General Models of Multilocus Evolution

Mark Kirkpatrick,*,[1] Toby Johnson[†,‡] and Nick Barton[†]

*Section of Integrative Biology, University of Texas, Austin, Texas 78712, †Institute of Cell, Animal and Population Biology, University of Edinburgh, Scotland EH9 3JT, United Kingdom and ‡Department of Zoology, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada

ABSTRACT

In 1991, Barton and Turelli developed recursions to describe the evolution of multilocus systems under arbitrary forms of selection. This article generalizes their approach to allow for arbitrary modes of inheritance, including diploidy, polyploidy, sex linkage, cytoplasmic inheritance, and genomic imprinting. The framework is also extended to allow for other deterministic evolutionary forces, including migration and mutation. Exact recursions that fully describe the state of the population are presented; these are implemented in a computer algebra package (available on the Web at http://helios.bto.ed.ac.uk/evolgen). Despite the generality of our framework, it can describe evolutionary dynamics exactly by just two equations. These recursions can be further simplified using a "quasi-linkage equilibrium" (QLE) approximation. We illustrate the methods by finding the effect of natural selection, sexual selection, mutation, and migration on the genetic composition of a population.

EVOLUTION involves simultaneous changes at many genetic loci. Modeling these changes is difficult because associations between alleles at different loci ("linkage disequilibria") cause the effects of selection acting on one locus to spill over onto other loci, generating indirect selection (see EWENS 1979, p. 195). These indirect effects influence the evolution of genes that are themselves under direct selection, altering the course of adaptation (*e.g.*, HILL and ROBERTSON 1966; BARTON 1983). Moreover, indirect selection determines the fate of modifier genes that have important effects even if they themselves are free of direct selection. Examples of such modifiers include female mating preference genes (FISHER 1952), modifiers of recombination (OTTO and MICHALAKIS 1998), and modifiers of the mutation rate (DAWSON 1999; SNIEGOWSKI *et al.* 2000).

The most obvious approach to modeling multilocus systems is simply to follow the frequencies of all possible genotypes. There are three basic drawbacks here. First, the number of genotypes grows exponentially with the number of loci, rapidly overwhelming both analytical and simulation approaches when there are even a modest number of genes. Second, the quantities that are often of most interest, such as allele frequencies and mean phenotypes, are obscured by working with genotypes. Third, approximations for the dynamic equations appear more naturally when we work with quantities other than genotype frequencies.

Several approximate approaches have been developed to deal with these problems.

1. The infinitesimal model assumes that very many genes influence the phenotype, such that each allele has an infinitesimal effect (FISHER 1918; BULMER 1980; TURELLI and BARTON 1994). In this limit, the genetic variance contributed by allelic variation at each locus is constant, and evolutionary change is due solely to changes in associations among loci. This is an accurate and general approximation for short-term change under strong selection, but cannot describe changes in allele frequencies over the longer term.

2. The hypergeometric model (KONDRASHOV 1984; BARTON 1992; DOEBELI 1996) also assumes that loci are unlinked and have equal effects, but allows the number of loci to be finite. However, the stability of solutions to this model is limited to certain selection regimes (SHPAK and KONDRASHOV 1999; BARTON and SHPAK 2000), which do not include many scenarios of evolutionary interest, such as stabilizing selection.

3. Another method introduced by FISHER (1953) models a population by following the inheritance of "junctions" between chromosome regions with different ancestries. This approach is well suited to describing the ancestry of samples of neutral genomes (HUDSON 1990) and models of hybridization, in which selection can be approximated as acting on the proportions of genetic material derived from different source populations (BAIRD 1995). It is again intractable over long timescales, however, since the number of junctions increases geometrically.

4. PRICE (1970) gave an exact and completely general equation, in which the average change in a trait is precisely equal to its covariance with relative fitness

---

[1] *Corresponding author:* Section of Integrative Biology C0930, University of Texas, Austin, Texas 78712. E-mail: kirkp@mail.utexas.edu

plus the change due to transmission. This takes the classical approach of quantitative genetics, by following only the phenotype and disregarding the (usually unknown) genetics that underlies it.

