

Self-Fertilization and the Evolution of Recombination

Denis Roze^{*,1} and Thomas Lenormand[†]

^{*}*Génétique et Evolution des Maladies Infectieuses, Institut de Recherche Pour le Développement, 34394 Montpellier, France and*
[†]*Génétique et Dynamique des Populations, Centre d'Ecologie Fonctionnelle et Evolutive (CNRS), 34293 Montpellier, France*

Manuscript received September 15, 2004
Accepted for publication February 9, 2005

ABSTRACT

In this article, we study the effect of self-fertilization on the evolution of a modifier allele that alters the recombination rate between two selected loci. We consider two different life cycles: under gametophytic selfing, a given proportion of fertilizations involves gametes produced by the same haploid individual, while under sporophytic selfing, a proportion of fertilizations involves gametes produced by the same diploid individual. Under both life cycles, we derive approximations for the change in frequency of the recombination modifier when selection is weak relative to recombination, so that the population reaches a state of quasi-linkage equilibrium. We find that gametophytic selfing increases the range of epistasis under which increased recombination is favored; however, this effect is substantial only for high selfing rates. Moreover, gametophytic selfing affects the relative influence of different components of epistasis (additive \times additive, additive \times dominance, dominance \times dominance) on the evolution of the modifier. Sporophytic selfing has much stronger effects: even a small selfing rate greatly increases the parameter range under which recombination is favored, when there is negative dominance \times dominance epistasis. This effect is due to the fact that selfing generates a correlation in homozygosity at linked loci, which is reduced by recombination.

SEX is the combination of two complementary events—meiosis and syngamy—resulting in the alternation of a diploid and a haploid phase during the life cycle; by contrast, an asexual life cycle can be viewed simply as a succession of mitoses. Both life cycles can be compared to understand the relative advantages of sexual *vs.* asexual reproduction, and this approach has been useful to identify the costs (MAYNARD SMITH 1971; WILLIAMS and MITTON 1973) and benefits of meiosis owing to the processes of segregation and recombination (see below). However, a whole variety of sexual life cycles does exist, depending in particular on how syngamy occurs. Although both the evolution of mating systems and the evolution of sex and recombination have received considerable attention, the interplay between mating systems and the evolution of recombination has received less attention.

The famous twofold cost of meiosis notoriously depends on the mating system (CHARLESWORTH 1980); it can, for example, vanish in selfing hermaphrodites. Similarly, the advantage of sexual reproduction due to segregation greatly varies depending on the mating system (UYENOYAMA and BENGTTSSON 1989; AGRAWAL and CHASNOV 2001; OTTO 2003). However, despite considerable work and progress in understanding when recombination may evolve and provide an advantage to

sexual reproduction, we still largely ignore how the mating system influences the evolution of recombination. Some simulations have been performed in this direction (CHARLESWORTH *et al.* 1977, 1979; HOLSINGER and FELDMAN 1983), but no clear conclusion has emerged as to exactly how nonrandom mating might affect the general conditions under which recombination is favored.

Different theories have been proposed to explain the evolution and maintenance of high rates of recombination in higher organisms. Apart from theories based on the idea of a mechanistic advantage of recombination (*e.g.*, through DNA repair), recombination is thought to be beneficial because it reduces linkage disequilibria (LD) between loci. This last hypothesis (termed a generative hypothesis) consists of a variety of possible underlying processes that can be classified according to the mechanisms generating LD (KONDRASHOV 1993; OTTO and LENORMAND 2002). First, LD can be produced by epistatic selection, in which case recombination is favored when epistasis is weak and negative (FELDMAN *et al.* 1980; KONDRASHOV 1982, 1988; BARTON 1995) and not too variable among pairs of loci (OTTO and FELDMAN 1997) or when epistasis fluctuates over short periods of time (CHARLESWORTH 1976; BARTON 1995). Second, LD can be produced by migration and directional selection, in which case recombination is favored when directional selection at different loci covaries negatively between habitats (LENORMAND and OTTO 2000). Third, LD can be produced by drift and directional selection, a process referred to as the Hill-Robertson effect (FELSENSTEIN

¹*Corresponding author:* School of Biological Sciences, University of Edinburgh, King's Bldgs., W. Mains Rd., Edinburgh EH9 3JT, United Kingdom. E-mail: denis.roze@ed.ac.uk

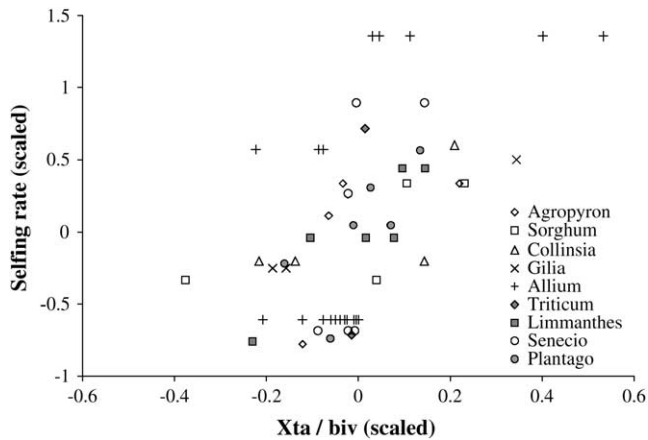


FIGURE 1.—Cytological data on selfing rates and numbers of chiasmata per bivalent at meiosis. Each symbol corresponds to a different species or subspecies. x -axis, deviation of the number of chiasmata per bivalent relative to the genus average, divided by the genus average; y -axis, deviation of the degree of selfing relative to the genus average, divided by the genus average. Data are from GRANT (1958) (Agropyron, Sorghum, Collinsia, and Gilia), VED BRAT (1965) (Allium), ZARCHI *et al.* (1972) (Triticum), ARROYO (1973) (Limnanthes), GIBBS *et al.* (1975) (Senecio), and SHARMA *et al.* (1992) (Plantago).

1974), in which case recombination is favored when populations are not too large (FISHER 1930; MULLER 1932; OTTO and BARTON 1997, 2001). These different sources of LD can also occur simultaneously, which affects the conditions for the evolution of recombination (LENORMAND and OTTO 2000; OTTO and BARTON 2001).

To date, most analytical models on the evolution of recombination have considered randomly mating populations; however, biological data and simulation models suggest that nonrandom mating may be an important factor. These data and models have focused on the case of partial self-fertilization. For example, cytological observations indicate that self-pollinating species of flowering plants tend to have higher chiasma frequencies than closely related outcrossing species, suggesting that recombination rates may correlate with the amount of self-fertilization (GRANT 1958; STEBBINS 1958; VED BRAT 1965; ZARCHI *et al.* 1972; ARROYO 1973; GIBBS *et al.* 1975; SHARMA *et al.* 1992). Figure 1 displays data collected from these studies. In general, quantitative estimates of rates of self-fertilization are difficult to obtain, but the different species can be classified into different broad classes, from “complete outcrossers” (for example, in the case of auto-incompatible species) to “complete selfers” (for example, in the case of cleistogamous species). From this classification, we attributed a number to each species, representing its degree of selfing. The y -axis of Figure 1 represents the difference between the degree of selfing of each species and the average over all measured species of the same genus, divided by the genus average. The x -axis represents the number of chiasmata per bivalent (measured in general at meta-

phase I of meiosis); again, each value corresponds to the deviation of each species relative to the genus average, divided by the genus average. Variables on both axes are strongly correlated in every genus and overall (Spearman Rho is 0.535 overall on the 57 points, $P < 10^{-4}$). However, it is important to note here that all these data have been obtained from male meioses; data from female meioses are thus needed to determine if the same correlation exists in both cases, given that the ratio of the number of chiasmata in male/female meioses also covaries with the mating system (LENORMAND and DUTEIL 2005).

Furthermore, simulation models have shown that self-fertilization increases the strength of indirect selection acting on recombination modifiers and that under some conditions modifiers increasing recombination decrease in frequency under random mating, but increase in frequency under partial selfing, even when the selfing rate is small (CHARLESWORTH *et al.* 1977, 1979; HOLSINGER and FELDMAN 1983). However, the mechanisms by which self-fertilization affects the evolution of recombination remain difficult to understand from these simulations.

In this article, we present an analytical study of the effects of partial selfing on the evolution of recombination, using multilocus techniques that have been applied previously to the study of the evolution of recombination (BARTON 1995; OTTO and FELDMAN 1997; LENORMAND 2003). Our model is deterministic (infinite population size), and therefore, the evolution of recombination will be driven by epistatic interactions between loci. We consider two different life cycles. Under gametophytic selfing, we assume that a given proportion of fertilizations involves gametes produced by the same *haploid* individual (generating zygotes, which are homozygous at all loci). Gametophytic selfing corresponds to the most extreme form of inbreeding, occurring, for example, in homosporous ferns (McCAULEY *et al.* 1985; SOLTIS and SOLTIS 1986). It has also been used by OTTO (2003) to consider the effects of inbreeding on segregation and the evolution of sex. Under gametophytic selfing, we assume that selection can occur during both the haploid and the diploid stages of the life cycle. The second life cycle corresponds to sporophytic selfing: in this case, a given proportion of fertilizations involves gametes produced by the same *diploid* individual (as in most selfing plants and animals). Under sporophytic selfing, we assume that selection occurs only during the diploid phase of the life cycle. We will see that gametophytic selfing tends to increase the range of epistasis values where recombination is favored; it also increases strongly selection pressures on recombination and affects the relative importance of the different components of epistasis (additive \times additive, additive \times dominance, dominance \times dominance). Under sporophytic selfing, dominance \times dominance epistasis generates a selective pressure on the evolution of recombination that is absent under random mating. If dominance \times

TABLE 1
Notation

i, j, k	Indices for loci; j and k are selected loci, and i modifies the recombination rate between j and k
$p_l, q_l = 1 - p_l$	Frequency of allele 1 (p_l) and of allele 0 (q_l) at locus l
X_l, X_l^*	Indicator variables that equal 1 if allele 1 is present at locus l on the maternally inherited chromosome (X_l) or paternally inherited chromosome (X_l^*) of a given individual and 0 otherwise
$\zeta_l = X_l - p_l$	Scaled indicator variable
ζ_U	Product of ζ -variables over the set of loci U
C_U, C_U^H	Genetic association between loci in the set U during the haploid phase (in the gametophytic selfing model), before and after haploid selection
$C_{U,V}^S, C_{U,V}^D$	Genetic association between loci in the sets U and V on homologous chromosomes of diploid individuals, before and after diploid selection (in the sporophytic selfing model, $C_{U,V}^S$ is noted $C_{U,V}$)
a_U^H, e_U^H	Effect of haploid selection acting on the set of loci U (in the gametophytic selfing model) and epistasis between these loci on a multiplicative scale
$a_{U,V}^D, e_{U,V}^D$	Effect of diploid selection acting on the sets of loci U and V present on homologous chromosomes and epistasis between these loci on a multiplicative scale; D superscripts are dropped in the sporophytic selfing model
s	Direct selection acting on loci j and k , under the fitness matrix given by Table 2
h	Dominance of alleles 1 at loci j and k , under the fitness matrix given by Table 2
$e_{a \times a}, e_{a \times d}, e_{d \times d}$	Additive \times additive, additive \times dominance, and dominance \times dominance epistasis between loci j and k , under the fitness matrix given by Table 2
α	Rate of self-fertilization (either gametophytic or sporophytic)
r_{ij}, r_{jk}	Recombination rate between loci i and j and between loci j and k (without modifier)
dr	Effect of the modifier on r_{jk}
h_M	Dominance of allele 1 at the modifier locus (locus i)
$t_{S,T}$	Probability that a gamete produced by an individual contains the set S of loci derived from one of its parents and the set T from the other
$\bar{t}_{S,T}$	Population average of $t_{S,T}$
$\delta t_{S,T i}, \delta t_{S,T i^*}$	Additive effect of the modifier on $t_{S,T}$ and effect of dominance at the modifier locus
ϵ	Scaling factor

dominance epistasis is negative, even a small rate of sporophytic selfing greatly increases the range of parameters under which recombination can increase.

