Theor Appl Genet (1995) 90:73–80

M. Lorieux · B. Goffinet · X. Perrier
D. González de León · C. Lanaud

Maximum-likelihood models for mapping genetic markers showing segregation distortion. 1. Backcross populations

Received: 11 August 1993 / Accepted: 2 April 1994

Abstract A maximum-likelihood approach is used in order to estimate recombination fractions between markers showing segregation distortion in backcross populations. It is assumed that the distortions are induced by viability differences between gametes or zygotes due to one or more selected genes. We show that Bailey's (1949) estimate stays consistent and efficient under more general assumptions than those defined by its author. This estimate should therefore be used instead of the classical maximum-likelihood estimate. The question of detection of linkage is also discussed. We show that the order of markers on linkage groups may be affected by segregation distortion.

Key words Genetic mapping · Segregation distortion · Maximum-likelihood · Linkage · Molecular markers

Introduction

Segregation distortion is a problem often encountered in mapping studies (Wendel and Parks 1984; Torres et al. 1985; Lyttle 1991; Schön et al. 1991; Zivy et al. 1992). It has been shown that the analysis of linkage may be influenced by deviations of single-locus segregation ratios from expected frequencies, and several authors have discussed methods to test for linkage or to estimate recombination frequencies between genes showing segregation distortion (Bailey 1949; Garcia-Dorado and Gallego 1992).

The most common disturbances to the expected frequencies of the phenotypic classes (1:1:1:1 for a backcross) are caused by: (1) linkage between the two loci, and (2) upsets in the formation or function of gametes or zygotes, leading to differential viability. A third source of disturbance is constituted by failures of manifestation that lead to assigning a proportion of individuals to phenotypic classes inappropriate to their genotypes; this effect cannot be met with molecular markers, apart from mistyping errors. Linkage can be distinguished from differential viability because it upsets the joint distribution without affecting the single-marker ratios.

The subject of this paper is to extend some of the pre-cited methods which were developed for genes to molecular markers, since the markers are not directly affected by viability effects, but show significant deviations in their segregation ratios, due to their linkage to genes affected by differential viability.

For backcrosses, Bailey's (1949) method can be used for dominant or codominant markers to estimate recombination fractions, since in both cases the four classes can be distinguished. We show here that this method often leads to a consistent and efficient estimate of the recombination frequency between two markers, even when these markers are not located on the genes that cause segregation distortion. An estimate is consistent, or asymptotically unbiased, if it converges to the "true" value of the parameter as the population size increases. It is efficient if no other estimate has a smaller variance. The bias of the classical estimate is derived in several cases of selection, in order to compare it with the standard error of Bailey's estimate.

Communicated by G. Wenzel

M. Lorieux[1] (✉) · C. Lanaud
CIRAD-BIOTROP, B.P. 5035, 34032 Montpellier Cedex 1, France

B. Goffinet
Station de Biométrie et d'Intelligence Artificielle, INRA, B.P. 27, 31326 Castanet-Tolosan Cedex, France

X. Perrier
CIRAD-FLHOR, B.P. 5035, 34032 Montpellier Cedex 1, France

D. González de León
CIMMYT, Lisboa 27, Colonia Juárez, Apdo. Postal 6-641, 06600 México, D.F., México

Present address:
[1]LRGAPT-ORSTOM, BP5045, 34032 Montpellier Cedex 1, France

Two-point models

Estimation of linkage

It is easy to show that when an allelic form at only one locus affects viability, then the best estimate of r is simply the classical estimate, which is the ratio of the number of recombinant individuals over the total number. In this case of selection, the proportionality between the expected frequencies of the parental and recombinant classes remains the same. Now, consider a coupling mating of the type $AB/ab \times ab/ab$, involving two markers A and B, exactly located on two genes, both affected by independent selections. This mating gives rise to four kinds of offspring, which are phenotypically: AB, ab (parentals) and Ab, aB (recombinants). Then, consider that the viability of A phenotypes relative to a is u, and that the viability of B phenotypes relative to b is v, with $0 < u < +\infty$, and $0 < v < +\infty$, using Bailey's (1949) notations. The case $u = v = 1$ is that of no selection, i.e., Mendelian segregation. The recombination fraction is r, and the population size is n. The observed and expected frequencies of phenotypic classes are given in Table 1. Note that gametic and zygotic selection leads to the same result in backcrosses, since the four phenotypic classes correspond to the four gametic types produced by the heterozygous parent.