5. Several independent developments extend Price's approach by following the mean, variance, and higher moments of the phenotypic distribution (*e.g.*, Barton and Turelli 1987; Bürger 1991; Shapiro *et al.* 1994). Each moment depends on higher moments, however, and so approximations are required to give a closed set of dynamical equations. Such approximations are accurate under restricted circumstances only. In contrast to thermodynamics, where molecular motions can be averaged out over macroscopic scales, genetic details do influence phenotypic evolution.

Barton and Turelli (1991, hereafter BT91) developed the quantitative genetic approach to provide a complete description of multilocus systems, with no restrictions on the relation between genotype and phenotype. This developed from work by Barton (1983, 1986), Barton and Turelli (1987), and Turelli and Barton (1990); it was paralleled by independent work of Christiansen (1987) and Bürger (1991). Barton and Turelli's (1991) approach contains three key elements. First, it gives a general representation of populations with multiple alleles and multiple loci. Second, it derives exact recursions for the effects of selection and recombination on allele frequencies and the associations between alleles at different loci. Third, it finds a "quasi-linkage equilibrium" approximation (QLE) that allows the recursion equations to be greatly simplified under some conditions.

Although the notation and framework of the BT91 approach are general in many respects, the methods it develops are restricted to certain forms of inheritance. Their notation can describe autosomal genes in randomly mating diploids and in nonrandomly mating haploids. It cannot, however, accommodate such complications as nonrandom mating in diploids, polyploidy, sex linkage, genome imprinting, and cytoplasmic inheritance. The main aim of this article is to show how the BT91 approach can be generalized to include all forms of inheritance. We also show how migration and mutation can be described in the same framework.

This article begins by presenting a notation that is sufficiently flexible to accommodate a variety of evolutionary forces and modes of inheritance. Next, we derive general recursion equations that describe how the genetic state of a population changes over the course of a generation. The following section shows how the selection and transmission coefficients that appear in the recursions are calculated for any particular situation. We then present the QLE approximation for the recursions, which greatly simplifies the equations. The QLE approach is illustrated with examples in the following section.

The notation and recursions set out in this article have been implemented in a set of *Mathematica* packages (Wolfram 1999) that are available on the Web at http://helios.bto.ed.ac.uk/evolgen. These packages use the general notation, in essentially the same form as in this article, and apply this notation to define functions appropriate for selection and recombination in diploids. Appendix D gives examples that show how the recursions can be computed automatically to give algebraic expressions for genetic changes in an arbitrary set of loci.

A GENERAL NOTATION FOR MULTILOCUS EVOLUTION

Here we lay out a notation that is sufficiently flexible to account for the different modes of inheritance and evolutionary forces that motivate the model. The section starts by introducing the concepts on which the notation relies, shows how the notation can describe genotypes and populations, and then describes the relation between phenotypes and genotypes.

Contexts and Positions, Selection, and Transmission: It is useful to begin by defining the word "gene" to mean a particular copy of a nonrecombining sequence at some locus in some individual. Thus two different genes may or may not reside at the same locus, and if they do they may or may not be in the same allelic state. A gene at a given locus can be found in any of a number of situations: It might be carried by a male or a female, it might have been inherited from a mother or a father, or it might reside in one deme or another. We refer to this collection of qualities as the gene's *context*. Context is a key concept in our notation and it is important in two ways. First, it determines how evolutionary forces act on the gene. A selection coefficient, for example, may depend on whether the gene is carried by a male or a female. If there is genomic imprinting, then the selection coefficient will also depend on the sex of the individual from which the gene was inherited. Second, a gene's context affects how it is transmitted. Consider two autosomal loci in a diploid individual. If there is no recombination between the loci during meiosis, the resulting gamete will carry copies of the genes that both descended from the individual's mother or father; with recombination, the gamete will carry one gene from the individual's mother and one from the father.

The information we need to specify a gene's context varies between models. In a model of a spatially structured population, the context will include geographical information. Likewise, in a life history model the context specifies the life stage of the individual carrying it. The context will include the sex of the individual carrying a gene in a model with two sexes, but not in a model of a hermaphroditic population.

