Q$.

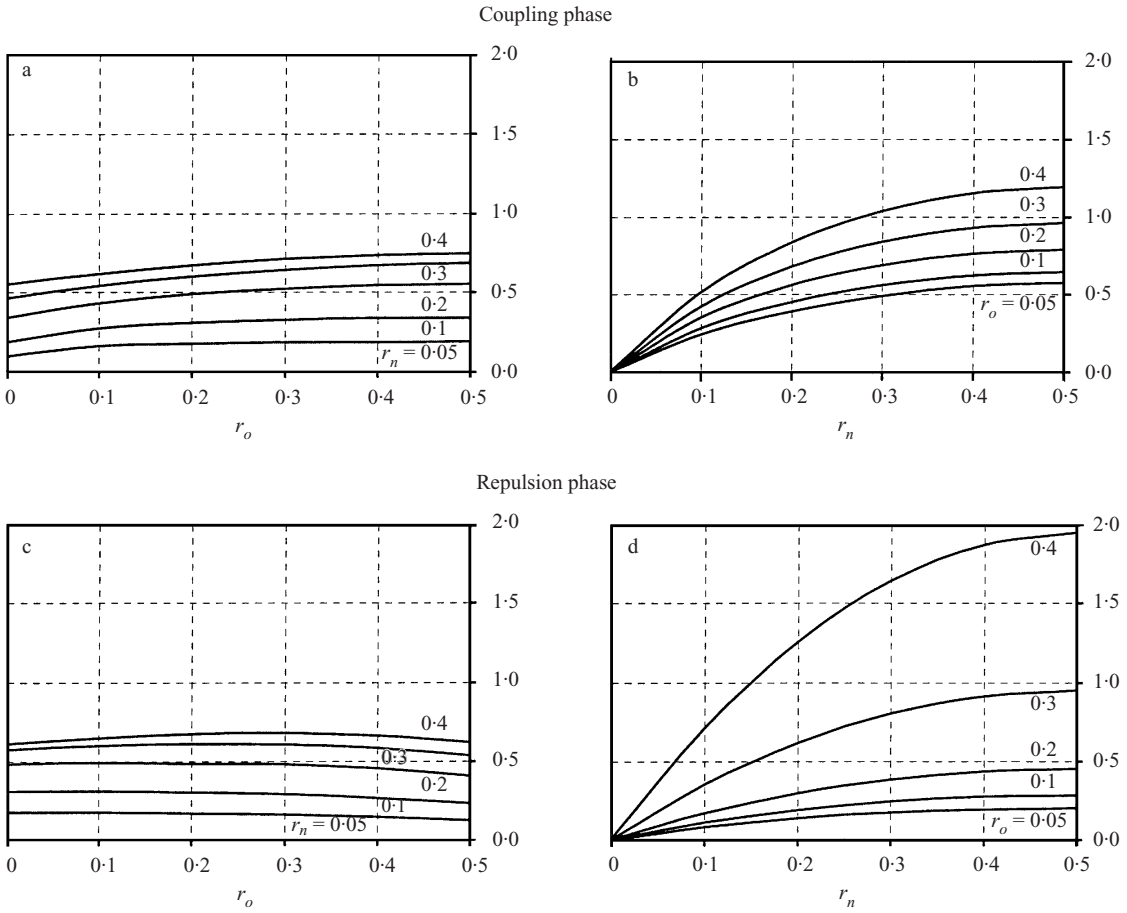

Fig. 2. Asymptotic sampling variance of the estimated recombination fraction between two marker, one from type B_1 (1:2:1) and the other from type C (3:1), as affected by the degree of linkage in alternative parents. The two dominant alleles, one from B_1 and the other from C, might be in a coupling phase (*a*, *b*) or in a repulsion phase in a parent (*c*, *d*). (*a*, *c*) The sampling variance of the recombination fraction for the parent including the non-null allele at B_1. (*b*, *d*) The sampling variance of the recombination fraction for the parent including the null allele at B_1.

degree of linkage in the other parent (Fig. 1*a*). When a type A marker is paired with a type B_3 (Fig. 1*b*) or C (Fig. 1*c*) marker, the sampling variance of the estimated recombination fraction increases significantly. In the A–B_3 or A–C pair, the sampling variance of the estimate is affected by the degree of linkage in the parent under consideration and also by the degree of linkage in the other parent. The sensitivity of the sampling variance of the estimate to the degree of linkage in the alternative parent is stronger when two markers are loosely rather than tightly linked in the parent under consideration.