When u and v are different from one, then the log-likelihood is, omitting an irrelevant constant,

$$L = (a+d)\log(1-r) + (b+c)\log r + (a+b)\log u + (a+c)\log v - n\log[(uv+1)(1-r) + (u+v)r]. \quad (1)$$

L is maximized when r, u and v are replaced by their maximum-likelihood estimates (MLEs), obtained by partially deriving L:

$$\begin{cases} \dfrac{\partial L}{\partial r} = \dfrac{a+d}{r-1} + \dfrac{b+c}{r} - n\dfrac{u+v-uv-1}{D} = 0 \\ \dfrac{\partial L}{\partial u} = \dfrac{a+b}{u} - n\dfrac{v-rv+r}{D} = 0 \\ \dfrac{\partial L}{\partial v} = \dfrac{a+c}{v} - n\dfrac{u-ru+r}{D} = 0. \end{cases} \quad (2)$$

Table 1 Expected and observed frequencies for a backcross in coupling, involving two genes, A and B, selected with intensities u and v. For matings in repulsion, r is replaced by $1-r$.

Phenotypes	AB	Ab	aB	ab
Expected	$n\frac{uv(1-r)}{D}$	$n\frac{ur}{D}$	$n\frac{vr}{D}$	$n\frac{(1-r)}{D}$
Observed frequencies	a	b	c	d

$D = (uv+1)(1-r) + (u+v)r$

The resolution of (2) gives the efficient estimate (Bailey 1949)

$$\hat{r}_B = \frac{\sqrt{bc}}{\sqrt{ad} + \sqrt{bc}} \quad (3a)$$

and

$$\hat{u} = \sqrt{\frac{ab}{cd}} \quad \hat{v} = \sqrt{\frac{ac}{bd}}. \quad (3b)$$

In practical situations, (3a) will be undefined if one of the observed frequencies is zero. Nevertheless, this problem is circumvented by solving (2) iteratively, e.g., by the Newton-Raphson's algorithm (see Edwards 1972).

Replacing the observations by their expectations, the asymptotic variance of the estimate (3a) can be expressed in a slightly different form than that given in Bailey (1949):

$$V_{\hat{r}_B} = r(1-r)[(uv+1)(1-r) + (u+v)r] [(uv+1)r + (u+v)(1-r)]/4nuv.$$

Fig. 1a shows the value of the asymptotic standard error $s_{\hat{r}_B} = \sqrt{V_{\hat{r}_B}}$ of $\hat{r}_B$, a function of r, u and v, for a backcross of 100 individuals (if the population size is n, $s_{\hat{r}_B}$ is obtained by multiplying the values of Fig. 1a by $\sqrt{100/n}$). For u and/or $v = 1$, $s_{\hat{r}_B}$ is equal to the standard error of the classical estimate $s_{\hat{r}_B} = \sqrt{r(1-r)/n}$. It appears that $s_{\hat{r}_B}$ is considerably increased for strong values of selection. To appreciate the pertinence of using Bailey's estimate instead of the classical estimate, we can compare $s_{\hat{r}_B}$ to the asymptotic bias, $B_{\hat{r}}$, of the classical estimate. This bias is obtained by replacing b and c by their expectations in the estimate, and then subtracting the "true" value of r, giving

$$B_{\hat{r}} = \frac{(u+v)r}{D} - r. \quad (4)$$

Fig. 1b shows the value of $B_{\hat{r}}$, as a function of r, u and v. The comparison of Fig. 1a and b clearly indicates that $s_{\hat{r}_B}$ is in all cases lower than the absolute value of $B_{\hat{r}}$, indicating that it is always advantageous to use Bailey's estimate for this model of selection.

After studying the estimation of linkage with segregation distortion, we now discuss the detection of linkage, which is directly related to its estimation.

Detection of linkage

Several methods exist to detect the existence of linkage between two loci. The χ^2 and the LOD score tests are very frequently used, and are asymptotically equivalent. In order to study the effects of segregation distortion on