Loci are referred to by lowercase italic letters. The

context is written in a series of subscripts whose elements carry the relevant information. In this article we use the convention that for diploid populations with two sexes, the first subscript of the context gives the sex of the individual carrying the gene, which we call its "sex of carrier." The second subscript gives the sex of the parent from which it was inherited, its "sex of origin." Sexes are denoted by "m" for male and "f" for female. For example, genes at a diploid locus i that are carried by females and that descended from a male (the female's father) are referred to as i_{fm}. These are then four possible contexts for genes at this locus. Contexts in hermaphrodites would not include the sex of carrier (since all individuals are the same sex), but would include the sex of origin to denote whether the gene was transmitted through an egg or a sperm. One would account for more than two sexes (as when modeling a plant population with tristyly) by simply allowing the subscripts to take more than two possible values. Subscripts can be added to denote other information, such as the deme or the family in which a gene resides.

We use the term *position* to refer to a particular locus in a particular context. An example of a position is the place in the genomes of females where genes inherited from males (their fathers) at locus i reside. Positions, like loci, are defined independently of the allelic states of the genes that reside there. With n diploid loci in a dioecious population, there are $4n$ positions that genes occupy (n loci × two sexes of carrier × two sexes of origin). Open-faced lowercase letters refer to single positions, for example $\mathbb{i} = i_{fm}$. Open-faced uppercase letters refer to sets of positions, *e.g.*, $\mathbb{A} = \{\mathbb{i}, \mathbb{j}\}$. A genome, denoted $\mathbb{G}$, is the set of all positions in an individual. With a single diploid locus i, for example, the genome for a male is $\mathbb{G} = \{i_{mm}, i_{mf}\}$, and for a female it is $\mathbb{G} = \{i_{fm}, i_{ff}\}$. The notation is summarized in Table 1.

Two fundamental kinds of events that occur during the course of a generation are *selection* and *transmission*. "Selection" accounts for variation in the contribution of different genotypes to the next stage in the life cycle. Fitnesses are assigned to either individuals or groups, depending on the form of selection. The simplest case is viability selection, in which case fitnesses are assigned to individuals. For sexual selection and assortative mating, we account for the relative contributions of mated pairs of male and female genotypes; here fitnesses are assigned to all possible kinds of pairs (BT91). Thus according to our use of the term, with nonrandom mating there can be selection even when all individual genotypes have equal survival, mating success, and fecundity. Group selection is described by assigning fitnesses to all possible combinations of genotypes that could comprise a group.

By "transmission" we mean an event that changes the context of a gene. A simple example is meiosis followed by syngamy: A gene that was carried by a female becomes a gene that was inherited from a female. The rules of transmission depend both on the gene's context and on the mode of inheritance obeyed by its locus (autosomal, Y-linked, cytoplasmic, etc.).

Describing genotypes and populations: The genotype of an individual at position $\mathbb{i}$ is represented by the indicator variable $X_{\mathbb{i}}$. With just two alleles per locus, $X_{\mathbb{i}}$ can take two values, which it is convenient to set at 0 or 1; for this special case, the frequency of allele 1 at position $\mathbb{i}$ is written $p_{\mathbb{i}}$ and the frequency of allele 0 as $q_{\mathbb{i}} = 1 - p_{\mathbb{i}}$. A fact that is useful later is that under these conventions, the expected value of $X_{\mathbb{i}}$ (averaging over all individuals in the population) is equal to $p_{\mathbb{i}}$. When there are more than two alleles, we can choose any distinct values to distinguish the alleles. If we are considering alleles that have additive effects on a quantitative trait, it is convenient to set the values equal to their effects so that the expectation of $X_{\mathbb{i}}$ equals the position's contribution to the mean value of the trait.

These conventions can be generalized if the situation demands it. When a position has pleiotropic effects on a set of k traits, the variable $X_{\mathbb{i}}$ becomes a vector of length k. In multiallelic models, vectors can also be used as an alternative to the scalar-value convention described in the last paragraph. We can define the indicator to be a vector of length equal to the number of alleles, all entries of which are zero except for the one corresponding to the allelic state. (This approach might be used in a model of genes with additive effects if, for example, alleles with the same effect mutate to other alleles at different rates.) Other conventions are also possible: For example, each position could be represented by a vector of length two, with the first giving the allelic effect and the second a label for the allele.