The influence of sex-specific differences on the estimate of recombination fraction is much more complex when two paired markers are each from a partially informative marker type. For example, the sampling variance of the estimated recombination fraction between type B_1 or B_2 and type C markers is affected by the degree of linkage in both parents and the linkage phase of the two markers in the parents (Fig. 2). The estimate of the recombination fraction of markers B_1 and C has a smaller sampling variance for the parent that includes the non-null allele at B_1 (Fig. 2*a*, *c*) than for the parent that includes the null allele at this marker (Fig. 2*b*, *d*). This contrast is much sharper when the two dominant alleles, one from B_1 and the other from C, are in a repulsion phase (Fig. 2*a*, *b*) rather than a coupling phase in a parent (Fig. 2*c*, *d*). Also, the sampling variance of the estimate of the recombination fraction in the parent containing the non-null allele at B_1 is not much affected by the degree of linkage in the alternative parent containing the null allele at this marker (Fig. 2*a*, *c*), whereas the sampling variance of the recombination fraction in the parent containing the null allele is highly sensitive to the degree of linkage in the parent containing the non-null allele (Fig. 2*b*, *d*).

The sampling variance of the estimate of recombination fraction between type B_3 and C markers is more sensitive to the difference of linkage between the two parents than to the degree of linkage itself (Fig. 3). The sampling variance of the estimate increases with the reduced difference of linkage between the two parents and tends towards infinity when the difference approaches zero. Also, if these two types of markers have a loose linkage, the between-parent difference of

https://doi.org/10.1017/S0016672301005389 Published online by Cambridge University Press