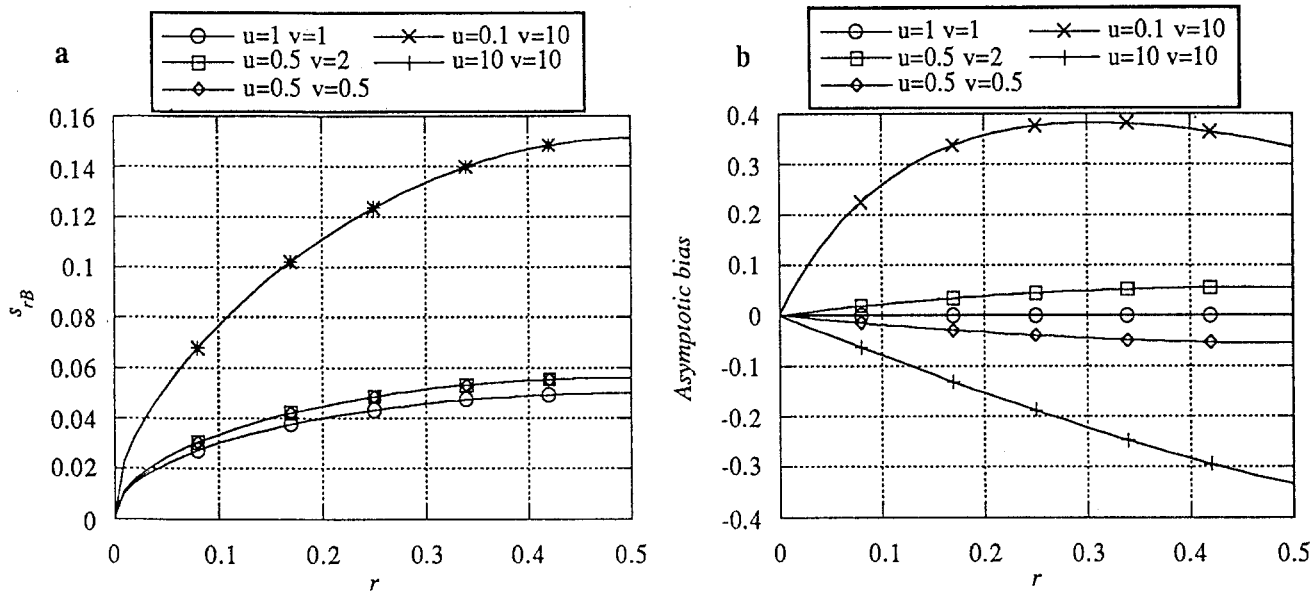


Fig. 1a,b Estimation of the recombination fraction, r, between two markers selected with intensities u and v for a backcross. **a** Asymptotic standard error of Bailey's estimate, against r, for a population size of 100 individuals. **b** Asymptotic bias of the classical estimate against r

the accuracy of the detection of linkage, one can derive the ELOD (for expected LOD score), i.e., the weighted average of the LOD scores calculated for each possible outcome in a given population. For a backcross, the LOD score evaluated at $\hat{r}$, the MLE of r, is equal to

$$Z_{\max}=\begin{cases} n[\log(2)+\hat{r}\log(\hat{r})+(1-\hat{r})\log(1-\hat{r})] & \text{if } \hat{r}>0 \\ n\log(2) & \text{if } \hat{r}=0 \end{cases}$$

and the ELOD is equal to

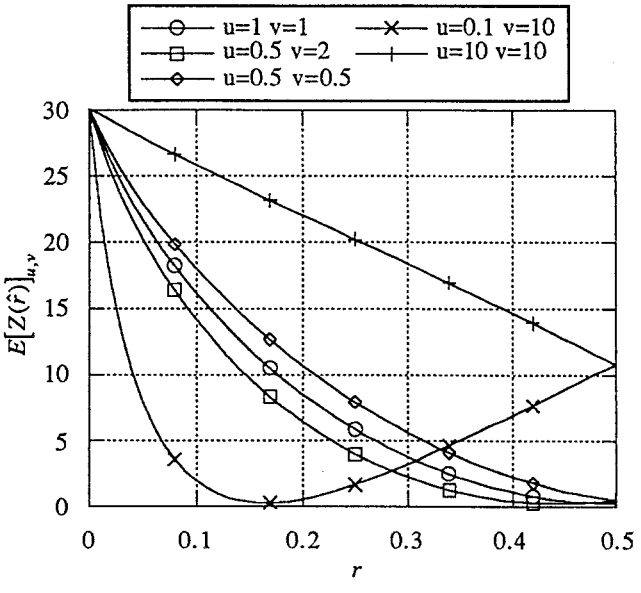
$$E[Z(\hat{r})]=\sum_{k=0}^{n} P(k;r)\times Z_k(\hat{r}) \qquad (5)$$

where k denotes the number of recombinant individuals in the population of size n, $\hat{r}$ is the estimated value of r, the $Z_k(\hat{r})$ are the LOD scores for each outcome k and a given value of $\hat{r}$, and the $P(k;r)$ are the weights, i.e., the probabilities of the outcomes k for a given value of r (Ott 1985). For a backcross, the $P(k;r)$ are given by the binomial probabilities $P(k;r)=C_n^k r^k(1-r)^{n-k}$. When differential viability occurs on the two markers, the expected frequency of recombinant individuals is $p=(u+v)r/[(uv+1)(1-r)+(u+v)r]$. Thus, the ELOD can be expressed as

$$E[Z(\hat{r})]_{u,v}=\sum_{k=0}^{n} C_n^k p^k(1-p)^{n-k}\times Z_k(\hat{r}). \qquad (6)$$