In general, an individual is represented by a vector **X** containing the values of his/her indicator variables for every position in the genome. It is useful below to extend this vector to include positions from more than one individual.

The genetic state of a population can be completely described by a set of statistical moments that we call *associations.* These include associations among genes within a haploid genome, which are conventionally referred to as *linkage disequilibria.* We use the more general term, however, since we are concerned with associations among arbitrary sets of positions, which may or may not be linked and may or may not be in a population that is at equilibrium. Indeed, we need to consider associations between positions that are in different individuals. The relation between different measures of linkage disequilibrium is summarized in the DISCUSSION.

Our notation allows multilocus moments to be defined in a variety of ways. The key quantities that determine the moments are a set of *reference values*. There is one reference value for each position, and the reference value for position $\mathbb{i}$ is denoted $\wp_{\mathbb{i}}$. (Note the distinction between an italic $p_{\mathbb{i}}$, which denotes an allele frequency, and a curly $\wp_{\mathbb{i}}$, which denotes a reference value.) To

TABLE 1

Summary of notation

Symbol	Meaning
Loci, contexts, and positions	
i, j	Individual loci
A, B	Sets of loci
m, f	Male, female; used to specify contexts
$\mathring{\imath} = i_{\mathrm{mf}}$	The position corresponding to locus i in the context {mf}; that is, genes at locus i in males inherited from females
$\mathbb{U} = \{\mathring{\imath}, \mathring{\jmath}, \ldots\}$	A set of positions
$\mathbb{U}\backslash\mathbb{V}$	The set $\mathbb{U}$ with the elements of the set $\mathbb{V}$ removed; defined only when $\mathbb{V}$ is a subset of $\mathbb{U}$
$\mathbb{G}$	The set of all positions within an individual genome; repeated elements are included if there are multiple alleles
$\mathbb{W}$	The set of all positions in the selection group
Summations	
$\Sigma_{\mathring{\imath} \in \mathbb{U}}$	A sum over all positions $\mathring{\imath}$ in the set $\mathbb{U}$
$\Sigma_{\mathbb{U} \subseteq \mathbb{A}}$	A sum over all subsets $\mathbb{U}$ of the set $\mathbb{A}$, including terms for the set $\mathbb{A}$ itself and the empty set $\varnothing$
$\Sigma^{*}_{\mathbb{U} \subseteq \mathbb{A}}$	A sum over all subsets $\mathbb{U}$ of the set $\mathbb{A}$, including a term for the set $\mathbb{A}$ itself but not for the empty set $\varnothing$
$\Sigma_{\mathbb{U}:U=A}$	A sum over all sets $\mathbb{U}$ such that $U = A$, that is, such that U and A are equal when the context information is removed
$\Sigma^{*}_{S+T=A}$	A sum over all nonnull partitions of the set A; for example, with $A = \{i, j\}$ there is one term with $S = i$ and $T = j$, and one term with $S = j$ and $T = i$
Allele frequencies and associations	
$X_{\mathring{\imath}}$	The indicator variable that labels the allelic state of position $\mathring{\imath}$
$\boldsymbol{X}$	A vector of X's that gives the genotype of an individual
$f(\boldsymbol{X})$	The frequency of genotype $\boldsymbol{X}$
$\wp_{\mathring{\imath}}$	The reference value for position $\mathring{\imath}$
$\zeta_{\mathring{\imath}} = X_{\mathring{\imath}} - \wp_{\mathring{\imath}}$	The deviation of an individual at position $\mathring{\imath}$ from the reference value
$E_{\boldsymbol{X}}[g(\boldsymbol{X})] = \Sigma_{\boldsymbol{X}} f(\boldsymbol{X}) g(\boldsymbol{X})$	The expectation of $g(\boldsymbol{X})$ over allelic states, $\boldsymbol{X}$
$m_{\mathring{\imath}} = E[X_{\mathring{\imath}}]$	The mean value of position $\mathring{\imath}$
$p_{\mathring{\imath}}$	The frequency of allele $X_{\mathring{\imath}} = 1$ at position $\mathring{\imath}$
$pq_{\mathring{\imath}}$	$p_{\mathring{\imath}}(1 - p_{\mathring{\imath}})$
$pq_{\mathbb{U}} = \Pi_{\mathring{\imath} \in \mathbb{U}}\, pq_{\mathring{\imath}}$	The product of allele frequencies over the set of positions $\mathbb{U}$
$\zeta_{\mathbb{U}} = \Pi_{\mathring{\imath} \in \mathbb{U}} \zeta_{\mathring{\imath}}$	The product of deviations over the set of positions $\mathbb{U}$
$D_{\mathbb{U}} = E_{\boldsymbol{X}}[\zeta_{\mathbb{U}}]$	The association between the set of positions $\mathbb{U}$
Phenotypes, fitness, and selection	
Z	The value of a phenotypic trait
$\mathbb{Z}$	The set of all positions influencing trait Z
W	Fitness
$\overline{W} = E_X[W]$	Mean fitness
$a_{\mathbb{U}}$	The selection coefficient for the set of positions $\mathbb{U}$, defined by Equation 7
Transmission	
$t_{\mathbb{A} \leftarrow \mathbb{B}}$	The proportion of genes at positions $\mathbb{A}$ that were transmitted from positions $\mathbb{B}$