METHODS

Throughout the article, we consider three linked loci, i, j , and k (present in that order along the chromosome), where j, k are selected loci, while locus i affects the recombination rate between j and k . Two alleles, noted 0 and 1, segregate at each locus. We use the multilocus formalism of BARTON and TURELLI (1991) (extended by KIRKPATRICK *et al.* 2002) to write recurrence equations for allele frequencies and for genetic associations, measuring the statistical associations among alleles at different loci and/or on different chromosomes. Definitions of the different parameters and variables are given in Table 1. Following previous usage, we denote p_l the frequency of allele 1 at locus l in the population, and $q_l = 1 - p_l$. We then define two indicator variables X_l and X_l^* , which equal 1 if allele 1 is present at locus l on the maternally inherited (X_l) or paternally inherited (X_l^*) chromosome of a given individual and 0 otherwise. In the haploid stage, we have only one X_l variable per haploid individual. ζ variables are defined as

$$\zeta_l = X_l - p_l, \quad \zeta_l^* = X_l^* - p_l, \quad (1)$$

and products of ζ -variables are given by

$$\zeta_U = \prod_{l \in U} \zeta_l, \quad \zeta_U^* = \prod_{l \in U} \zeta_l^*, \quad (2)$$

where U is a set of loci, the possible sets being $\emptyset, \{i\}, \{j\}, \{k\}, \{i, j\}, \{i, k\}, \{j, k\}$, and $\{i, j, k\}$. By convention, $\zeta_{\emptyset} = \zeta_{\emptyset}^* = 1$. Genetic associations C_U measured in the haploid phase (before haploid selection) are defined as

$$C_U = E[\zeta_U], \quad (3)$$

where U is a set of loci, and where E means the average over all individuals in the population. Genetic associations in diploid organisms are given by

$$C_{U,V} = E[\zeta_U \zeta_V^*], \quad (4)$$

where U and V are sets of loci. For example, C_{ij} measures the linkage disequilibrium between loci i and j in the haploid phase, while $C_{ij\emptyset}$ measures the linkage disequilibrium between i and j on maternal chromosomes, during the diploid phase. Throughout, we assume no sex-specific selection nor any effect of the sex-of-origin of chromosomes, and therefore we will always have $C_{U,V} = C_{V,U}$.

Gametophytic selfing: The life cycle under gametophytic selfing is represented in Figure 2A. Each generation starts after meiosis, at the beginning of the haploid phase. In this phase, the population can be described by seven haploid genotype frequencies (there are eight genotypes, and their frequencies sum to 1) or by the three allele frequencies p_i, p_j , and p_k and the four gametic linkage disequilibria C_{ij}, C_{ik}, C_{jk} , and C_{ijk} . We assume that selection can occur during the haploid phase, followed by syngamy. At syngamy, a proportion α of

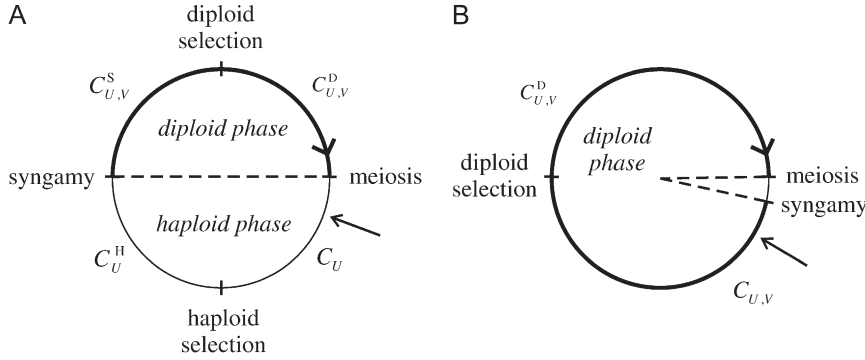


FIGURE 2.—Model life cycles. (A) Gametophytic selfing; (B) sporophytic selfing. In each case, the arrow marks the start of the life cycle, at which point we measure allele frequencies and genetic associations.

zygotes are formed by two gametes produced by the same haploid individual (and are thus homozygous at all loci), while a proportion $1 - \alpha$ are formed by gametes sampled at random. Finally, selection occurs between diploid organisms, followed by meiosis. Changes in allele frequencies and genetic associations caused by selection have been derived in KIRKPATRICK *et al.* (2002), while recursions for meiosis in the presence of a recombination modifier are given in BARTON (1995). Therefore, the only new part here corresponds to the recursions for syngamy. In the following, we use superscripts H, S, and D to denote variables measured after haploid selection, syngamy, and diploid selection, respectively.

Haploid selection: To represent haploid selection, BARTON and TURELLI (1991) and KIRKPATRICK *et al.* (2002) define selection coefficients a_U^H such that the fitness of a haploid individual, relative to the average fitness in the population, can be written as

$$1 + \sum_U a_U^H (\zeta_U - C_U). \quad (5)$$

a_U^H represents the effect of selection acting on the set of loci U ; since we assume that selection acts only at loci j and k , we have only three coefficients a_j^H , a_k^H , and a_{jk}^H , where a_j^H and a_k^H represent direct selection at loci j and k , while a_{jk}^H represents the interaction between these loci.

Recursions for allele frequencies and genetic associations, using these a_U^H coefficients are given by Equations 9, 10, and 15 in KIRKPATRICK *et al.* (2002). In particular, the change in frequency of the recombination modifier during haploid selection is given by

$$\Delta_H p_i = \sum_U a_U^H C_{Ui} = a_j^H C_{ij} + a_k^H C_{ik} + a_{jk}^H C_{ijk}. \quad (6)$$

Syngamy: Allele frequencies do not change during syngamy. Associations after syngamy are given by

$$C_{U,V}^S = \alpha C_{U,V}^H + (1 - \alpha) C_U^H C_V^H, \quad (7)$$

where α is the rate of gametophytic selfing, and U and V are sets of loci. For instance, $C_{ij,\emptyset}^S = C_{ij}^H$, $C_{ij}^S = \alpha C_{ij}^H$, while $C_{jk,jk}^S = \alpha C_{jk,jk}^H + (1 - \alpha)(C_{jk}^H)^2$. Repeated indices in associations can be eliminated using the relation

$$C_{Ull} = p_l q_l C_U + (1 - 2p_l) C_{Ul} \quad (8)$$

(KIRKPATRICK *et al.* 2002, Equation 5), so that, for example, $C_{jkk} = p_j q_j p_k q_k + (1 - 2p_j)(1 - 2p_k) C_{jk}$.

Diploid selection: Diploid selection is represented by defining $a_{U,V}^D$ selection coefficients, such that the fitness of a diploid organism, relative to the average fitness of the population, can be written as

$$1 + \sum_{U,V} a_{U,V}^D (\zeta_{U,V} - C_{U,V}^S). \quad (9)$$

The sum in Equation 9 is over all sets of loci U and V , including the empty set. Again, $a_{U,V}^D$ represents the effect of selection acting on the set of loci (U, V) ; for example $a_{j\emptyset}^D$ and $a_{k\emptyset}^D$ represent direct selection acting on loci j and k , while a_{jj}^D and a_{kk}^D represent the effects of dominance at these loci (in the additive case, $a_{jj}^D = a_{kk}^D = 0$). In principle, the notation allows for different selective effects depending upon the sex-of-origin of chromosomes (for example, it is possible to have $a_{jj}^D \neq a_{jj}^D$), but because we assume no sex-of-origin effect, $a_{U,V}^D = a_{V,U}^D$ for all sets U and V . Interactions between pairs of genes at loci j and k are represented by coefficients $a_{jk,\emptyset}^D$ and a_{jk}^D ; without *cis-trans* effects, $a_{jk,\emptyset}^D = a_{jk}^D$. Finally, interactions among three and four alleles at loci j and k are represented by coefficients a_{jkk}^D , a_{jjk}^D , and a_{jjkk}^D .

The exact expressions of $a_{U,V}^D$ coefficients are complicated functions of the fitnesses and frequencies of the different genotypes, but they take simpler forms when calculated to first order in the selective differences. Throughout the article, we consider the case of directional selection acting on loci j and k and use the symmetric fitness matrix given in Table 2, where s is the selection coefficient of alleles 1 at loci j and k , h is their dominance coefficient, and $e_{a \times a}$, $e_{a \times d}$, and $e_{d \times d}$ measure additive \times additive, additive \times dominance, and dominance \times dominance epistasis between loci j and k . Without epistasis ($e_{a \times a} = e_{a \times d} = e_{d \times d} = 0$) and under random mating, this selection regime does not generate any gametic linkage disequilibrium between loci j and k ($C_{jk} = 0$). $a_{U,V}^D$ coefficients can be obtained by equating the fitnesses of the different genotypes with expressions derived from Equation 9; APPENDIX A gives expressions for the $a_{U,V}^D$ coefficients under the fitness matrix given in Table 2 to different orders in s , $e_{a \times a}$, $e_{a \times d}$, and $e_{d \times d}$.

Recursions for allele frequencies and genetic associa-

TABLE 2
Example of fitness matrix for diploid selection

	00	01	11
00	1	$1 + hs$	$1 + s$
01	$1 + hs$	$(1 + hs)^2 + e_{a \times a}$	$(1 + hs)(1 + s) + 2e_{a \times a} + e_{a \times d}$
11	$1 + s$	$(1 + hs)(1 + s) + 2e_{a \times a} + e_{a \times d}$	$(1 + s)^2 + 4e_{a \times a} + 4e_{a \times d} + e_{d \times d}$

The entries correspond to the different genotypes at loci j and k ; $e_{a \times a}$, $e_{a \times d}$, and $e_{d \times d}$ correspond to additive \times additive, additive \times dominance, and dominance \times dominance epistatic effects, respectively.

tions after diploid selection, in terms of the $a_{U,V}^D$ coefficients, are given by Equations 9, 10, and 15 in KIRKPATRICK *et al.* (2002). In particular, the change in frequency of the recombination modifier is given by

$$\Delta_D p_i = \sum_{U,V} a_{U,V}^D C_{U,V}^S, \quad (10)$$

where the sum is over all sets U and V , including the empty set.

Meiosis: We denote the recombination rate between loci i and j as r_{ij} . The recombination rate between j and k depends on the individual's genotype at locus i : individuals having genotype 00, 01, and 11 at this locus have recombination rate r_{jk} , $r_{jk} + h_M dr$, and $r_{jk} + dr$, respectively, between loci j and k . Therefore, dr measures the effect of the recombination modifier, while h_M is the dominance of allele 1 at locus i . We assume no interference among crossing-over events.

Meiosis does not change allele frequencies. A general recursion for genetic associations in the presence of a recombination modifier is given in BARTON (1995). The method consists in defining coefficients $t_{S,T}$, measuring the probability that a gamete produced by an individual contains the set S of loci derived from one of its parents and the set T from the other. For example, $t_{i,jk}$ is the proportion of gametes containing locus i from one parent (either mother or father) and loci j and k from the other. In the presence of a recombination modifier, some $t_{S,T}$ will depend on the individual genotype at locus i and can be written under the form

$$t_{S,T} = \bar{t}_{S,T} + \delta t_{S,T|i}(\zeta_i + \zeta_i^*) + \delta t_{S,T|i,i} \zeta_{ii}, \quad (11)$$

where $\bar{t}_{S,T}$ is the population average, $\delta t_{S,T|i}$ represents the additive effect of the modifier on $t_{S,T}$, while $\delta t_{S,T|i,i}$ represents the effect of dominance at the modifier locus. For all S and T , $\bar{t}_{S,T}$, $\delta t_{S,T|i}$, and $\delta t_{S,T|i,i}$ can be expressed as functions of r_{ij} , r_{jk} , dr , h_M , and p_i (BARTON 1995). Genetic associations after meiosis are then given by

$$\begin{aligned} C'_X &= E \left[\sum_{S+T=X}^* t_{S,T} \zeta_{S,T} \right] \\ &= \sum_{S+T=X}^* \left[\bar{t}_{S,T} C_{S,T}^D + \delta t_{S,T|i} (C_{S_i,T}^D + C_{S,T_i}^D) + \delta t_{S,T|i,i} C_{S_i,T_i}^D \right] \end{aligned} \quad (12)$$

(where E means the average over all individuals, just before meiosis). To go from the first to the second line, one uses Equation 11 and replaces expectations of products of ζ -variables by associations before meiosis. As in BARTON (1995), the asterisk in Equation 12 means that the sum is over all distinct partitions of the set X . For example, for $X = ijk$, the partitions $\{S = i, T = jk\}$ and $\{S = jk, T = i\}$ contribute to a single term in the sum. This stands in contrast to sums like in Equation 10, where, for example, $\{U = jk, V = j\}$ and $\{U = j, V = jk\}$ represent two different terms.

Sporophytic selfing: The life cycle under sporophytic selfing is represented in Figure 2B. Each generation starts after fertilization; selection then occurs (in the diploid phase), followed by meiosis and syngamy. A proportion α of fertilizations involves gametes produced by the same diploid individual, while a proportion $1 - \alpha$ involves gametes sampled at random from the whole population. Under sporophytic selfing, the population cannot be described only in terms of haplotype frequencies: one has to keep track of 36 diploid genotype frequencies or of 3 allele frequencies and of 32 genetic associations (measured at the beginning of the life cycle, in juvenile diploids): 9 associations between pairs of genes (such as $C_{i,i}$, $C_{ij,\emptyset}$, $C_{i,h}$), 10 associations between three genes (such as $C_{ijk,\emptyset}$, $C_{ij,i}$, $C_{ih,j}$), 9 associations between four genes (such as $C_{ijk,i}$, $C_{ij,ij}$, $C_{ih,jk}$), 3 associations between five genes ($C_{ijk,ij}$, $C_{ijk,ik}$, $C_{ijk,jk}$), and the association between the six genes, $C_{ijk,ijk}$. Because of this higher number of variables, the analysis of the model becomes more tedious, but nevertheless leads to simple results, at least to leading order in the dominance \times dominance epistasis.