Figure 2 shows the values of $E[Z(\hat{r})]_{u,v}$, as a function of r, u and v, for a population size of 100 individuals. We see from this figure that, for the tested value of $r=0.5$, the LOD score ignoring selection will be overestimated when both markers are under selection. This means that segregation distortion may generate false-positive linkages, leading to the aggregation of two or several linkage groups. On the other hand, for $r\neq 0.5$, false-negative linkages may appear, leading to the division of a linkage group into several groups. Thus, the determination of linkage groups may be biased by segregation distortion. This effect could seriously reduce the utility of the map, for example for detecting QTLs (quantitative trait loci). However, the LOD score will be well estimated when only one marker is under selection (data not shown).

Fig. 2 Expected LOD score as a function of the recombination fraction, r between two markers selected with intensities u and v (backcross population)



To circumvent the problem of bias in the detection of linkage, it is possible to derive an unbiased LOD score test, which takes into account the selection parameters

$$Z(\hat{r}_B, \hat{u}, \hat{v}) = (a+d)\log(1-\hat{r}_B) + (b+c)\log(\hat{r}_B) - n\log\left[\frac{(1-\hat{r}_B)(\hat{u}\hat{v}+1)+\hat{r}_B(\hat{u}+\hat{v})}{\hat{u}+\hat{v}+\hat{u}\hat{v}+1}\right] \quad (7)$$

where $\hat{r}_B$, $\hat{u}$ and $\hat{v}$ are obtained by solving (2) or by using (3a) and (3b). Under the null hypothesis (i.e., $r = 0.5$), this test is asymptotically equivalent to the well-known χ^2 suggested by Mather (1957) to test the independence of two segregations, conditional on the marginal frequencies (one degree of freedom). A simple formula for this test is

$$\chi_1^2 = \frac{n(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)}. \quad (8)$$

Note that the deviation at locus B may be simply due to its linkage to A. To test the significance of the departure of v from unity, Bailey suggests to use the following χ^2 test

$$\chi_1^2 = \frac{n(ac-bd)^2}{(a+b)(a+d)(b+c)(c+d)}. \quad (9)$$

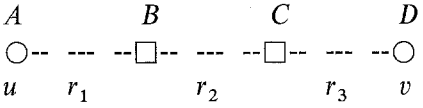
This χ^2 is with one degree of freedom since, of the three available, two were used in estimating u and r.

We now discuss the more general case of the analysis of linkage between two markers linked to two selected genes but not exactly located on them.

Four-point models

Estimation of linkage

Consider two molecular markers, B and C, both showing a significant deviation in their segregation ratios, due to their linkage to two genes, A and D, both affected by a gametic or a zygotic selection u and v, and flanking the interval defined by B and C. Let the recombination fractions between A and B, B and C, and C and D be r_1, r_2 and r_3, respectively. We can represent this situation by the following model



(model 1)

where □ denotes markers, and ○ denotes the genes under selection. Suppose that only the segregations of the two markers are observable. Then, the expected and observed frequencies of the four classes are those of Table 2. Five parameters (r_1, r_2, r_3, u and v) have to be estimated, but only three degrees of freedom are available. We are interested only in estimating r_2, so we can carry out the following convenient transformations

$$r_2 = r$$

$$\alpha = r_3 - vr_3 - 1$$

$$\beta = ur_1 - r_1 + 1.$$

The expected frequencies of Table 2 become

$$f(BC) = n(r-1)\frac{\alpha\beta}{D}\left(1+\frac{v+1}{\alpha}\right)\left(1-\frac{u+1}{\beta}\right)$$

$$f(Bc) = nr\frac{\alpha\beta}{D}\left(1-\frac{u+1}{\beta}\right)$$

$$f(bC) = nr\frac{\alpha\beta}{D}\left(1-\frac{v+1}{\alpha}\right)$$

$$f(bc) = n(r-1)\frac{\alpha\beta}{D}.$$

Let us put $\gamma = [1+(v+1)/\alpha]$ and $\delta = [1-(u+1)/\beta]$. As the sum of the expected frequencies is equal to n, i.e.,

Table 2 Expected and observed frequencies for a backcross in coupling, involving two markers, B and C, flanked by two genes, A and D, selected with intensities u and v