describe a population, we first make a change of variables, such that the allelic state of an individual gene at position $\mathring{\imath}$ is measured relative to the reference value for that position, $\wp_{\mathring{\imath}}$:

$$\zeta_{\mathring{\imath}} = X_{\mathring{\imath}} - \wp_{\mathring{\imath}}. \tag{1}$$

Choice of the reference values is up to the investigator. Typically it is useful to define $\wp_{\mathring{\imath}}$ as the expected value of $X_{\mathring{\imath}}$ among zygotes. The reference value then has a particularly simple meaning for biallelic models. Defined that way, if there are no differences in allele frequencies between positions at a locus (*e.g.*, between males and females), the reference value is equal to the frequency of allele 1 at that locus ($\wp_{\mathring{\imath}} = \wp_{i} = p_{i}$). The new indicator variable $\zeta_{\mathring{\imath}}$ then takes the values $1 - p_{i}$ and $-p_{i}$.

Next we define the product of all the ζ's in the set of positions $\mathbb{A}$:
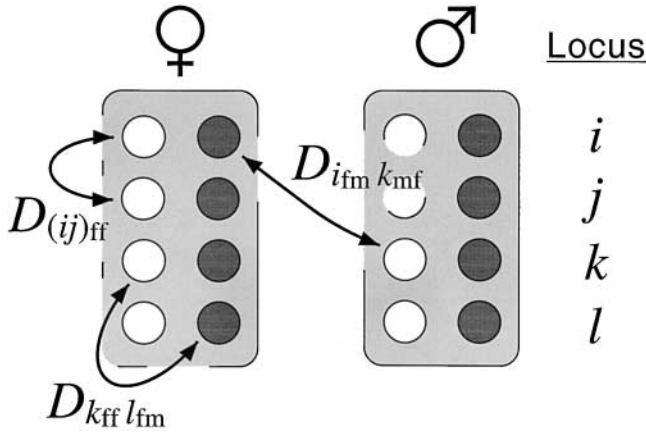
FIGURE 1.—A model of a dioecious population with four autosomal loci. Open circles are genes inherited from a female; solid circles are genes from a male. Three of the associations between the 16 positions are shown.

$$\zeta_A = \Pi_{i \in A} \, \zeta_i. \tag{2}$$

The symbol $\in$ indicates that the product includes one term for each element in the set $\mathbb{A}$. If we choose an individual at random from the population, then ζ_A is a random variable. The association between the alleles at the positions in set $\mathbb{A}$ is defined as the expectation of ζ_A taken over the whole population,

$$D_A = E_X[\zeta_A], \tag{3}$$

where $E_X[\cdot]$ denotes an expectation over the distribution of genotype frequencies. The D_A are therefore moments, that is, measures of statistical association. As an example of the notation, the association between alleles at loci i and j in a diploid male, one inherited from his female parent (mother) and the other from his male parent (father), is written $D_{i_{mf}j_{mm}}$. Products over empty sets are defined to be 1, so that $D_\varnothing = 1$. The D's are the same as BT91's C's. Figure 1 illustrates the notation. [We assume here and below that the indicators X_i have scalar values (0 or 1, say). If they are vectors, then the pairwise $D_{[i,j]}$ are matrices, and nth-order associations are tensors of rank n.]