As in the previous section, recursions for diploid selection are obtained by defining $a_{U,V}^D$ selection coefficients and using Equations 9, 10, and 15 in KIRKPATRICK *et al.* (2002). Recursions for meiosis and syngamy are obtained as follows. Recursions for associations between genes present on the same chromosome are not affected by selfing and are given by

$$C'_{X,\emptyset} = \sum_{S+T=X}^* \left[\bar{t}_{S,T} C_{S,T}^D + \delta t_{S,T|i} (C_{S_i,T}^D + C_{S,T_i}^D) + \delta t_{S,T|i,i} C_{S_i,T_i}^D \right], \quad (13)$$

which is equivalent to Equation 12. Recursions for asso-

ciations between genes present on homologous chromosomes are then given by

$$\begin{aligned}
C'_{XY} = & (1 - \alpha) C'_{X\emptyset} C'_{Y\emptyset} \\
& + \frac{\alpha}{2} \sum_{S+T=X}^* \sum_{U+V=Y}^* [\bar{t}_{ST} \bar{t}_{UV} (C_{SU,TV}^D + C_{SV,TU}^D) \\
& + (\delta t_{sT} \bar{t}_{UV} + \delta t_{UV} \bar{t}_{sT}) \\
& \times (C_{SU,TV}^D + C_{SU,Tv}^D + C_{Sv,TU}^D + C_{SV,TU}^D) \\
& + (\delta t_{sT} \bar{t}_{UV} + \delta t_{UV} \bar{t}_{sT}) (C_{SU,TV}^D + C_{SV,TU}^D)] + o(dr).
\end{aligned} \tag{14}$$

This expression is obtained using the same method as in BARTON (1995): one considers all the possible recombination events that can have produced each of the two chromosomes of a newly formed zygote; if this zygote has been produced by selfing, one thus has to sum over all products t_{sT} and t_{UV} (such that $S + T = X$ and $U + V = Y$).

Approximations: To obtain recursions on allele frequencies and genetic associations, we implemented the general expressions described above in a Mathematica 5.0 notebook (available upon request); these full recursions are complicated expressions, which are not given here for space reasons. From these expressions, useful approximations can be obtained when selection is weak relative to recombination: in this case, genetic associations equilibrate fast relative to the change in allele frequencies, and the population quickly reaches a state described as *quasi-linkage equilibrium* (QLE), where associations change slowly over time (BARTON and TURELLI 1991; NAGYLAKI 1993). In principle, self-fertilization must not be too high for the QLE approximation to hold, because selfing reduces the effectiveness of recombination. Expressions of genetic associations at QLE, for given allele frequencies, are obtained by solving recursions for associations at equilibrium (that is, solving equations of the form $C'_X = C_X$ under gametophytic selfing, and $C'_{XY} = C_{XY}$ under sporophytic selfing). This way, associations can be expressed in terms of allele frequencies and of the different parameters of the model.

We consider different assumptions concerning the strength of selection and the frequency of self-fertilization; to do this, we introduce a small scaling term ε and assume that the modifier effect dr is of order ε , while selection coefficients a_{UV} and the selfing rate α can be of different orders in ε . At QLE, genetic associations will be of different orders, depending on the assumptions made. To obtain approximations to leading orders of associations at QLE, we perform a perturbation analysis: we first do not make any assumption on the order of magnitude of genetic associations and look for terms that are of order 1 at QLE (they are obtained by solving recursions expressed to the order 1). We then assume that all associations are of order ε , except those that were found to be of order 1 (which are replaced by their expression to the order 1, plus a term in ε), and solve recursions expressed to the order ε , to find associations that are effectively of order ε at QLE. We repeat

the same procedure for the orders ε^2 , ε^3 , \dots , until all associations needed to express the change in frequency of the modifier have been obtained.

RESULTS

Gametophytic selfing: Under gametophytic selfing we obtained that, as under random mating, recombination only decreases when selection and epistasis are of the same order of magnitude (as is apparent from our results). To find cases where recombination increases, one has to assume that epistasis is weak relative to selection. Therefore, throughout this section, we assume that selection coefficients a_j^H , a_k^H , $a_{j\emptyset}^D$, $a_{k\emptyset}^D$, a_{jj}^D , and a_{kk}^D are of order ε , while a_{jk}^H , $a_{jk\emptyset}^D$, $a_{jk,k}^D$, $a_{jk,j}^D$, $a_{jk,k}^D$, and $a_{jk,jk}^D$ are of order ε^2 . We also assume that the modifier has a small effect: dr is of order ε . We do not make any assumption on α . In this case, after solving recursions to express genetic associations at QLE, one obtains that C_{ij} and C_{ik} are of order ε^4 , C_{jk} is of order ε^2 , and C_{ijk} is of order ε^3 . Recursions for these associations to leading orders in ε are given in APPENDIX B.

The coefficients $a_{jk\emptyset}^D$, $a_{jk,k}^D$, $a_{jk,j}^D$, $a_{jk,k}^D$, and $a_{jk,jk}^D$ measure additive epistatic effects (as relative fitness in Equation 9 is written as a sum over these coefficients); to the order ε^2 , they depend both on multiplicative epistatic parameters (Table 2, $e_{a \times a}$, $e_{a \times d}$, and $e_{d \times d}$) and on direct selection. However, these coefficients can be written in the form

$$\begin{aligned}
a_{jk\emptyset}^D &= a_{j\emptyset}^D a_{k\emptyset}^D + e_{jk\emptyset}^D, & a_{jk,k}^D &= a_{j\emptyset}^D a_{k\emptyset}^D + e_{jk,k}^D, \\
a_{jk,j}^D &= a_{jj}^D a_{k\emptyset}^D + e_{jk,j}^D
\end{aligned} \tag{15}$$

$$a_{jk,k}^D = a_{k,k}^D a_{j\emptyset}^D + e_{jk,k}^D, \quad a_{jk,jk}^D = a_{jj}^D a_{k,k}^D + e_{jk,jk}^D, \tag{16}$$

where e_{UV}^D coefficients represent multiplicative epistatic effects (for example, $e_{jk,k}^D = e_{d \times d}$). This can be seen from the expressions given in APPENDIX A. Similarly, in the case of haploid selection, a_{jk}^H can be written in the form $a_j^H a_k^H + e_{jk}^H$, where e_{jk}^H represents multiplicative epistasis (BARTON and TURELLI 1991).

General solution at QLE: The change in frequency at the modifier locus is obtained from Equations B1, B2, B3, B5, and B7 in APPENDIX B:

$$\Delta p_i = A_j C_{ij} + A_k C_{ik} + (E + A_j A_k) C_{ijk} + o(\varepsilon^5), \tag{17}$$

with

$$A_j \equiv a_j^H + (1 + \alpha) a_{j\emptyset}^D + \alpha(1 - 2p_j) a_{jj}^D, \tag{18}$$

$$A_k \equiv a_k^H + (1 + \alpha) a_{k\emptyset}^D + \alpha(1 - 2p_k) a_{kk}^D, \tag{19}$$

and

$$\begin{aligned}
E \equiv & e_{jk}^H + (1 + \alpha) e_{jk\emptyset}^D + 2\alpha e_{jk,k}^D + 2\alpha e_{jk,j}^D (1 - 2p_j) \\
& + 2\alpha e_{jk,k}^D (1 - 2p_k) + \alpha e_{jk,jk}^D (1 - 2p_j) (1 - 2p_k) \\
& + \alpha(1 - \alpha) [a_{j\emptyset}^D + (1 - 2p_j) a_{jj}^D] [a_{k\emptyset}^D + (1 - 2p_k) a_{kk}^D].
\end{aligned} \tag{20}$$

QLE values of C_{ij} , C_{ik} , and C_{ijk} (denoted with a hat) are obtained from the recursions given in APPENDIX B. Assuming that the modifier has an additive effect ($h_M = 1/2$), we obtain

$$\hat{C}_{ijk} = -\frac{drE pq_{ijk}}{2r_{jk}r_{ijk}(1-\alpha)} + o(\epsilon^3) \quad (21)$$

$$\hat{C}_{ij} = \left[\frac{A_k}{r_{ij}(1-\alpha)} - (a_k^H + a_{k,\emptyset}^D) \right] \hat{C}_{ijk} + o(\epsilon^4) \quad (22)$$

$$\hat{C}_{ik} = \left[\frac{A_j}{r_{ik}(1-\alpha)} - (a_j^H + a_{j,\emptyset}^D) \right] \hat{C}_{ijk} + o(\epsilon^4), \quad (23)$$

where $pq_{ijk} = p_i q_i p_j q_j p_k q_k$. In Equation 21, r_{ijk} is the probability that at least one crossing over occurs between loci i , j , and k , in the absence of the modifier. Since we assume no interference between recombination events between i and j and between j and k , we have $r_{ijk} = 1 - (1 - r_{ij})(1 - r_{jk})$. Equations 17–23 give the frequency change at the modifier locus at QLE,

$$\Delta p_i = -\frac{drE(E + \Gamma)pq_{ijk}}{2r_{jk}r_{ijk}(1-\alpha)} + o(\epsilon^5), \quad (24)$$

where

$$\Gamma \equiv A_j A_k \left(\frac{1}{r_{ij}(1-\alpha)} + \frac{1}{r_{ik}(1-\alpha)} + 1 \right) - A_j (a_k^H + a_{k,\emptyset}^D) - A_k (a_j^H + a_{j,\emptyset}^D). \quad (25)$$

Equation 24 shows that, as under random mating, recombination always decreases when direct selection and epistasis are of the same order of magnitude. Indeed in this case Γ becomes negligible relative to E , and Δp_i has the same sign as $-dr$. In the absence of selfing ($\alpha = 0$), Equation 24 corresponds to the result obtained by BARTON (1995) for the change in frequency of a recombination modifier in a panmictic population (a modifier increasing recombination increases in frequency when $\Gamma < E < 0$). Finally, Δp_i under arbitrary dominance at the modifier locus is given by Equation 24, multiplied by a factor $2[h_M + p_i(1 - 2h_M)]$.

QLE under haploid selection: When selection occurs only during the haploid phase of the life cycle, Equation 24 becomes

$$\Delta p_i = -\frac{dre_{jk}^H pq_{ijk}}{2r_{jk}r_{ijk}(1-\alpha)} \left[e_{jk}^H + a_j^H a_k^H \left(\frac{1}{r_{ij}(1-\alpha)} + \frac{1}{r_{ik}(1-\alpha)} - 1 \right) \right] + o(\epsilon^5). \quad (26)$$

Equation 26 takes the same form as under random mating (Equation 12 in BARTON 1995), all recombination rates (dr , r_{ij} , r_{ik} , r_{jk} , and r_{ijk}) being multiplied by a factor $1 - \alpha$; this reflects the fact that selfing reduces the effectiveness of recombination. As a result, the minimum value of epistasis for a modifier increasing recombination to be selected, given by the term between brackets in Equation 26, decreases as α increases. Therefore,

under haploid selection, gametophytic selfing increases the range of negative epistasis values over which increased recombination is selected.

We tested the validity of Equation 26, using a model of directional selection, where haploid individuals carrying zero, one, and two alleles 1 at loci j and k have fitnesses 1, $1 + s$, and $(1 + s)^2 + e$, respectively. In this case, when s is of order ϵ and e of order ϵ^2 , one obtains that $a_j^H a_k^H = s^2 + o(\epsilon^2)$, and $e_{jk}^H = e + o(\epsilon^2)$; thus, Equation 26 predicts that a modifier increasing recombination should increase in frequency when

$$-s^2 \left(\frac{1}{r_{ij}(1-\alpha)} + \frac{1}{r_{ik}(1-\alpha)} - 1 \right) < e < 0. \quad (27)$$

Figure 3 compares this prediction with deterministic simulations. In the simulations we iterate exact recursions for genotype frequencies; we first let the system equilibrate in the absence of the modifier and then introduce the modifier in small frequency and see if it increases (Figure 3, solid circles) or decreases (Figure 3, open circles) in frequency. Simulations confirm that gametophytic selfing increases the range of (negative) epistasis values over which increased recombination is favored; they also show that Equation 27 gives good predictions when selection is weak (Figure 3A, $s = -0.01$), except when the selfing rate is close to 1, while it does not give such good predictions for stronger selection (Figure 3B, $s = -0.05$), when the selfing rate is moderate to high. This is due to the fact that selfing reduces the effectiveness of recombination and therefore reduces the range of selection coefficients and recombination rates under which the QLE approximation is valid.