Phenotypes at the four loci	Recombination between B and C	Expected frequencies	Observed frequencies
ABCD, aBCd, ABCd, aBCD	No	$n\frac{(r_2-1)(v+r_3-vr_3)(ur_1-r_1-u)}{D^a}$	e
ABcD, aBcd, ABcd, aBcD	Yes	$n\frac{r_2(r_3-vr_3-1)(ur_1-r_1-u)}{D}$	f
AbCD, abCd, AbCd, abCD	Yes	$n\frac{r_2(v+r_3-vr_3)(ur_1-r_1+1)}{D}$	g
AbcD, abcd, Abcd, abcD	No	$n\frac{(r_2-1)(r_3-vr_3-1)(ur_1-r_1+1)}{D}$	h

[a] $D = (uv-u-v+1)(r_1r_3 - r_{AD}) + uv + 1$

$\alpha\beta[(r-1)(\gamma\delta+1)+r(\gamma+\delta)]/D=n$, we can write

$$f(BC)=n\frac{(r-1)\gamma\delta}{(r-1)(\gamma\delta+1)+r(\gamma+\delta)}$$

$$f(Bc)=n\frac{r\delta}{(r-1)(\gamma\delta+1)+r(\gamma+\delta)}$$

$$f(bC)=n\frac{r\gamma}{(r-1)(\gamma\delta+1)+r(\gamma+\delta)}$$

$$f(bc)=n\frac{(r-1)}{(r-1)(\gamma\delta+1)+r(\gamma+\delta)}.$$

Then, the number of parameters to be estimated is equal to the number of degrees of freedom. Under this condition, Bailey (1951) showed that, subject to the condition of solubility, the maximum-likelihood estimates of the parameters can be obtained by setting the observations equal to the expectations

$$\begin{cases} n\dfrac{(r-1)\gamma\delta}{(r-1)(\gamma\delta+1)+r(\gamma+\delta)}=e \\ n\dfrac{r\delta}{(r-1)(\gamma\delta+1)+r(\gamma+\delta)}=f \\ n\dfrac{r\gamma}{(r-1)(\gamma\delta+1)+r(\gamma+\delta)}=g \\ n\dfrac{(r-1)}{(r-1)(\gamma\delta+1)+r(\gamma+\delta)}=h. \end{cases}$$

Solving this system for r leads to the maximum-likelihood estimate

$$\hat{r}=\frac{\sqrt{fg}}{\sqrt{eh}+\sqrt{fg}} \qquad (10)$$

which is identical to (3a), and has the same variance since γ and δ are the estimates of u and v.

However, it can be shown that, if a selected gene is located between the markers, then Bailey's estimate is biased. This situation is summarized by the model

$$\begin{array}{ccccccc} A & & & B & & C & \\ \square & -- & --- & --\bigcirc-- & --- & -- & \square \\ & & r_1 & u & r_2 & & \end{array} \qquad \text{(model 2)}$$

where $\square$ denotes markers, and $\bigcirc$ denotes the gene under selection. When only the segregations of the two markers are observable, the expected and observed frequencies of the four classes are those of Table 3. The asymptotic bias of Bailey's estimate in this situation is given by

$$B_{\hat{r}_B}=\frac{\sqrt{bc}}{\sqrt{bc}+\sqrt{ad}}-r$$

Table 3 Expected and observed frequencies for a backcross in coupling, involving two markers, A and C, flanking a gene, B, selected with intensity u

Phenotypes	Expected frequencies	Observed frequencies
ABC, AbC	$n\dfrac{(1-r_1)(1-r_2)u+r_1r_2}{u+1}$	a
ABc, Abc	$n\dfrac{(1-r_1)r_2u+r_1(1-r_2)}{u+1}$	b
aBC, abC	$n\dfrac{r_1(1-r_2)u+(1-r_1)r_2}{u+1}$	c
aBc, abc	$n\dfrac{r_1r_2u+(1-r_1)(1-r_2)}{u+1}$	d

where a, b, c and d are replaced by their expectations (Table 3). Figure 3 shows the values of $B_{\hat{r}_B}$ for the model where the selected gene, B, is located exactly midway to the two markers, A and B, assuming no interference. Note that the classical estimate is consistent in this situation, since

$$\frac{b+c}{n}=r_1+r_2-2r_1r_2$$

where b and c are replaced by their expectations. If there is no interference, this quantity is equal to the recombination fraction between A and C. In practice, the failure of Bailey's estimate in this model should not seriously affect linkage analysis: consider a linkage group with m markers, defining $m-1$ intervals, and l selected linked genes, with no more than one gene per interval. Suppose that the order of markers is known. Then, if we use Bailey's estimate, $m-l-1$ recombination fractions will be well estimated, and l will be biased. If the classical estimate is used, any recombination fraction of an interval located between the genes will be biased. If we assume that the number of selected genes per linkage group is relatively small, Bailey's estimate will on average, provide, better estimates of the distances than the classical estimate. We therefore suggest the use of Bailey's estimate when several markers on a linkage group show segregation distortion.