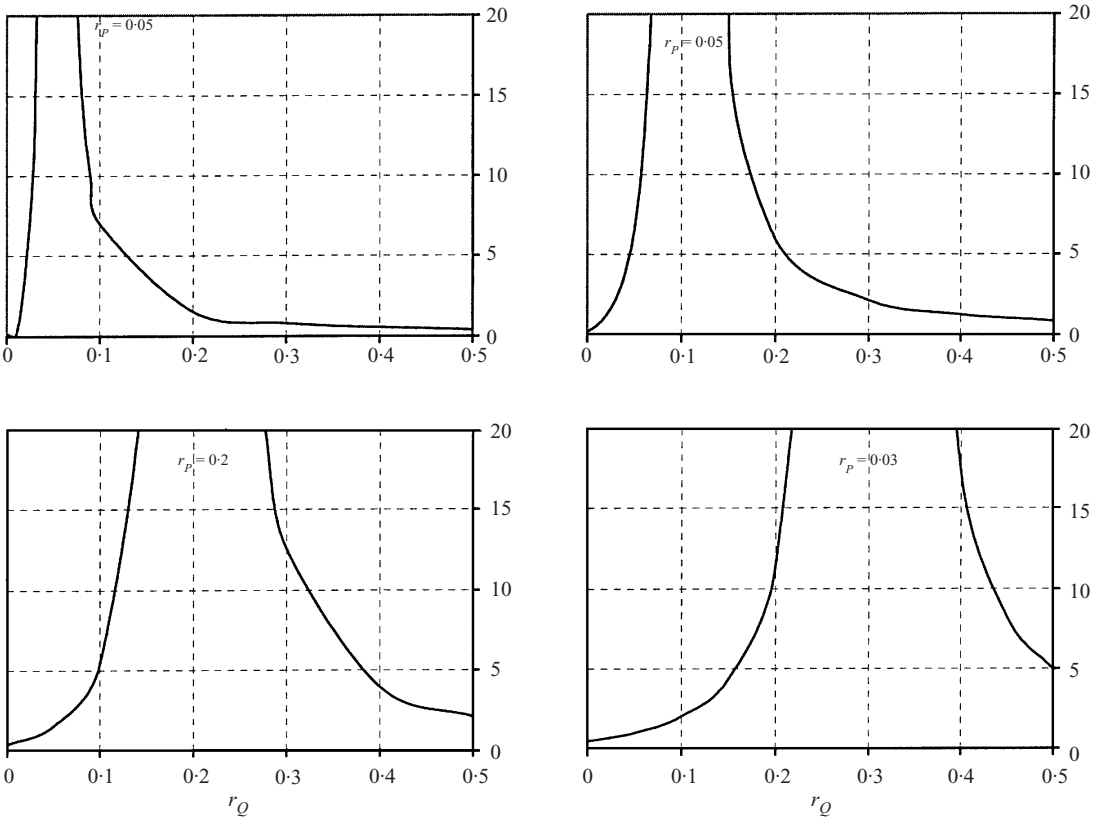

Fig. 3. Asymptotic sampling variance of the estimated recombination fraction between two markers, one from type B_3 (1:2:1) and the other from type C (3:1), as affected by the degree of linkage in alternative parents.

linkage must be very large to generate a reasonably small sampling variance. Given these properties, parent-specific linkage analysis is not recommended for type B_3 and C markers unless there is strong evidence for a difference between their linkage in the two parents.

When two type C markers are paired, the two parents have three different allelic configuration combinations for two dominant alleles: (1) coupling phase × coupling phase; (2) coupling phase × repulsion phase; and (3) repulsion phase × repulsion phase. The sampling variance of the estimate of recombination fraction cannot be evaluated for the third combination because the information matrix is singular. The amount and change pattern of the sampling variance of the estimate are different in the first two combinations (Fig. 4). When two dominant alleles from a marker are in a coupling phase in both parents, the sampling variance of the estimate is larger than when they are in a coupling phase in one parent (Fig. 4*a*) but in a repulsion phase in the other (Fig. 4*b*). If two dominant alleles are in a coupling phase, the sampling variance is only dependent on the degree of linkage in the alternative parent, being unaffected by the degree of linkage in the parent under consideration. When two dominant alleles are in a coupling phase in one parent but in a repulsion phase in the other, the estimate of recombination fraction has different sampling variances for the two parents. The estimate from the parent with two dominant alleles in a coupling phase has smaller sampling variances than the estimate from the parent with two dominant alleles in a repulsion phase. Generally, small sampling variances of the estimate for the mixed linkage phase between the two parents mean that two dominant markers can provide a precise estimate of recombination fraction when the sex-specific difference of linkage is assumed. This is in sharp contrast to the situation in which two dominant markers are not effective for the estimate of recombination fraction when the same linkage is assumed between the two parents (Maliepaard *et al*., 1997).

4. Discussion

Linkage analysis for molecular markers is more challenging for a full-sib family derived from outbred parents than for a family of two inbred parental lines, for two reasons. Firstly, there are uncertainties in an outbred family about the number of alleles per marker locus and the linkage phase between different markers. Secondly, the two parents used to make a full-sib family are usually genetically diverged because the original purpose of the mating is to exploit possible heterosis. Thus, the differences in map distances of markers likely occur between the two parents, as

https://doi.org/10.1017/S0016672301005389 Published online by Cambridge University Press