QLE under diploid selection: In the case of diploid selection, it can be seen from Equations 20 and 24 that self-fertilization generates indirect selection on the recombination modifier even in the absence of epistasis (when all $e_{i,j}^{D,V}$ coefficients equal zero). For example, without epistasis or dominance, $E = \alpha(1 - \alpha)a_{j,\emptyset}^D a_{k,\emptyset}^D$, and Equation 24 shows that a modifier increasing recombination will always be selected against (when $a_{j,\emptyset}^D$ and $a_{k,\emptyset}^D$ have the same sign). Indeed, self-fertilization generates positive linkage disequilibrium between loci j and k , even in the absence of epistasis; at QLE, one obtains (without epistasis or dominance)

$$\hat{C}_{jk} = \frac{\alpha a_{j,\emptyset}^D a_{k,\emptyset}^D p_j q_j p_k q_k}{r_{jk}} + o(\epsilon^2). \quad (28)$$

Linkage disequilibrium C_{jk} is due to the fact that, even under a completely multiplicative fitness matrix, the marginal fitnesses of the different types of chromosomes are not multiplicative, because under partial selfing genotype frequencies are not given by simple products of gamete frequencies.

To consider the effects of dominance and epistasis under diploid selection, we used the fitness matrix given

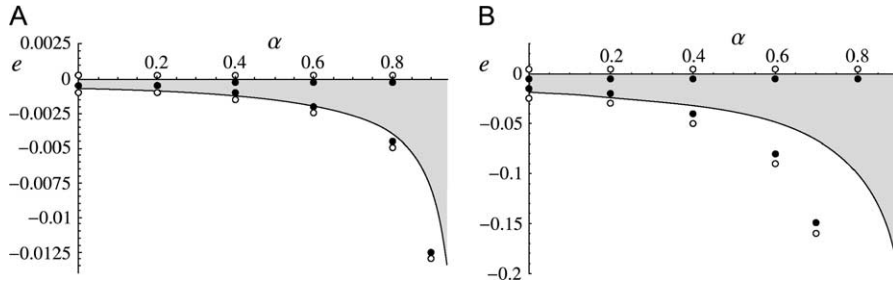


FIGURE 3.—Range of epistasis under which increased recombination is favored, under directional haploid selection, as a function of the rate of gametophytic selfing, α . Equation 27 predicts that recombination should increase in the shaded area. Circles correspond to simulation results: solid circle, a modifier increasing recombination increases in frequency when rare; open circle, it decreases in frequency. Parameter values are: (A) $s = -0.01$ and (B) $s = -0.05$, $r_{ij} = 0.2$, $r_{jk} = 0.2$. In the simulations $dr = 0.01$, $h_M = \frac{1}{2}$, and $u = 10^{-5}$.

in Table 2. To obtain simpler expressions, we considered the case where alleles 1 at loci j and k are deleterious ($s < 0$) and produced by recurrent mutations occurring at a small rate u (this small mutation rate should not affect significantly our expressions of associations at QLE, calculated in the absence of mutation); in this case, p_j and p_k remain small. By neglecting terms in p_j^2 and p_k^2 in Equation 24, one obtains that a modifier increasing recombination increases in frequency if $e_{a \times a}$ lies between

$$e_{a \times a}^{\min} = -\frac{1}{(1 + 3\alpha)} [4\alpha e_{a \times d} + \alpha e_{d \times d} + (\alpha(1 - h)^2 - h^2)s^2] - \frac{[\alpha(1 - h) + h]^2 s^2}{(1 - \alpha)(1 + 3\alpha)} \left(\frac{1}{r_{ij}} + \frac{1}{r_{ik}} \right) \quad (29)$$

$$e_{a \times a}^{\max} = -\frac{\alpha}{1 + 3\alpha} [4e_{a \times d} + e_{d \times d} + (1 - \alpha)(1 - h)^2 s^2]. \quad (30)$$

Figure 4 represents the range of epistasis values under which a modifier increasing recombination increases in frequency, for different values of α , at mutation-selection equilibrium. When α is small (Figure 4, 10^{-5}), recombination increases only when $e_{a \times a}$ is weak and negative, almost independently of $e_{a \times d}$ and $e_{d \times d}$. As α increases, mutant homozygotes become more frequent, and $e_{a \times d}$ and $e_{d \times d}$ play a more important role; however, the parameter range under which recombination is selected remains narrow. In some sense, there is still a requirement for negative epistasis, as at least one of the parameters $e_{a \times a}$, $e_{a \times d}$, and $e_{d \times d}$ must be negative for increased recombination to evolve. At high selfing rates, Equations 29 and 30 and simulations show that increased recombination is favored over a wider range of epistasis values: the space between the planes in Figure 4A increases (not shown). When $\alpha = 1$, the modifier becomes neutral again, however, as recombination has no effect in this case.

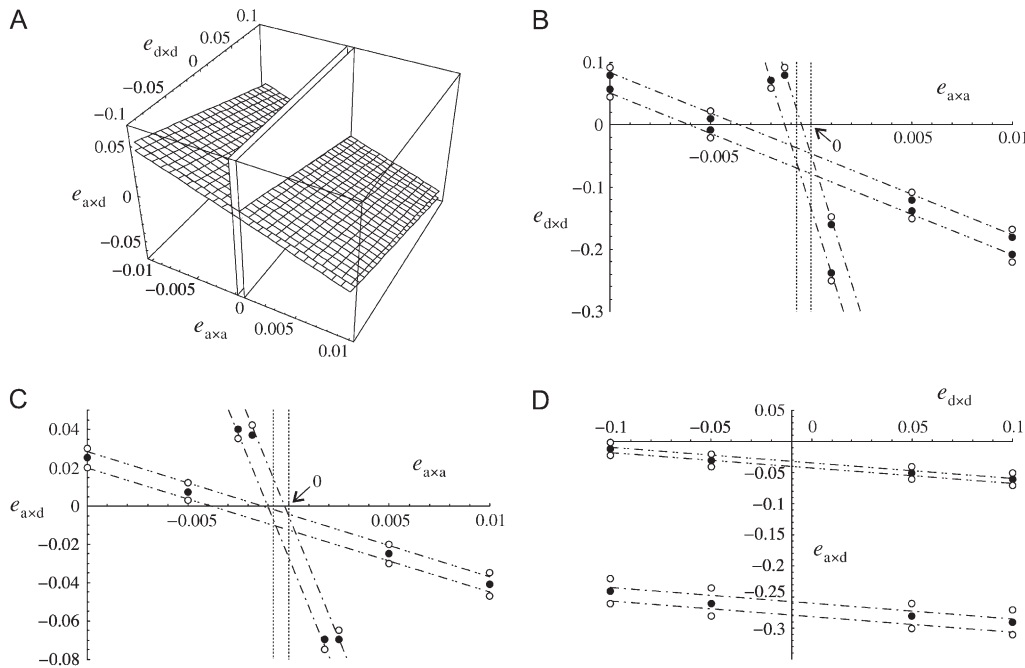


FIGURE 4.—Range of epistasis values under which a modifier increasing recombination increases in frequency, under gametophytic selfing, and under the fitness matrix given in Table 2. (A) Increased recombination is selected between the two vertical planes when $\alpha = 10^{-5}$ and between the two other planes when $\alpha = 10^{-1}$. (B–D) Regions of $e_{a \times a}$, $e_{a \times d}$, and $e_{d \times d}$, where increased recombination is selected, when $\alpha = 10^{-5}$ (dotted lines), $\alpha = 10^{-2}$ (dashed-dotted lines), and $\alpha = 10^{-1}$ (dashed-double-dotted lines). Each time, a modifier increasing recombination is predicted to increase in frequency between the two lines (obtained from Equations 29 and 30). Circles correspond to simulation results for $\alpha = 10^{-2}$ and 10^{-1} :

solid circle, the modifier increases in frequency when rare; open circle, it decreases in frequency. Parameter values are: (B) $s = -0.1$, $h = 0.1$, $r_{ij} = 0.2$, $r_{ik} = 0.2$, $e_{a \times d} = 0.01$; (C) $e_{d \times d} = 0.01$; and (D) $e_{a \times a} = 0.01$. In the simulations, $dr = 0.01$, $h_M = \frac{1}{2}$, and $u = 10^{-5}$.

TABLE 3
Expressions for the change in frequency at the recombination modifier locus at QLE,
under sporophytic selfing

	Weak selection	Weak epistasis
$\alpha = O(1)$	$\Delta p_i = a_{jk,jk} C_{ijk,jk} + o(\epsilon^2)$	$\Delta p_i = a_{jj} C_{ij,j} + a_{kk} C_{ik,k} + a_{jk,jk} C_{ijk,jk} + o(\epsilon^3)$
$\alpha = O(\epsilon)$	$\Delta p_i = a_{jk,\emptyset} C_{ijk,\emptyset} + a_{jk,jk} C_{ijk,jk} + o(\epsilon^3)$	$\Delta p_i = a_{jk,jk} C_{ijk,jk} + o(\epsilon^4)$
$\alpha = O(\epsilon^2)$	$\Delta p_i = a_{jk,\emptyset} C_{ijk,\emptyset} + o(\epsilon^3)$	$\Delta p_i = a_{j,\emptyset} C_{ij,\emptyset} + a_{k,\emptyset} C_{ik,\emptyset} + a_{jk,\emptyset} C_{ijk,\emptyset} + a_{jk,jk} C_{ijk,jk} + o(\epsilon^5)$

The modifier effect dr is assumed to be of order ϵ ; under weak selection, all $a_{U,V}$ coefficients are of order ϵ , while under weak epistasis $a_{j,\emptyset}$, $a_{k,\emptyset}$, $a_{ij,j}$, and $a_{ik,k}$ are of order ϵ and $a_{jk,\emptyset}$, $a_{j,k}$, $a_{jk,j}$, $a_{jk,k}$, and $a_{jk,jk}$ are of order ϵ^2 . The different rows correspond to different orders of magnitude of the selfing rate α .

Sporophytic selfing: Results under sporophytic selfing are qualitatively different from those under gametophytic selfing. In particular, modifiers increasing recombination can increase in frequency even when direct selection and epistasis are of the same order of magnitude (which is not the case under random mating and under gametophytic selfing). Since selection occurs only during the diploid phase, we drop D superscripts from a_{UV}^D coefficients. In the following, we again assume that epistasis is weak relative to direct selection: a_{jk} , $a_{j,k}$, $a_{jk,j}$, $a_{jk,k}$, and $a_{jk,jk}$ are of order ϵ^2 , while $a_{j,\emptyset}$, $a_{k,\emptyset}$, $a_{ij,j}$, and $a_{ik,k}$ are of order ϵ . We also give approximations for the case where all selection coefficients are of the same order. Concerning the rate of self-fertilization, we consider the cases where α is of order 1, of order ϵ , and of order ϵ^2 . Finally, we still assume that the modifier effect dr is of order ϵ . Recursions for associations that influence the change in frequency of the modifier, for the different cases that we consider, are given in APPENDIX C. Table 3 gives the leading-order terms of Δp_i at QLE, under weak selection and weak epistasis, and for the different assumptions on the order of α . In the following, we give only results for $h_M = 1/2$ (additive modifier); results are qualitatively similar under arbitrary h_M , and, in many cases, the change in frequency at the modifier locus can be obtained simply by multiplying the expressions obtained by a factor $2[h_M + p_i(1 - 2h_M)]$; we indicate when this is the case. Expressions for arbitrary dominance of the modifier effect are given in APPENDIX C.

of order 1: Under weak epistasis and strong selfing, the change in frequency of the modifier takes the form

$$\Delta p_i = a_{jj} \hat{C}_{ij,j} + a_{kk} \hat{C}_{ik,k} + a_{jk,jk} \hat{C}_{ijk,jk} + o(\epsilon^3), \quad (31)$$

where again the hat stands for genetic associations at QLE. Indeed, under sporophytic selfing, the association $C_{ijk,jk}$ is generated only by selfing and by the effect of the modifier (even when selection is absent) and is therefore of order ϵ . When $h_M = 1/2$, $\hat{C}_{ijk,jk}$ is given by

$$\hat{C}_{ijk,jk} = -\frac{dr \alpha (1 - \alpha) (1 - 2r_{jk}) [2 - \alpha (1 - r_{ij})^2 (1 - \rho_{jk})]}{[2 - \alpha (1 - \rho_{jk})]^2 [2 - \alpha (1 - \rho_{ij}) (1 - \rho_{jk})]} p q_{ijk} + o(\epsilon) \quad (32)$$