Detection of linkage

In the previous section, we have shown that for a majority of situations, represented by model 1, Bailey's estimate is consistent and efficient. Since this estimate is based on the expected frequencies which take into account the selection parameters u and v, it is directly related to the LOD score $Z(\hat{r},\hat{u},\hat{v})$ (equation 7), or to the χ^2 test of independence defined by Mather (equation 8). Consequently, these two linkage tests can still be used under model 1.

For model 2, it can be shown that both tests are unbiased under the null hypothesis of independence (i.e., $r = 0.5$). This result is not surprising because Bailey's estimate is not biased for $r = 0.5$ (Fig. 3). Therefore, either one or the other test can be used under very general conditions to detect linkage.

Ordering markers

The accuracy of any maximum-likelihood method of ordering loci (see, for example, Lander and Green 1987; Lathrop and Lalouel 1988) is directly related to the quality of the estimation of the recombination frequencies. Since the algorithms usually used to estimate these frequencies do not take into account the possibility of segregation distortion, it is necessary to study if such disturbances can affect the determination of the order of the markers of a linkage group.

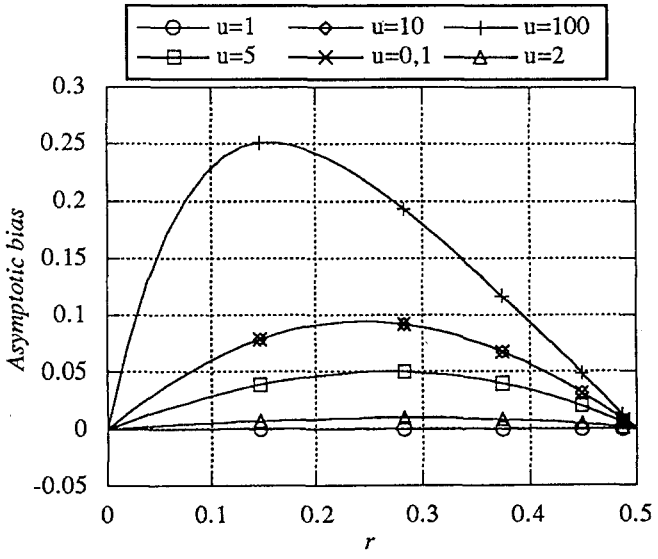


Fig. 3 Asymptotic bias of Bailey's estimate of the recombination fraction, r, between two markers flanking a gene selected with intensity u (backcross population, $n = 100$)

Consider three linked markers A, B and C. Let us call r_1, r_2 and r_3 the three recombination fractions between A and B, B and C, and A and C, respectively. Suppose that A and B are exactly located on two selected genes, and let u and v be the selection parameters of these two genes. Assuming no interference, the likelihoods of the three orders can be expressed as

$$e^L(ABC) = [(1-r_1)(1-r_2)]^{(a+b)}[(1-r_1)r_2]^{(c+d)} \times [r_1(1-r_2)]^{(e+f)}[r_1r_2]^{(g+h)}u^{(a+c+e+g)}v^{(a+c+f+h)}/D_1^n \quad (11)$$

$$e^L(ACB) = [(1-r_3)(1-r_2)]^{(a+b)}[(1-r_3)r_2]^{(g+h)} \times [r_3(1-r_2)]^{(e+f)}[r_3r_2]^{(c+d)}u^{(a+c+e+g)}v^{(a+c+f+h)}/D_2^n \quad (12)$$

$$e^L(BAC) = [(1-r_1)(1-r_3)]^{(a+b)}[(1-r_1)r_3]^{(c+d)} \times [r_1(1-r_3)]^{(g+h)}[r_1r_3]^{(e+f)}u^{(a+c+e+g)}v^{(a+c+f+h)}/D_1^n \quad (13)$$

where all parameters are replaced by their estimates, and where $D_1 = 1 + uv + r_1(u + v - uv - 1)$, $D_2 = (1 - r_2 - r_3 + 2r_2r_3)(uv + 1) + r_1(u + v)$, and $D_3 = D_1$. Putting an additional selection parameter, w, on marker C allows us to suppose that the two selected genes are on two adjacent markers, or alternatively on the two bordering markers. This leads to the observed and expected frequencies of Table 4, for the three possible orders ABC, ACB and BAC. The LOD score which tests the relative likelihoods of two orders is defined as the $\log_{10}$ of the ratio of the corresponding likelihoods, which are expressed in a similar manner to (11), (12) and (13). If $u = v = w = 1$, the LOD scores are simply the classical LODs. If only one parameter is different from one, the classical estimates of the recombination fractions stay consistent. Consequently, the LOD scores are unchanged, meaning that a single selected gene does not