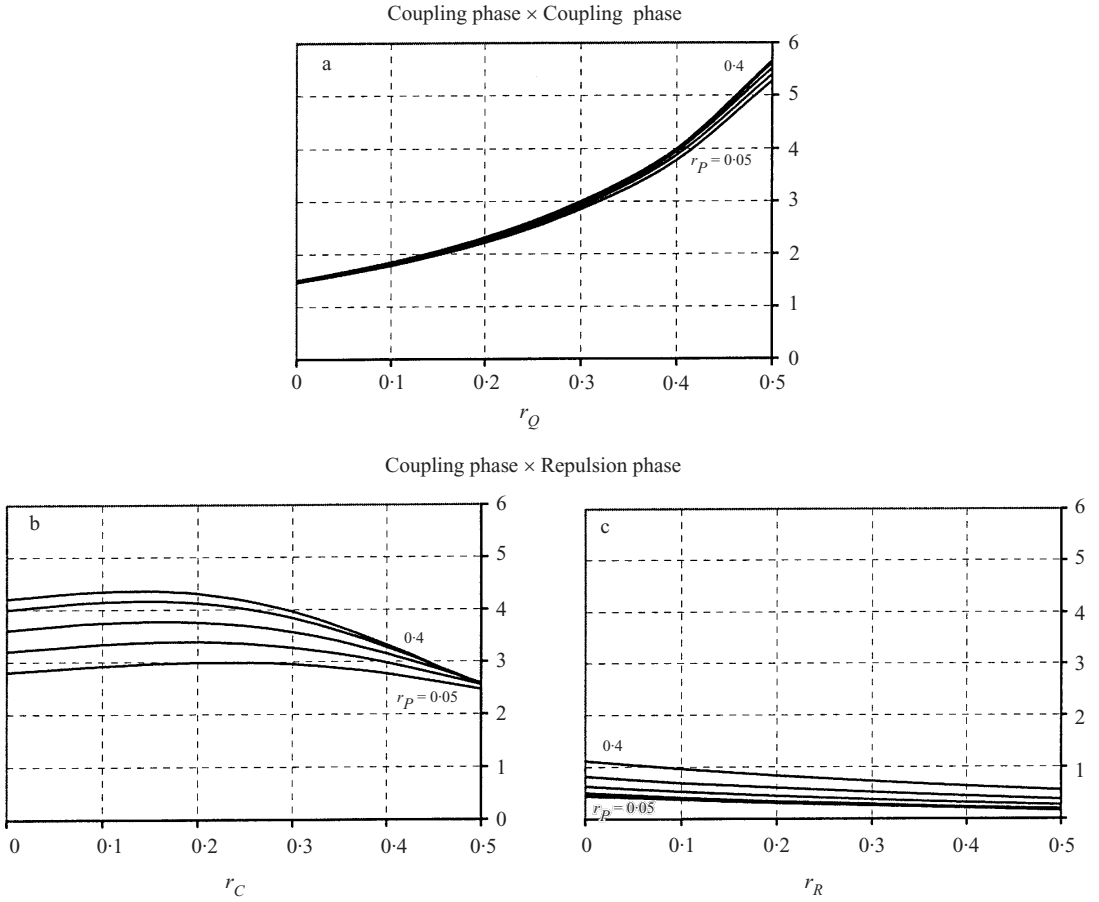

Fig. 4. Asymptotic sampling variance of the estimated recombination fraction between two type C (3:1) markers, as affected by the degree of linkage in alternative parents.

suggested by a number of molecular studies in a variety of organisms (see Introduction). Different map distances in homologous pairs of chromosomes can be caused by differences in DNA sequence, DNA content and chromosomal rearrangements. The linkage analysis strategies in the current literature have not considered sex-specific differences in recombination fraction within a framework of multilocus linkage phase inferences, and so cannot accurately manipulate the increasingly accumulative data of molecular markers from various laboratories.

In this paper, we propose a general EM algorithm for estimating recombination fractions specifically for different parents for a mixed set of molecular markers segregating as 1:1:1:1, 1:2:1, 3:1 or 1:1 in a full-sib family. It is likely that these types of markers occur simultaneously in an outbred population and can be characterized using a variety of marker systems, such as microsatellite, RAPD and AFLP. Compared with previous linkage analyses by Ritter *et al.* (1990), Arus *et al.* (1994), Ritter and Salamini (1996), Maliepaard *et al.* (1997) and Ridout *et al.* (1998), our analysis has three major advantages for analysing linkage between different marker types. First, a general procedure has been implemented for linkage analysis of various types of markers based on multilocus linkage phase inferences. The previous analyses presented all possible pairs of different types of markers and gave individual formulae for estimating recombination fraction between a pair of markers by assuming one of the four possible combinations of linkage phases in the two parents. Such a strategy is not very efficient for computer implementation. Second, our analysis considers sex-specific differences in recombination fraction, which can eliminate the bias in estimating linkage generated by traditional methods that assume an identical rate of recombination between two parents. Finally, the previous analyses cannot clearly specify linkage phases between those markers of symmetrical cross types between two parents (i.e. types B_3 or C). Assume that parents P and Q have a symmetrical cross type at two markers, one from type B_3 and the other from type C (*aabo* × *aabo*). For this cross type, linkage phase assignments