(APPENDIX C), where $\rho_{ij} = 2r_{ij}(1 - r_{ij})$, $\rho_{jk} = 2r_{jk}(1 - r_{jk})$, and again $p q_{ijk} = p_i q_j p_j q_k q_k$. Equation 32 shows that $\hat{C}_{ijk,jk}$

has the same sign as $-dr$; this reflects the fact that, under partial selfing, the correlation in homozygosity between loci j and k is lower among individuals carrying an allele that increases recombination between j and k . The associations $C_{ij,j}$ and $C_{ik,k}$ are generated by $C_{ijk,jk}$ and by selection at loci j and k and are of order ϵ^2 ; at QLE, they are given by

$$\hat{C}_{ij,j} = a_{kk} \frac{\alpha}{2 - \alpha} \hat{C}_{ijk,jk} + o(\epsilon^2), \quad \hat{C}_{ik,k} = a_{jj} \frac{\alpha}{2 - \alpha} \hat{C}_{ijk,jk} + o(\epsilon^2) \quad (33)$$

(APPENDIX C). Other associations of the form $C_{U,V}$ produce terms of higher order in the expression of Δp_i . Equations 31 and 33 lead to

$$\Delta p_i = \left(\frac{2 + \alpha}{2 - \alpha} a_{jj} a_{kk} + e_{jk,jk} \right) \hat{C}_{ijk,jk} + o(\epsilon^2), \quad (34)$$

where $e_{jk,jk}$ equals $a_{jk,jk} - a_{jj} a_{kk}$ and measures multiplicative dominance \times dominance epistasis. Because $\hat{C}_{ijk,jk}$ has the sign of $-dr$, we obtain that a modifier increasing recombination increases in frequency if

$$e_{jk,jk} < -\frac{2 + \alpha}{2 - \alpha} a_{jj} a_{kk}. \quad (35)$$

Applying the coefficients from APPENDIX A, this condition becomes, under the fitness matrix given in Table 2,

$$e_{d \times d} < -\frac{2 + \alpha}{2 - \alpha} (1 - 2h)^2 s^2. \quad (36)$$

When epistasis and direct selection are of the same order of magnitude, this condition becomes $\sim e_{d \times d} < 0$: recombination is favored when double homozygotes are less fit than expected on the basis of single homozygotes, because recombination during sporophytic selfing then produces fewer double homozygotes. This mechanism thus provides an advantage to alleles that increase recombination, in the presence of negative dominance \times dominance epistasis. Condition (36) indicates that the maximum value of $e_{d \times d}$ for recombination to be selected is in fact slightly negative; this is due to the fact that, by reducing the proportion of double homozygotes, recombination also decreases the variance in fitness, which reduces the efficiency of natural selection and

therefore gives a cost to recombination. In the absence of dominance \times dominance epistasis ($e_{d \times d} = 0$), the advantage of recombination mentioned above is absent, and this cost selects against recombination. These mechanisms by which recombination evolves under sporophytic selfing are very different from those under random mating, as is discussed later.

α of order ε : We now assume that the selfing rate is small: α is of order ε . In this case, $\hat{C}_{ijk,jk}$ becomes of order ε^2 and is given by Equation 32, expressed to the first order in α :

$$\hat{C}_{ijk,jk} = -\frac{1}{2}\alpha dr(1 - 2r_{jk})pq_{ijk} + o(\varepsilon^2). \quad (37)$$

Solving recursions for genetic associations indicates that, at QLE, $\hat{C}_{ijk,\emptyset}$ is of order ε^3 , while other associations involving a single i index are of higher order. As a consequence, the change in frequency at the modifier locus is simply given by

$$\Delta p_i = a_{jk,jk}\hat{C}_{ijk,jk} + o(\varepsilon^4). \quad (38)$$

However, if epistasis terms are stronger (*i.e.*, if all $a_{U,V}$ coefficients are of order ε), $\hat{C}_{ijk,\emptyset}$ becomes of the same order as $\hat{C}_{ijk,jk}$, while other associations involving a single i index are of higher order. In that case, we obtain

$$\Delta p_i = a_{jk,\emptyset}\hat{C}_{ijk,\emptyset} + a_{jk,jk}\hat{C}_{ijk,jk} + o(\varepsilon^3). \quad (39)$$

Solving the recursion for $C_{ijk,\emptyset}$ gives

$$\hat{C}_{ijk,\emptyset} = -\frac{dra_{jk,\emptyset}pq_{ijk}}{2r_{ijk}r_{jk}} + o(\varepsilon^2) \quad (40)$$

(APPENDIX C). Indeed, selfing affects $\hat{C}_{ijk,\emptyset}$ only through terms of order ε^3 , which are neglected here. Equations 10, 37, and 40 lead to

$$\Delta p_i = -\frac{dr}{2}\left[\frac{a_{jk,\emptyset}^2}{r_{ijk}r_{jk}} + \alpha a_{jk,jk}(1 - 2r_{jk})\right]pq_{ijk} + o(\varepsilon^3). \quad (41)$$

Therefore, under weak selection and weak selfing, we find that a modifier increasing recombination increases in frequency when dominance \times dominance epistasis is sufficiently negative,

$$\alpha a_{jk,jk} < -\frac{a_{jk,\emptyset}^2}{r_{ijk}r_{jk}(1 - 2r_{jk})}, \quad (42)$$

which, using Equations 54 and 55, gives

$$\alpha e_{d \times d} < -\frac{[e_{a \times a} + (p_j + p_k)e_{a \times d} + p_j p_k e_{d \times d}]^2}{r_{ijk}r_{jk}(1 - 2r_{jk})}. \quad (43)$$

Under arbitrary dominance at the modifier locus, Equation 41 is multiplied by a factor $2[h_M + p_i(1 - 2h_M)]$, which does not affect condition (42). When epistasis is weak ($a_{jk,jk}$ and $a_{jk,\emptyset}$ of order ε^2), this condition becomes $\sim a_{jk,jk} < 0$, in agreement with Equation 38.

α of order ε^2 : Finally, when the selfing rate is of order ε^2 , $\hat{C}_{ijk,jk}$ and $\hat{C}_{ijk,\emptyset}$ are of order ε^3 , $\hat{C}_{ij,\emptyset}$ and $\hat{C}_{ik,\emptyset}$ are of order

ε^4 , while other associations involving a single i index are of higher order, giving

$$\Delta p_i = a_{j,\emptyset}\hat{C}_{ij,\emptyset} + a_{k,\emptyset}\hat{C}_{ik,\emptyset} + a_{jk,\emptyset}\hat{C}_{ijk,\emptyset} + a_{jk,jk}\hat{C}_{ijk,jk} + o(\varepsilon^5). \quad (44)$$

At QLE, $\hat{C}_{ijk,jk}$ is still given by Equation 37, while other associations in (44) are given by

$$\hat{C}_{jk,\emptyset} = -\frac{dre_{jk,\emptyset}pq_{ijk}}{2r_{ijk}r_{jk}} + o(\varepsilon^3) \quad (45)$$

$$\hat{C}_{ij,\emptyset} = -a_{k,\emptyset}\left(1 - \frac{1}{r_{ij}}\right)\hat{C}_{ijk} + o(\varepsilon^4) \quad (46)$$

$$\hat{C}_{ik,\emptyset} = -a_{j,\emptyset}\left(1 - \frac{1}{r_{ik}}\right)\hat{C}_{ijk} + o(\varepsilon^4) \quad (47)$$

(see APPENDIX C), giving for the change in frequency of the modifier,

$$\Delta p_i = -\frac{1}{2}dr\left[\frac{e_{jk,\emptyset}}{r_{ijk}r_{jk}}\left(e_{jk,\emptyset} + a_{j,\emptyset}a_{k,\emptyset}\left(\frac{1}{r_{ij}} + \frac{1}{r_{ik}} - 1\right)\right) + \alpha a_{jk,jk}(1 - 2r_{jk})\right]pq_{ijk} + o(\varepsilon^5). \quad (48)$$

We checked this expression against numerical simulations, using the fitness matrix given by Table 2, again assuming that alleles 1 at loci j and k are deleterious ($s < 0$) and maintained at mutation-selection equilibrium, so that p_j and p_k remain small. In this case, the expressions given in APPENDIX A give: $a_{j,\emptyset}a_{k,\emptyset} \approx s^2 h^2 + o(\varepsilon^2)$, $e_{jk,\emptyset} \approx e_{a \times a} + o(\varepsilon^2)$, and $a_{jk,jk} \approx (1 - 2h)^2 s^2 + e_{d \times d} + o(\varepsilon^2)$. From this and from Equation 48, we expect that a modifier increasing recombination should increase in frequency if

$$e_{d \times d} < -(1 - 2h)^2 s^2 + \frac{e_{a \times a}}{\alpha(1 - 2r_{jk})r_{jk}r_{ik}}\left[e_{a \times a} + s^2 h^2\left(\frac{1}{r_{ij}} + \frac{1}{r_{ik}} - 1\right)\right]. \quad (49)$$

Figure 5 compares this prediction with simulation results obtained by iterating exact recursions for genotype frequencies. When α is very small (Figure 5, 10^{-5}), recombination increases only when $e_{a \times a}$ is weakly negative, almost independently of $e_{d \times d}$, as under random mating; as α increases, the range of $e_{a \times a}$ under which recombination is selected greatly increases, as long as $e_{d \times d} < 0$. As α becomes strong (Figure 5, 10^{-1}), approximation (49) still gives correct results and shows that recombination increases as long as $e_{d \times d}$ is lower than a limit value (without lower bound).

The axes of Figure 5 are not drawn to the same scale, the scale for $e_{d \times d}$ being much larger than that for $e_{a \times a}$; we used these different scales to be able to represent the narrow parameter range under which recombination is selected for under random mating or when selfing is very small ($\alpha = 10^{-5}$), while at the same time presenting simulation results for strong absolute values of $e_{d \times d}$. However, drawing the figure in this way may give a false impression of the parameter range where recombination is favored. Figure 6 presents similar graphs, but

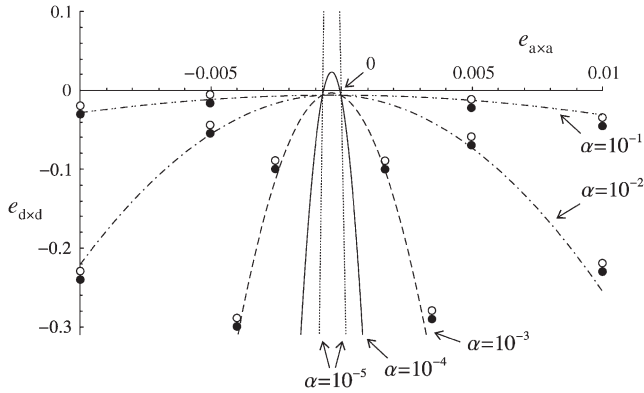


FIGURE 5.—Maximum value of dominance \times dominance epistasis ($e_{d \times d}$) for a modifier increasing recombination to be selected, under sporophytic selfing and under the fitness matrix given in Table 2, as a function of additive \times additive epistasis ($e_{a \times a}$). The curves correspond to our analytical result for weak epistasis and weak selfing (Equation 49), for $\alpha = 10^{-5}$ (dotted curves), 10^{-4} (solid curve), 10^{-3} (dashed curve), 10^{-2} (dashed-dotted curve), and 10^{-1} (dashed-double-dotted curve). Circles correspond to simulation results for $\alpha = 10^{-3}$, 10^{-2} , and 10^{-1} : solid circle, the modifier increases in frequency when rare; open circle, it decreases in frequency. Parameter values are: $r_{ij} = r_{jk} = 0.2$, $s = -0.1$, $h = 0.1$; in the simulations $dr = 0.01$, $h_M = 1/2$, $e_{a \times d} = 0.01$, $u = 10^{-5}$.

where axes are drawn to the same scale. Decreasing the strength of selection (from $s = -0.1$ in Figure 6A to $s = -0.05$ in Figure 6B) reduces the area in which recombination is selected under random mating (this area is too small to be represented in Figure 6B), while it increases the area in which recombination is selected for under selfing (the threshold value for $e_{d \times d}$ is higher in Figure 6B than in Figure 6A).

The QLE approximation does not hold under strong selfing. Comparisons between the prediction of Equation 49 and simulations are shown in Figure 7, for $s = -0.1$ and $\alpha = 0.3, 0.6, 0.9$, and 0.99 . As α gets large, the QLE approximation becomes clearly wrong: Equation 49 predicts that a modifier increasing recombination increases in frequency below the curves in Figure 7 (that is, for sufficiently negative $e_{d \times d}$), while simulations

show that, when selfing is strong, recombination can increase even when $e_{d \times d}$ is positive. Under strong selfing, selection on the recombination modifier becomes less dependent on $e_{d \times d}$ and more dependent on $e_{a \times a}$, which in most cases has to be negative (recombination increases in Figure 7, C and D, left sides).