Table 4 Expected and observed frequencies for a backcross in coupling, involving three markers, A, B and C, which are under selection of intensities u, v and w. All expected frequencies have to be multiplied by n, the population size

Phenotypes	Expected frequencies (order ABC)	Expected frequencies (order ACB)	Expected frequencies (order BAC)	Observed frequencies
ABC	$uvw(1-r_1)(1-r_2)/D_1$	$uvw(1-r_3)(1-r_2)/D_2$	$uvw(1-r_1)(1-r_3)/D_3$	a
abc	$(1-r_1)(1-r_2)/D_1$	$(1-r_3)(1-r_2)/D_2$	$(1-r_1)(1-r_3)/D_3$	b
ABc	$uv(1-r_1)r_2/D_1$	uvr_3r_2/D_2	$uv(1-r_1)r_3/D_3$	c
abC	$w(1-r_1)r_2/D_1$	wr_3r_2/D_2	$w(1-r_1)r_3/D_3$	d
Abc	$ur_1(1-r_2)/D_1$	$ur_3(1-r_2)/D_2$	ur_1r_3/D_3	e
aBC	$vwr_1(1-r_2)/D_1$	$vwr_3(1-r_2)/D_2$	vwr_1r_3/D_3	f
AbC	uwr_1r_2/D_1	$uw(1-r_3)r_2/D_2$	$uwr_1(1-r_3)/D_3$	g
aBc	vr_1r_2/D_1	$v(1-r_3)r_2/D_2$	$vr_1(1-r_3)/D_3$	h

$D_1 = (1-r_1)(1-r_2)(uvw+1) + r_1(1-r_2)(u+vw) + (1-r_1)r_2(w+uv) + r_1r_2(v+uw)$
$D_2 = (1-r_2)(1-r_3)(uvw+1) + r_2(1-r_3)(v+uw) + (1-r_2)r_3(u+vw) + r_2r_3(w+uv)$
$D_3 = (1-r_1)(1-r_3)(uvw+1) + r_1(1-r_3)(v+uw) + (1-r_1)r_3(w+uv) + r_1r_3(u+vw)$

Table 5 Expected LOD scores comparing the orders of three markers, *A*, *B* and *C*. $E[Z]$: ELODs calculated under hypothesis of no distortion; $E[Z(u, v, w)]$: ELODs under hypothesis of selection on markers *A* and *B* ($w = 1$) or *A* and *C* ($v = 1$). The values are calculated for $r_1 = r_2 = 0.1$ and for $n = 100$

u	v	w	$E[Z]$ *ABC/ACB*	$E[Z]$ *ABC/BAC*	$E[Z]$ *BAC/ACB*	$E[Z(u,v,w)]$ *ABC/ACB*	$E[Z(u,v,w)]$ *ABC/BAC*	$E[Z(u,v,w)]$ *BAC/ACB*
1	1	1	6.35	6.35	0	6.35	6.35	0
10	10	1	11.18	1.58	9.61	3.05	4.51	−1.46
20	20	1	12.31	0.81	11.49	2.01	6.12	−4.11
100	100	1	13.60	0.17	13.43	0.54	11.47	−10.92
10	0.1	1	0.63	14.87	−14.24	12.45	16.37	−3.91
20	0.05	1	0.02	15.94	−15.92	13.43	19.77	−6.34
100	0.01	1	4.44	8.87	−4.42	7.10	19.90	−12.79
10	1	10	1.29	1.29	0	3.68	3.68	0
20	1	20	−0.20	−0.20	0	3.29	3.29	0
100	1	100	−2.04	−2.04	0	2.96	2.96	0
10	1	0.1	4.77	4.77	0	13.03	13.03	0
20	1	0.05	−1.08	−1.08	0	16.15	16.15	0
100	1	0.01	−17.48	−17.48	0	20.57	20.57	0

modify the determination of order. On the other hand, if two parameters are different from one, then the classical estimates of the recombination fractions are severely biased, whereas Bailey's estimates are rarely biased; for example, if u and v are different from one, the following conclusions can be shown:

if the classical estimates are used, then:
$\hat{r}_1$ will be biased for all three orders,
$\hat{r}_2$ will be biased if the order is *BAC* or *ACB*,
$\hat{r}_3$ will be biased if the order is *ABC* or *ACB*.

if Bailey's estimates are used:
$\hat{r}_1$ will be consistent for all three orders,
$\hat{r}_2$ will be biased if the order is *BAC*,
$\hat{r}_3$ will be biased if the order is *ABC*.