$$\frac{a}{a}\left\|\frac{b}{o}\times\frac{a}{o}\right\|\frac{b}{a} \quad \text{and} \quad \frac{a}{o}\left\|\frac{b}{a}\times\frac{a}{a}\right\|\frac{b}{o} \tag{15}$$

cannot be distinguished when the recombination fraction of the two markers is postulated to be identical between the two parents (Maliepaard *et al.*,

https://doi.org/10.1017/S0016672301005389 Published online by Cambridge University Press

1997). However, this problem can be well solved using our analysis by taking sex-specific difference into account.

Our linkage analysis method also has the strength of simultaneously estimating linkage and parental linkage phase over all markers from a chromosome. The characterization of parental linkage phase configuration is based on the calculation of posterior probabilities conditional on a given marker data set. The highest posterior probability corresponds to the most likely linkage phase configuration. But this strategy will be computationally burdened when many markers are considered at the same time. For example, when there are ten 1:1:1:1-segregating markers, the number of linkage phase configurations will be $4^9 = 262{,}144$. A feasible approach to manipulating many markers includes two steps: (1) characterizing the linkage phase configuration over a small number of selected markers (five or so) that are more informative; and (2) using these phase-determined markers as anchors to characterize the linkage phases of the rest of the markers. This approach can also be used to determine the best order of markers on the same chromosome. On the framework map of informative markers that can be effectively ordered with existing software (Lander *et al*., 1987), one can use a stepwise analysis procedure to put those less informative markers at right positions.

In this study, the precision of the estimate of sex-specific recombination fractions is assessed using a numerical analysis of the asymptotic sampling variance of the estimate based on the information function. The precision analysis under the assumption of sex-specific differences in linkage is performed for several representative marker pairs and compared with that under the assumption of identical linkage between two parents. The sampling variance of the estimate is expected to increase when the sex-specific difference is characterized, although the rate of increase depends on marker pair types and allelic configurations. The sex-specific linkage analysis of marker pairs B_3 and C might generate huge sampling errors if the difference in linkage between two parents is small. However, for marker pairs C and C, sex-specific linkage analysis has high precision, especially when two dominant alleles are in a coupling phase in one parent but a repulsion phase in the other. In practice, the simultaneous use of our sex-specific linkage analysis and traditional linkage analysis is recommended. Sex-specific linkage analysis is expected to be reasonably precise when an actual difference occurs in linkage between two parents.

Most current mapping studies have used sex-averaged maps in the search for genes. Given large differences in the level of meiotic recombination between the female and male parents in both animals and plants, such an analytical strategy is not effective and probably provides misleading information for gene mapping. Now, a more powerful linkage analysis for simultaneously estimating recombination fractions specifically for each of the two parents for different marker types is available, permitting the precise mapping of genes of various interests in an outbred full-sib family based on sex-specific linkage maps.

We thank two anonymous reviewers for their useful comments on this manuscript. This work is partly supported by grant GM 45344 from the National Institutes of Health and grant #9801300 from USDA Plant Genome Program To Z.-B.Z.

References

Arus, P., Olarte, C., Romero, M. & Vargas, F. (1994). Linkage analysis of 10 isozyme genes in F_1 segregating almond progenies. *Journal of the American Society for Horticultural Science* **119**, 339–344.

Burt, A., Bell, G. & Harvey, P. H. (1991). Sex differences in recombination. *Journal of Evolutionary Biology* **4**, 259–277.