DISCUSSION

Our model shows that self-fertilization has important effects on the evolution of recombination. Both sporophytic and gametophytic selfing generate correlations in homozygosity at different loci (associations such as $C_{ijk,jk}$). While such correlations do not depend on recombination under gametophytic selfing, they are decreasing functions of recombination under sporophytic selfing, as individuals who recombine more produce fewer double homozygotes when they self. This mechanism generates an association $C_{ijk,jk}$ between the recombination modifier (locus i) and loci j and k under sporophytic selfing, even in the absence of selection at these loci. By contrast, under random mating, all associations between the recombination modifier and other loci are built by selection acting at these loci and by the modifier effect. This association $C_{ijk,jk}$ has important qualitative and quantitative effects. For example, in the case where direct selection and epistasis are of the same order, the evolution of modifiers that increase recombination becomes possible (a necessary condition being that dominance \times dominance epistasis is negative) while it is not the case under random mating (BARTON 1995). For example, in the case of recurrent deleterious mutations, increased recombination can evolve only when additive \times additive epistasis is weakly negative when mating is random. This condition changes very fast as the rate of sporophytic selfing increases (Figure 5), selfing rates as low as 10^{-2} having substantial effects. As selfing increases, the condition for the evolution of recombination depends more and more on $e_{d \times d}$ (which has to be negative, without a lower bound) and less and less on $e_{a \times a}$ (in particular, recombination can increase when

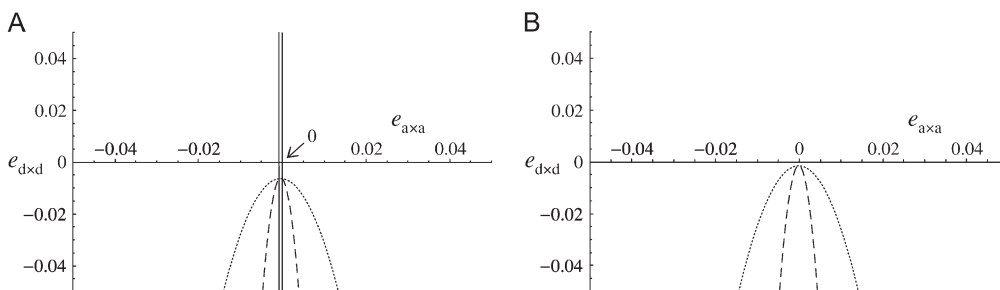


FIGURE 6.—Values of $e_{a \times a}$ and $e_{d \times d}$ for which a modifier increasing recombination increases in frequency, under sporophytic selfing and under the fitness matrix given in Table 2. Solid lines, $\alpha = 0$; dashed lines, $\alpha = 0.01$; dotted lines, $\alpha = 0.1$. The modifier increases in frequency between the solid lines in A and below the dashed and dotted lines in A and B. Parameter values are: (A) $s = -0.1$ and (B) $s = -0.05$, $h = 0.1$, $r_{ij} = 0.2$, $r_{jk} = 0.2$.

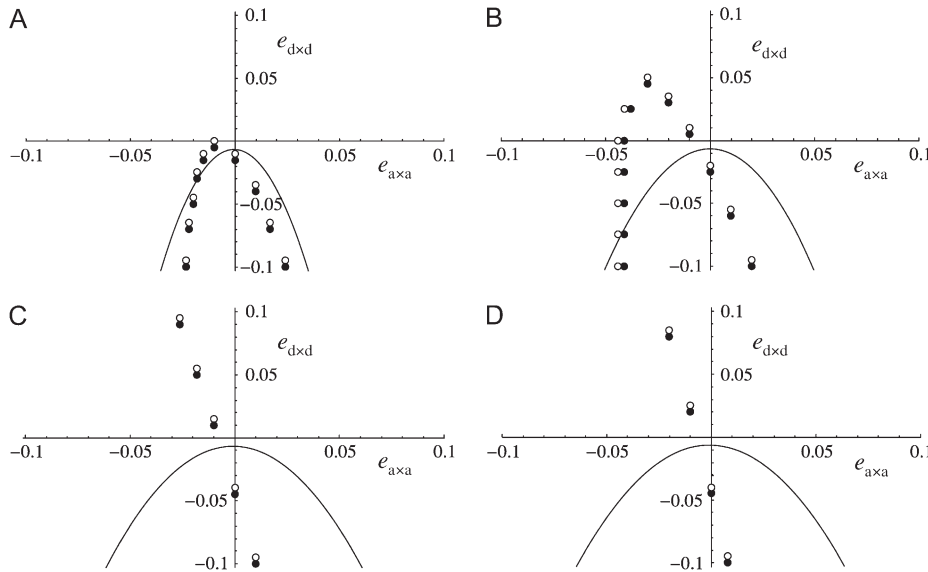


FIGURE 7.—Values of $e_{a \times a}$ and $e_{d \times d}$ for which a modifier increasing recombination increases in frequency, under stronger sporophytic selfing and under the fitness matrix given in Table 2. (A) $\alpha = 0.3$, (B) $\alpha = 0.6$, (C) $\alpha = 0.9$, and (D) $\alpha = 0.99$. Other parameters are: $s = -0.1$, $h = 0.1$, $e_{a \times d} = 0.01$, $r_{ij} = 0.2$, $r_{jk} = 0.2$, $dr = 0.01$, $h_M = 0.5$, $u = 10^{-5}$. The curves correspond to Equation 49, and the circles show simulation results; solid circle, the modifier increases in frequency when rare; open circle, it decreases in frequency.

$e_{a \times a}$ is positive, as long as $e_{d \times d}$ is sufficiently negative, as can be seen in Figure 5). Importantly, self-fertilization not only increases the parameter range under which recombination is favored, but also strongly increases the strength of selection on recombination, because when α is strong, $C_{ijh,jk}$ is of higher order than the other associations between the modifier and the selected loci.

These results are in agreement with simulation results obtained by CHARLESWORTH *et al.* (1979) and confirmed by HOLSINGER and FELDMAN (1983). Charlesworth *et al.* considered a fitness matrix where dominance \times dominance epistatic effects are measured by four coefficients k_1 , k_2 , k_3 , and k_4 (Table 2 in CHARLESWORTH *et al.* 1979). It is possible to show that, under their fitness matrix, the coefficient $a_{jkh,jk}$ equals $-k_1 - k_2 - k_3 - k_4$, to the first order in selective differences. Charlesworth *et al.* found that, when all k_i are equal (and equal to k), increased recombination is selected when k is positive, provided that self-fertilization occurs (even at a small rate); under random mating, recombination does not increase. This result agrees with our Equations 36 and 42.

The mechanisms by which recombination evolves under sporophytic selfing (when $e_{d \times d} < 0$) and under random mating are also qualitatively different. Under random mating, and when epistasis is negative, recombination decreases the average fitness of offspring, while it increases their variance in fitness. This increase in variance drives the evolution of the modifier, provided that the initial decrease in fitness is not too high (which can be the case only when epistasis is weak). Recombination has the opposite effect under sporophytic selfing when $e_{d \times d} < 0$. Indeed, because sporophytic selfing generates an excess of double homozygotes ($C_{jkh,jk} > 0$), it increases the variance in fitness. Recombination, however, reduces this excess of double homozygotes ($C_{jkh,jk}$ decreases as r_{jk} increases, as can be seen from Equation

C3) and thereby reduces the variance in fitness. However, when $e_{d \times d} < 0$, individuals who recombine more produce offspring that are more fit on average (because they produce fewer double homozygotes through selfing). As a result, conditions for higher rates of recombination to evolve are reversed: while recombination increases when epistasis is weakly negative under random mating, it increases under selfing when $e_{d \times d}$ is more negative than a limit value (Equation 35).

Under gametophytic selfing, correlations in homozygosity at different loci do not depend on recombination, because selfing occurs among identical gametes (produced by a haploid gametophyte). Under this mating system, the association $C_{ijh,jk}$ is generated by selection at loci j and k and by the modifier effect, and results are more similar to those obtained under random mating (this can be seen from Equation 17, which gives the change in frequency of the modifier under gametophytic selfing, and which takes the same form as under random mating). Gametophytic selfing has, however, two important effects. First, it decreases the effectiveness of recombination, thereby increasing hitchhiking effects. As a consequence, the range of epistasis under which increased recombination can evolve is increased—this effect, however, is important mostly at high selfing rates (Figure 3). Second, by increasing homozygosity, gametophytic selfing increases the effects of additive \times dominance and of dominance \times dominance epistasis under diploid selection (Figure 4). In the case of deleterious mutations, the combination of $e_{a \times a}$, $e_{a \times d}$, and $e_{d \times d}$ determines when modifiers that increase recombination can evolve (rather than just $e_{a \times a}$), and, in some cases, such modifiers evolve when $e_{a \times a}$ is positive.

We expect that results for other sources of inbreeding, such as sib mating or population structure, should be qualitatively similar to those obtained under sporophy-

tic selfing; indeed, inbreeding (and population structure) generates correlations in homozygosity, which in general decrease as recombination rates increase (gametophytic selfing being here an exception). Therefore, the result that negative dominance \times dominance epistasis selects for increased recombination would probably be obtained from other models of inbreeding. These different cases remain to be investigated using specific models.

Our results stress the fact that the different components of epistasis play different roles in the evolution of recombination under inbreeding: for example, under sporophytic selfing, increased recombination can evolve when $e_{a \times a}$ is positive, as long as $e_{d \times d}$ is sufficiently negative. Therefore, it appears important to measure these different components separately (not only $e_{a \times a}$) to assess the deterministic theories for the evolution of recombination. At present, most experimental studies measuring epistasis between mutations affecting fitness have been performed on haploid organisms (DE VISSER *et al.* 1996, 1997a,b; ELENA and LENSKI 1997; ELENA 1999; DE LA PEÑA *et al.* 2000; WLOCH *et al.* 2001) or on diploids made homozygous at all loci (DE VISSER and HOEKSTRA 1998; PETERS and KEIGHTLEY 2000; WHITLOCK and BOURGUET 2000), in which case the different components of epistasis cannot be separated. However, some studies on diploid plants have tested for dominance \times dominance epistasis (WILLIS 1993; FALCONER and MACKAY 1996 and references therein); in these studies, different components of fitness are measured over successive generations of inbreeding. In the absence of dominance \times dominance epistasis, the decline in fitness with inbreeding is expected to be linear, while negative $e_{d \times d}$ leads to a greater than linear decrease (CROW and KIMURA 1970, p. 80). Some of these studies found linear decreases, while others obtained accelerating decreases. However, this method is biased against finding negative dominance \times dominance epistasis, because the worst genotypes can be lost by selection during the course of the experiment; therefore, linear decreases in fitness with inbreeding do not prove that dominance \times dominance epistasis is absent.

In a recent article, OTTO (2003) studied the hypothesis that sex might have evolved to promote segregation. She found that segregation evolves only under a restricted range of parameters with random mating, but under a much wider range in the presence of inbreeding (represented in her model by gametophytic selfing). With inbreeding, selection on a modifier of segregation is also much stronger, and she concluded that sexual reproduction is more likely to have been driven by the benefits of segregation than by the benefits of recombination. However, our model demonstrates that under sporophytic selfing, selection on a recombination modifier is also much stronger than under random mating (when there is dominance \times dominance epistasis); therefore, which of the two effects of sex (segregation

or recombination) brings more benefits is not so clear; models including both effects are needed to answer this question, in parallel with measures of dominance \times dominance epistasis.

We thank Sally Otto, Sergey Gavrilets, François Rousset, Mike Whitlock, and an anonymous reviewer for helpful comments and discussions on previous versions of this manuscript. This work was supported by grant Action Concertée Incitative Jeunes Chercheurs no. 0693 from the French Ministry of Research.