The behavior of the LOD scores can be studied by calculating their values, the ELODs, when the observations (*a* to *h*) are replaced by their expectations. Table 5 shows the values of ELODs for $r_1 = r_2 = 0.1$, assuming that the true order is *ABC*. The $E[Z]$ are the expectations of the classical LOD scores, using the classical estimates of the recombination fractions. The $E[Z(u, v, w)]$ are the expectations of the LOD scores computed with formulas analogous to (11), (12) and (13), based on the expected frequencies of Table 4, and with Bailey's estimates of the recombination fractions. ELODs are given for different values of u, v and w, and for a population size of 100. The first line of Table 5 indicates the values of ELODs when no selection occurs ($u = v = w = 1$), while the other lines give the values for several models of selection. This table clearly indicates that when severe selection occurs on the two bordering markers, *A* and *C*, the signs of the classical LOD scores may be inverted, leading to false conclusions about order. On the other hand, the signs of the LOD scores using Bailey's estimate are not inverted. Therefore, the calculation of the likelihoods of the orders in case of segregation distortion should use Bailey's estimate, in conjunction with formulas taking into account the selection parameters. If $\hat{r}_B$ is not defined, due to the nullity of one recombinant class, then systems equivalent to (2) have to be solved iteratively.

Discussion

Maximum-likelihood estimates were derived for a two-point analysis of the recombination fraction in case of segregation distortion, for backcross populations. Their properties and usefulness are discussed below.

When segregation distortions are observed on two linked markers, we do not know if these distortions are due to their linkage with one or several selected genes. Tests such as (9) can be used in order to answer this question. If it has been found that only one selected gene is present, it is not necessary to use Bailey's estimate. Nevertheless, this estimate remains fully efficient even when no, or one, locus is selected. Moreover, Bailey's estimate was shown to be more often consistent than the classical estimate, even when the markers are not located on the genes selected. Hence, Bailey's estimate will have to be used for all linkage analyses and order determinations with segregation distortions in backcross populations. It should be noted that other types of selection may occur, where the rows or columns of the 2×2 contingency table are not entirely selected, but only for certain genotypes. These other types of selection could be due to epistatic effects such as complementary genes. For example, only the *aabb* genotypes may be in disfavor. In such cases, Bailey's estimate and Mather's χ^2 are biased. As a great number of such cases may occur, it would be very cumbersome to test for each possible model of selection. Moreover, the number of degrees of freedom available (three) would be insufficient to estimate r in certain cases. It can be shown that the χ^2 which tests the independence of two loci (Mather 1957) is not usable in such cases; thus, new methods have to be developed. A possible method of investigation in these cases is to take into account the

prior probability of linkage, i.e., Bayesian estimates (Neumann 1990).

Acknowledgements Many thanks for helpful discussions are due to L.B. Grivet and P.J.L. Lagoda.

References

Bailey NTJ (1949) The estimation of linkage with differential viability, II and III. Heredity 3:220–228

Bailey NTJ (1951) Testing the solubility of maximum likelihood equations in the routine application of scoring methods. Biometrics 7:268–274

Edwards AWF (1972) Likelihood. The John Hopkins University Press, Baltimore

Garcia-Dorado A, Gallego A (1992) On the use of the classical tests for detecting linkage. J. Hered 83:143–146

Lander ES, Green P (1987) Construction of multilocus genetic linkage maps in humans. Proc Natl Acad Sci USA 84:2363–2367

Lathrop GM, Lalouel J-M (1988) Efficient computations in multilocus linkage analysis. Am J Hum Genet 42:498–505

Lyttle TW (1991) Segregation distorters. Annu Rev Genet 25:511–557

Mather K (1957) The measurement of linkage in heredity. Methuen and Co., London

Neumann PE (1990) Two-locus linkage analysis using recombinant inbred strains and Bayes' theorem. Genetics 126:277–284

Ott J (1985) Analysis of human genetic linkage. John Hopkins Press, Baltimore

Schön CG, Hayes PM, Blake TK, Knapp SJ (1991) Gametophytic selection in a winter × spring barley cross. Genome 34:918–922

Torres AM, Mau-Lastovicka T, Williams TE, Soost RK (1985) Segregation distortion and linkage of *Citrus* and *Poncirus* isozyme genes. J Hered 76:289–294

Wendel JF, Parks CR (1984) Distorted segregation and linkage of alcohol dehydrogenase ec-1.1.1.1 genes in camellia-japonica theaceae. Biochem Genet 22:739–748

Zivy M, Devaux P, Blaisonneaux J, Jean R, Thiellement H (1992) Segregation distortion and linkage studies in microspore-derived double haploid lines of *Hordeum vulgare* L. Theor Appl Genet 83:919–924