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood from incomplete data via EM algorithm. *Journal of the Royal Statistical Society*, *Series B* **39**, 1–38.

de Vicente, M. C. & Tankslay, S. D. (1991). Genome-wide reduction in recombination of backcross progeny derived from male versus female gametes in an interspecific cross of tomato. *Theoretical and Applied Genetics* **83**, 173–178.

Garcia-Dorado, A. & Gallego, A. (1992). On the use of the classical tests for detecting linkage. *Journal of Heredity* **83**, 143–146.

Graner, A., Jahoor, A., Schondelmaier, J., Siedler, H., Pillen, K., Fischbeck, G., Wenzel, G. & Herrmann, R. G. (1991). Construction of an RFLP map of barley. *Theoretical and Applied Genetics* **83**, 250–256.

Grattapaglia, D. & Sederoff, R. (1994). Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. *Genetics* **137**, 1121–1137.

Groover, A. T., Williams, C. G., Devey, M. E., Lee, L. M. & Neale, D. B. (1995). Sex-related differences in meiotic recombination frequency in *Pinus taeda*. *Journal of Heredity* **86**, 157–158.

Jiang, C. J. & Zeng, Z.-B. (1997). Mapping quantitative trait loci with dominant and missing markers in various crosses from two inbred lines. *Genetica* **101**, 47–58.

Korol, A. B., Preygel, I. A. & Preygel, S. I. (1994). *Recombination Variability and Evolution: Algorithms of Estimation and Population-Genetic Models*. London: Chapman & Hall.

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

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

Lange, K. (1997). *Mathematical and Statistical Methods in Genetic Analysis*. New York: Springer.

Lathrop, G. M., Lalouel, J. M., Julier, C. & Ott, J. (1985). Multilocus linkage analysis in humans: detection of linkage and estimation of recombination. *American Journal of Human Genetics* **37**, 482–498.

Ling, S. (2000). *Constructing Genetic Maps for Outbred Experimental Crosses*. Ph.D. Dissertation. University of California. Berkeley, CA, USA.

Maliepaard, C., Jansen, J. & van Ooijen, J. W. (1997). Linkage analysis in a full-sib family of an outbreeding plant species: overview & consequences for applications. *Genetical Research* **70**, 237–250.

Maliepaard, C., Alston, F. H., van Arkel, G., Brown, L. M., Chevreau, E., Dunemann, F., Evans, K. M., Gardiner, S., Guilford, P., van Heusden, A. W., Janse, J., Laurens, F., Lynn, J. R., Manganaris, A. G., den Nijs, A. P. M., Periam, N., Rikkerink, E., Roche, P, Ryder, C., Sansavini, S., Schmidt, H., Tartarini, S., Verhaegh, J. J., Vrielink-van Ginkel, M. & King, G. J. (1998). Aligning male and female linkage maps of apple (*Malus pumila* Mill.) using multi-allelic markers. *Theoretical and Applied Genetics* **97**, 60–73.

Mather, K. (1951). *The Measurement of Linkage in Heredity*. London: Methuen.

Moran, G. F., Bell, J. C. & Hilliker, A. J. (1983). Greater meiotic recombination in male vs. female gametes in *Pinus radiata*. *Journal of Heredity* **74**, 62.

Morgan, T. H. (1912). Complete linkage in the second chromosome of the male of *Drosophila*. *Science* **36**, 719–720.

Plomion, C. & O'Malley, D. M. (1996). Recombination rate differences for pollen parents and seed parents in *Pinus pinaster*. *Heredity* **77**, 341–350.

Ridout, M. S., Tong, S., Vowden, C. J. & Tobutt, K. R. (1998). Three-point linkage analysis in crosses of allogamous plant species. *Genetical Research* **72**, 111–121.

Ritter, E., Gebhardt, C. & Salamini, F. (1990). Estimation of recombination frequencies and construction of RFLP linkage maps in plants from crosses between heterozygous parents. *Genetics* **125**, 645–654.

Ritter, E. & Salamini, F. (1996). The calculation of recombination frequencies in crosses of allogamous plant species with applications to linkage mapping. *Genetical Research* **67**, 55–65.