LITERATURE CITED

- AGRAWAL, A., and J. CHASNOV, 2001 Recessive mutations and the maintenance of sex in structured populations. *Genetics* **158**: 913–917.
- ARROYO, M. T. K., 1973 Chiasma frequency evidence on the evolution of autogamy in *Limnanthes floccosa* (Limnanthaceae). *Evolution* **27**: 679–688.
- BARTON, N. H., 1995 A general model for the evolution of recombination. *Genet. Res.* **65**: 123–144.
- BARTON, N. H., and M. TURELLI, 1991 Natural and sexual selection on many loci. *Genetics* **127**: 229–255.
- CHARLESWORTH, B., 1976 Recombination modification in a fluctuating environment. *Genetics* **83**: 181–195.
- CHARLESWORTH, B., 1980 The cost of sex in relation to the mating system. *J. Theor. Biol.* **84**: 655–671.
- CHARLESWORTH, D., B. CHARLESWORTH and C. STROBECK, 1977 Effects of selfing on selection for recombination. *Genetics* **86**: 213–226.
- CHARLESWORTH, D., B. CHARLESWORTH and C. STROBECK, 1979 Selection for recombination in partially self-fertilizing populations. *Genetics* **93**: 237–244.
- CROW, J. F., and M. KIMURA, 1970 *An Introduction to Population Genetics Theory*. Harper & Row, New York.
- DE LA PEÑA, M., S. F. ELENA and A. MOYA, 2000 Effect of deleterious mutation-accumulation on the fitness of RNA bacteriophage MS2. *Evolution* **54**: 686–691.
- DE VISSER, J. A. G. M., and R. F. HOEKSTRA, 1998 Synergistic epistasis between loci affecting fitness: evidence in plants and fungi. *Genet. Res.* **71**: 39–49.
- DE VISSER, J. A. G. M., R. F. HOEKSTRA and H. VAN DEN ENDE, 1996 The effect of sex and deleterious mutations on fitness in *Chlamydomonas*. *Proc. R. Soc. Lond. Ser. B* **263**: 193–200.
- DE VISSER, J. A. G. M., R. F. HOEKSTRA and H. VAN DEN ENDE, 1997a An experimental test for synergistic epistasis and its application in *Chlamydomonas*. *Genetics* **145**: 815–819.
- DE VISSER, J. A. G. M., R. F. HOEKSTRA and H. VAN DEN ENDE, 1997b Test of interaction between genetic markers that affect fitness in *Aspergillus niger*. *Evolution* **51**: 1499–1505.
- ELENA, S. F., 1999 Little evidence for synergism among deleterious mutations in a nonsegmented RNA virus. *J. Mol. Evol.* **49**: 703–707.
- ELENA, S. F., and R. E. LENSKI, 1997 Test of synergistic interactions among deleterious mutations in bacteria. *Nature* **390**: 395–397.
- FALCONER, D. S., and T. F. C. MACKAY, 1996 *Introduction to Quantitative Genetics*. Addison Wesley Longman, Harlow, UK.
- FELDMAN, M. W., F. B. CHRISTIANSEN and L. D. BROOKS, 1980 Evolution of recombination in a constant environment. *Proc. Natl. Acad. Sci. USA* **77**: 4838–4841.
- FELSENSTEIN, J., 1974 The evolutionary advantage of recombination. *Genetics* **78**: 737–756.
- FISHER, R. A., 1930 *The Genetical Theory of Natural Selection*. Oxford University Press, Oxford.
- GIBBS, P. E., C. MILNE and M. VARGAS CARILLO, 1975 Correlation between the breeding system and recombination index in five species of *Senecio*. *New Phytol.* **75**: 619–626.
- GRANT, V., 1958 The regulation of recombination in plants. *Cold Spring Harbor Symp. Quant. Biol.* **23**: 337–363.
- HOLSINGER, K. E., and M. W. FELDMAN, 1983 Linkage modification with mixed random mating and selfing: a numerical study. *Genetics* **103**: 323–333.

- KIRKPATRICK, M., T. JOHNSON and N. H. BARTON, 2002 General models of multilocus evolution. *Genetics* **161**: 1727–1750.
- KONDRASHOV, A. S., 1982 Selection against harmful mutations in large sexual and asexual populations. *Genet. Res.* **40**: 325–332.
- KONDRASHOV, A. S., 1988 Deleterious mutations and the evolution of sexual reproduction. *Nature* **336**: 435–440.
- KONDRASHOV, A. S., 1993 Classification of hypotheses on the advantage of amphimixis. *J. Hered.* **84**: 372–387.
- LENORMAND, T., 2003 The evolution of sex dimorphism in recombination. *Genetics* **163**: 811–822.
- LENORMAND, T., and J. DUTEIL, 2005 Recombination difference between sexes: a role for haploid selection. *PLoS Biol.* **3** (3): e63.
- LENORMAND, T., and S. P. OTTO, 2000 The evolution of recombination in a heterogeneous environment. *Genetics* **156**: 423–438.
- MAYNARD SMITH, J., 1971 What use is sex? *J. Theor. Biol.* **30**: 319–335.
- MCCAULEY, D. E., D. P. WHITTIER and L. M. REILLY, 1985 Inbreeding and the rate of self-fertilization in a grape fern, *Botrychium dissectum*. *Am. J. Bot.* **72**: 1978–1981.
- MULLER, H. J., 1932 Some genetic aspects of sex. *Am. Nat.* **66**: 118–138.
- NAGYLAZI, T., 1993 The evolution of multilocus systems under weak selection. *Genetics* **134**: 627–647.
- OTTO, S. P., 2003 The advantages of segregation and the evolution of sex. *Genetics* **164**: 1099–1118.
- OTTO, S. P., and N. H. BARTON, 1997 The evolution of recombination: removing the limits to natural selection. *Genetics* **147**: 879–906.
- OTTO, S. P., and N. H. BARTON, 2001 Selection for recombination in small populations. *Evolution* **55**: 1921–1931.
- OTTO, S. P., and M. W. FELDMAN, 1997 Deleterious mutations, variable epistatic interactions, and the evolution of recombination. *Theor. Popul. Biol.* **51**: 134–147.
- OTTO, S. P., and T. LENORMAND, 2002 Resolving the paradox of sex and recombination. *Nat. Rev. Genet.* **3**: 252–261.
- PETERS, A. D., and P. D. KEIGHTLEY, 2000 A test for epistasis among induced mutations in *Caenorhabditis elegans*. *Genetics* **156**: 1635–1647.
- SHARMA, N., P. KOUL and A. K. KOUL, 1992 Genetic systems of six species of *Plantago* (*Plantaginaceae*). *Plant Syst. Evol.* **181**: 1–9.
- SOLTIS, D. E., and P. S. SOLTIS, 1986 Electrophoretic evidence for inbreeding in the fern *Botrychium virginianum* (*Ophioglossaceae*). *Am. J. Bot.* **73**: 588–592.
- STEBBINS, G. L., 1958 Longevity, habitat and release of genetic variability in the higher plants. *Cold Spring Harbor Symp. Quant. Biol.* **23**: 365–377.
- UYENOYAMA, M. K., and B. BENGTSSON, 1989 On the origin of meiotic reproduction: a genetic modifier model. *Genetics* **123**: 873–885.
- VED BRAT, S., 1965 Genetic systems in *Allium*. III. Meiosis and breeding systems. *Heredity* **20**: 325–339.
- WHITLOCK, M. C., and D. BOURGUET, 2000 Factors affecting the genetic load in *Drosophila*: synergistic epistasis and correlations among fitness components. *Evolution* **54**: 1654–1660.
- WILLIAMS, G. C., and J. B. MITTON, 1973 Why reproduce sexually? *J. Theor. Biol.* **39**: 545–554.
- WILLIS, J. H., 1993 Effects of different levels of inbreeding on fitness components in *Mimulus guttatus*. *Evolution* **47**: 864–876.
- WLOCH, D. M., R. H. BORTS and R. KORONA, 2001 Epistatic interactions of spontaneous mutations in haploid strains of the yeast *Saccharomyces cerevisiae*. *J. Evol. Biol.* **14**: 310–316.
- ZARCHI, Y., G. SIMCHEN, J. HILLEL and T. SCHAAP, 1972 Chiasmata and the breeding system in wild populations of diploid wheats. *Chromosome* **38**: 77–94.

Communicating editor: M. K. UYENOYAMA

APPENDIX A

Under the fitness matrix given by Table 2, to the first order in s , $e_{a \times a}$, $e_{a \times d}$, and $e_{d \times d}$ (weak selection), $a_{U,V}^D$ coefficients are given by

$$a_{j\emptyset}^D = hs + (1 - 2h)sp_j + p_k[2e_{a \times a} + (2p_j + p_k)e_{a \times d} + p_jp_ke_{d \times d}] \quad (A1)$$

$$a_{k\emptyset}^D = hs + (1 - 2h)sp_k + p_j[2e_{a \times a} + (2p_k + p_j)e_{a \times d} + p_jp_ke_{d \times d}] \quad (A2)$$

$$a_{jj}^D = (1 - 2h)s + p_k(2e_{a \times d} + p_ke_{d \times d}) \quad (A3)$$

$$a_{kk}^D = (1 - 2h)s + p_j(2e_{a \times d} + p_je_{d \times d}) \quad (A4)$$

$$a_{jk\emptyset}^D = a_{j,k}^D = e_{a \times a} + (p_j + p_k)e_{a \times d} + p_jp_ke_{d \times d} \quad (A5)$$

$$a_{j\emptyset}^D = e_{a \times d} + p_je_{d \times d}, \quad a_{jk,k}^D = e_{a \times d} + p_je_{d \times d}, \quad a_{jk,jk}^D = e_{d \times d}. \quad (A6)$$

To the second order in s , and to the first order in $e_{a \times a}$, $e_{a \times d}$, and $e_{d \times d}$ (weak epistasis), $a_{U,V}^D$ coefficients are given by

$$a_{j\emptyset}^D = s[h + p_j(1 - 2h)](1 - sT_j) + p_k[2e_{a \times a} + (p_j + p_k)e_{a \times d} + p_jp_ke_{d \times d}] \quad (A7)$$

$$a_{k\emptyset}^D = s[h + p_k(1 - 2h)](1 - sT_k) + p_j[2e_{a \times a} + (p_j + p_k)e_{a \times d} + p_jp_ke_{d \times d}] \quad (A8)$$

$$a_{jj}^D = (1 - 2h)s(1 - sT_j) + p_k(2e_{a \times d} + p_ke_{d \times d}) \quad (A9)$$

$$a_{kk}^D = (1 - 2h)s(1 - sT_k) + p_j(2e_{a \times d} + p_je_{d \times d}) \quad (A10)$$

$$a_{jk\emptyset}^D = a_{j,k}^D = e_{a \times a} + (p_j + p_k)e_{a \times d} + p_jp_ke_{d \times d} + s^2[h + p_j(1 - 2h)][h + p_k(1 - 2h)] \quad (A11)$$

$$a_{j\emptyset}^D = e_{a \times d} + p_je_{d \times d} + (1 - 2h)s^2[h + p_k(1 - 2h)] \quad (A12)$$

$$a_{jk,k}^D = e_{a \times d} + p_je_{d \times d} + (1 - 2h)s^2[h + p_j(1 - 2h)] \quad (A13)$$

$$a_{jk,jk}^D = e_{d \times d} + (1 - 2h)s^2 \quad (A14)$$

with

$$T_j = p_j[2h + p_j(1 - 2h)] + (1 - 2h)F(p_jq_j + p_kq_k) \quad (A15)$$

$$T_k = p_k[2h + p_k(1 - 2h)] + (1 - 2h)F(p_jq_j + p_kq_k) \quad (A16)$$

and F is the neutral inbreeding coefficient, equal to $\alpha/(2 - \alpha)$ under sporophytic selfing and α under gametophytic selfing.

APPENDIX B

Here we give recursions for gametophytic selfing under weak epistasis: a_j^H , a_k^H , $a_{j\emptyset}^D$, $a_{k\emptyset}^D$, a_{jj}^D , and a_{kk}^D are of order ϵ , while a_{jk}^H , $a_{jk\emptyset}^D$, $a_{j,k}^D$, $a_{j,k}^D$, $a_{j,k}^D$, and $a_{j,k}^D$ are of order ϵ^2 .

Haploid selection: The change in allele frequency at the modifier locus (locus i) due to selection during the haploid phase is given by

$$\Delta_H p_i = a_j^H C_{ij} + a_k^H C_{ik} + a_{jk}^H C_{ijk} \quad (\text{B1})$$

(from KIRKPATRICK *et al.* 2002, Equation 10). The recursions for linkage disequilibria to leading orders in ϵ are given by

$$C_{ij}^H = C_{ij} + a_k^H C_{ijk} + o(\epsilon^4) \quad (\text{B2})$$

$$C_{ik}^H = C_{ik} + a_j^H C_{ijk} + o(\epsilon^4) \quad (\text{B3})$$

$$C_{jk}^H = C_{jk} + e_{jk}^H p_j q_j p_k q_k + o(\epsilon^2) \quad (\text{B4})$$

$$C_{ijk}^H = C_{ijk} + o(\epsilon^3) \quad (\text{B5})$$

(from KIRKPATRICK *et al.* 2002, Equations 9 and 15).

Diploid selection: The change in allele frequency at the modifier locus during diploid selection is given by

$$\Delta_D p_i = \sum_{U,V} a_{U,V}^D C_{U_i,V}^S, \quad (\text{B6})$$

which, using Equations 7 and 8, gives

$$\Delta_D p_i = [a_{j\emptyset}^D + \alpha \Lambda_j] C_{ij}^H + [a_{k\emptyset}^D + \alpha \Lambda_k] C_{ik}^H + [E^D + (a_{j\emptyset}^D + \alpha \Lambda_j)(a_{k\emptyset}^D + \alpha \Lambda_k)] C_{ijk}^H + o(\epsilon^5) \quad (\text{B7})$$

with

$$\Lambda_j = a_{j\emptyset}^D + a_{jj}^D(1 - 2p_j), \quad \Lambda_k = a_{k\emptyset}^D + a_{kk}^D(1 - 2p_k) \quad (\text{B8})$$

and

$$E^D = \alpha(1 - \alpha)\Lambda_j\Lambda_k + (1 + \alpha)e_{jk\emptyset}^D + 2\alpha e_{j,k}^D + 2\alpha e_{j,k}^D(1 - 2p_j) + 2\alpha e_{j,k}^D(1 - 2p_k) + \alpha e_{j,k}^D(1 - 2p_j)(1 - 2p_k). \quad (\text{B9})$$

Genetic associations after diploid selection are given by

$$C_{ij\emptyset}^D = C_{ij}^H + (a_{k\emptyset}^D + \alpha \Lambda_k) C_{ijk}^H + o(\epsilon^4) \quad (\text{B10})$$

$$C_{ij}^D = \alpha C_{ij}^H + \alpha(a_{k\emptyset}^D + \Lambda_k) C_{ijk}^H + o(\epsilon^4) \quad (\text{B11})$$

$$C_{ik\emptyset}^D = C_{ik}^H + (a_{j\emptyset}^D + \alpha \Lambda_j) C_{ijk}^H + o(\epsilon^4) \quad (\text{B12})$$

$$C_{i,k}^D = \alpha C_{ik}^H + \alpha(a_{j\emptyset}^D + \Lambda_j) C_{ijk}^H + o(\epsilon^4) \quad (\text{B13})$$

$$C_{jk\emptyset}^D = C_{jk}^H + E^D p_j q_j p_k q_k + o(\epsilon^2) \quad (\text{B14})$$

$$C_{j,k}^D = \alpha C_{jk}^H + [E^D - (1 - \alpha)(e_{jk\emptyset}^D - e_{j,k}^D)] p_j q_j p_k q_k + o(\epsilon^2). \quad (\text{B15})$$

Finally, we have

$$C_{ijk\emptyset}^D = C_{ijk}^H + o(\epsilon^3), \quad C_{ij,k}^D = C_{ik,j}^D = C_{j,k,i}^D = \alpha C_{ijk}^H + o(\epsilon^3). \quad (\text{B16})$$

Recombination: Recombination does not affect allele frequencies. Recursions for genetic associations in the presence of the recombination modifier are the same as in BARTON (1995),

$$C'_{ij} = (1 - r_{ij}) C_{ij\emptyset}^D + r_{ij} C_{ij}^D + o(\epsilon^4) \quad (\text{B17})$$

$$C'_{ik} = (1 - r_{ik}) C_{ik\emptyset}^D + r_{ik} C_{i,k}^D + o(\epsilon^4) \quad (\text{B18})$$

$$C'_{jk} = (1 - r_{jk}) C_{jk\emptyset}^D + r_{jk} C_{j,k}^D + o(\epsilon^2) \quad (\text{B19})$$

$$C'_{ijk} = [\alpha + (1 - \alpha)(1 - r_{ijk})] C_{ijk}^H - dr[h_M + p_i(1 - 2h_M)] p_j q_j (C_{jk\emptyset}^D - C_{j,k}^D) + o(\epsilon^3), \quad (\text{B20})$$

where, r_{ijk} is the probability that at least one crossing over occurs between loci i , j , and k , in the absence of the modifier.

APPENDIX C

In the absence of selection, selfing increases homozygosity and generates correlations in homozygosity at different loci, leading to positive associations between identical sets of loci present on homologous chromosomes. Equilibrium values for these associations under sporophytic selfing are obtained by solving recursions derived from Equations 13 and 14, with $C_{XY}^D = C_{XY}$, and all $\delta t_{s,T}$ (which represent the modifier effect) equal to zero. One obtains for the association between homologous genes at a single locus,

$$\hat{C}_{i,i} = Fp_iq_i, \quad (\text{C1})$$

where $F = \alpha/(2 - \alpha)$ measures the deviation from Hardy-Weinberg equilibrium at a single locus, due to selfing. Associations between pairs of homologous genes at two loci are given by

$$\hat{C}_{ij,ij} = G_{ij}p_iq_i p_jq_j, \quad (\text{C2})$$

where

$$G_{ij} = \frac{\alpha}{2 - \alpha} \frac{2 - \alpha - 2(2 - 3\alpha)r_{ij}(1 - r_{ij})}{2 - \alpha[1 - 2r_{ij}(1 - r_{ij})]}, \quad (\text{C3})$$

measuring the correlation between homozygosity at two linked loci (again due to selfing). Finally, one has

$$\hat{C}_{ijk,ijk} = H_{ijk}p_iq_i p_jq_j p_kq_k \quad (\text{C4})$$

with

$$H_{ijk} = \frac{\alpha X}{2 - \alpha(1 - \rho_{ij})(1 - \rho_{jk})}, \quad (\text{C5})$$

where

$$X = F + (1 - \rho_{ij})(1 - \rho_{jk})(1 - F) + (1 - \rho_{ij})\rho_{jk}G_{ij} + \rho_{ij}(1 - \rho_{jk})G_{jk} + \rho_{ij}\rho_{jk}G_{ik} \quad (\text{C6})$$

and $\rho_{ij} = 2r_{ij}(1 - r_{ij})$ and $\rho_{jk} = 2r_{jk}(1 - r_{jk})$.

α of order 1: When α is of order 1, $\hat{C}_{ijk,ijk}$ is of order ε ; indeed, $C_{ijk,ijk}$ is generated only by selfing and by the modifier effect. To linear order in dr , one obtains from Equations 13 and 14

$$\begin{aligned} C'_{ijk,ijk} &= \frac{\alpha}{2} [1 - 2r_{jk}(1 - r_{jk})] C_{ijk,ijk} \\ &\quad - \alpha dr [F - G_{ij} - G_{ik} + H_{ijk} + (h_M + p_i(1 - 2h_M))(1 - 3F + G_{ij} + G_{ik} + G_{jk} - H_{ijk})] (1 - 2r_{jk}) p q_{ijk} \end{aligned} \quad (\text{C7})$$

with $p q_{ijk} = p_i q_i p_j q_j p_k q_k$. Solving Equation C7 for $C'_{ijk,ijk} = C_{ijk,ijk}$ gives

$$\hat{C}_{ijk,ijk} = -\frac{4\alpha(1 - \alpha) dr [(h_M + p_i(1 - 2h_M))X_1 + X_2] (1 - 2r_{jk}) p q_{ijk}}{2 - \alpha[1 - 2r_{jk}(1 - r_{jk})]} + o(dr) \quad (\text{C8})$$

with

$$X_1 = \frac{1}{2 - \alpha(1 - \rho_{ij})(1 - \rho_{jk})} \quad (\text{C9})$$

$$X_2 = \frac{\alpha \rho_{ij}(1 - \rho_{jk})}{4 - \alpha(1 - \rho_{jk})[2 + (1 - \rho_{ij})(2 - \alpha(1 - \rho_{jk}))]} \quad (\text{C10})$$

Except dr , all terms in the fraction of Equation C8 are positive when r_{ij} and r_{jk} are in the range $[0, \frac{1}{2}]$, and α , h_M , and p_i are in $[0, 1]$; therefore, $\hat{C}_{ijk,ijk}$ has the same sign as $-dr$.

$\hat{C}_{ij,j}$ and $\hat{C}_{ik,k}$ are of order ε^2 and are obtained by solving the recursions

$$C'_{ij,j} = \frac{\alpha}{2} (C_{ij,j} + a_{k,k} C_{ijk,ijk}) + o(\varepsilon^2) \quad (\text{C11})$$

$$C'_{ik,k} = \frac{\alpha}{2} (C_{ik,k} + a_{j,j} C_{ijk,ijk}) + o(\varepsilon^2), \quad (\text{C12})$$

giving at QLE,

$$\hat{C}_{ij,j} = a_{i,k} \frac{\alpha}{2 - \alpha} \hat{C}_{ijk,jk} + o(\varepsilon^2), \quad \hat{C}_{ik,k} = a_{j,j} \frac{\alpha}{2 - \alpha} \hat{C}_{ijk,jk} + o(\varepsilon^2). \quad (\text{C13})$$

Under weak epistasis, $\hat{C}_{ijk,j}$, $\hat{C}_{ij,jk}$, $\hat{C}_{ijk,k}$, and $\hat{C}_{ij,ik}$ are also of order ε^2 , while other associations involving a single i index are of higher order.

α of order ε : When α is of order ε , $C_{ijk,jk}$ becomes of order ε^2 and is given by

$$\hat{C}_{ijk,jk} = -\alpha dr[h_M + p_i(1 - 2h_M)](1 - 2r_{jk})pq_{jk} + o(\varepsilon^2) \quad (\text{C14})$$

(from Equation C8). Under weak epistasis, $C_{ijk,\emptyset}$ is of order 3 at QLE, while other associations involving a single i index are of higher order. Under weak selection (when all $a_{U,V}$ coefficients are of order ε), $C_{ijk,\emptyset}$ is of order ε^2 and is obtained by solving the recursions

$$C'_{jk,\emptyset} = (1 - r_{jk})C_{jk,\emptyset} + [a_{jk,\emptyset} + (a_{j,k} - a_{jk,\emptyset})r_{jk}]pq_{jk} + o(\varepsilon) \quad (\text{C15})$$

$$C'_{ijk,\emptyset} = (1 - r_{ij})(1 - r_{jk})C_{ijk,\emptyset} - dr[h_M + p_i(1 - 2h_M)][C_{jk,\emptyset} + (a_{jk,\emptyset} - a_{j,k})pq_{jk}]p_iq_i + o(\varepsilon^2), \quad (\text{C16})$$

giving at QLE,

$$\hat{C}_{jk,\emptyset} = \frac{a_{jk,\emptyset} + (a_{j,k} - a_{jk,\emptyset})r_{jk}}{r_{jk}}pq_{jk} + o(\varepsilon) \quad (\text{C17})$$

$$\hat{C}_{ijk,\emptyset} = -\frac{dr[h_M + p_i(1 - 2h_M)]a_{jk,\emptyset}pq_{jk}}{r_{ijk}r_{jk}} + o(\varepsilon^2) \quad (\text{C18})$$

with $r_{ijk} = 1 - (1 - r_{ij})(1 - r_{jk})$.

α of order ε^2 : When α is of order ε^2 , $\hat{C}_{ijk,jk}$ is of order ε^3 and is given by Equation C14. Under weak epistasis, $\hat{C}_{ijk,\emptyset}$ is of order ε^3 and is obtained by solving

$$C'_{jk,\emptyset} = (1 - r_{jk})C_{jk,\emptyset} + [e_{jk,\emptyset} + (a_{j,k} - a_{jk,\emptyset})r_{jk}]pq_{jk} + o(\varepsilon^2) \quad (\text{C19})$$

$$C'_{ijk,\emptyset} = (1 - r_{ij})(1 - r_{jk})C_{ijk,\emptyset} - dr[h_M + p_i(1 - 2h_M)][C_{jk,\emptyset} + (a_{jk,\emptyset} - a_{j,k})pq_{jk}]p_iq_i + o(\varepsilon^3), \quad (\text{C20})$$

giving at QLE,

$$\hat{C}_{ijk,\emptyset} = -\frac{dr[h_M + p_i(1 - 2h_M)]e_{jk,\emptyset}pq_{jk}}{r_{ijk}r_{jk}} + o(\varepsilon^3). \quad (\text{C21})$$

Finally, $\hat{C}_{ij,\emptyset}$ and $\hat{C}_{ik,\emptyset}$ are of order ε^4 and are obtained by solving

$$C_{ij,\emptyset} = (1 - r_{ij})(C_{ij,\emptyset} + a_{k,\emptyset}C_{ijk,\emptyset}) + o(\varepsilon^4) \quad (\text{C22})$$

$$C_{ik,\emptyset} = (1 - r_{ik})(C_{ik,\emptyset} + a_{j,\emptyset}C_{ijk,\emptyset}) + o(\varepsilon^4). \quad (\text{C23})$$

Other associations involving a single i index are of higher order in ε .