The associations have particularly simple interpretations when the reference points are chosen to be the current allele frequencies ($\wp_i = p_i$). Moments for single positions vanish: $D_i = 0$. Associations between pairs of positions are equal to the covariance in the allelic state of genes at those positions. Departures from Hardy-Weinberg proportions are measured by D's involving pairs of positions at the same locus that have the same sex of carrier but different sexes of origin; for example, $D_{i_{mf}i_{mm}}$ for locus i in males. When there are no sex differences in allele frequencies, $D_{i_{mf}j_{mf}}$ is equal to the conventional measure of pairwise linkage disequilibrium (also called gametic phase disequilibrium) between loci i and j. Moments involving more than two

loci are measures of higher-order associations in the population. Following the standard terminology for statistical moments, we say that the associations are "centered" when the reference values are set equal to the current allele frequencies.

Equations 1–3 provide a recipe for translating genotypic frequencies into a set of reference values $\wp$ and associations D that completely describe the genetic state of a population. The reverse translation is of course also possible. For example, with biallelic loci and the reference values defined to be equal to the current allele frequencies ($\wp_i = p_i$), the frequency of genotype X is

$$f(X) = \prod_{i \in G} [X_i p_i + (1 - X_i) q_i]$$
$$+ \Sigma^*_{U \subseteq G} \{ D_U (-1)^{(|U| - \Sigma_{i \in U} X_i)} \prod_{i \in G \setminus U} [X_i p_i + (1 - X_i) q_i] \}, \tag{4}$$

where $|\mathbb{U}|$ means the number of positions in set $\mathbb{U}$, and $\mathbb{G} \backslash \mathbb{U}$ stands for the positions in set $\mathbb{U}$ that are left after those in set $\mathbb{V}$ are taken away. The first term in Equation 4, a product that includes one term for each position in the genome, gives the genotype frequency that would be found in the absence of any associations. The second term accounts for the effects of the associations. $\mathbb{G}$ is a set of all positions in a genome whose sex of carrier is the same as that of X. Expressions more general than Equation 4 that allow for multiple alleles and arbitrary definitions for the reference values can be derived using results that are developed below.

Summations of the kind seen in the second term of Equation 4 make frequent appearances in this article. The sum includes one term for each possible subset $\mathbb{U}$ of positions in the set $\mathbb{G}$, including $\mathbb{G}$ itself. When an asterisk appears, as in Equation 4, the sum does not include a term in which $\mathbb{U}$ equals the empty set, $\varnothing$. Thus, if $\mathbb{G}$ consists of the two positions $\dot{i}$ and $\dot{j}$, the sum in Equation 4 will have three terms as $\mathbb{U}$ takes on the values $\{\dot{i}\}$, $\{\dot{j}\}$, and $\{\dot{i}, \dot{j}\}$. When the summation symbol is not followed by an asterisk, the sum does include a term in which $\mathbb{U} = \varnothing$.

The notation allows for more than two alleles per locus. It does become more complicated in that event, however, because the extra degrees of freedom require us to account for associations with repeated positions. With three alleles, for example, the allele frequencies at a position are described by the two variables D_i and D_{ii}. However, when there are only two alleles per locus, associations containing repeated positions can be expressed in terms of associations with no repeated positions. (This article focuses mainly on biallelic loci, which is perhaps not a severe restriction as loci can be defined as single-nucleotide sites. Readers who are interested in loci with multiple alleles should consult BT91, or the documentation with the *Mathematica* packages, for more details about those models.)

Two basic equations for simplifying associations for biallelic loci are useful later. From Equations 1–3

$$D_{\text{U}ii} = p_i q_i D_\text{U} + (1 - 2\,p_i)D_{\text{U}i},$$

$$\zeta_i^2 = p_i q_i + \zeta_i(1 - 2\,p_i) \tag{5}$$

(see BT91, Equation 5). Here and throughout, expressions of the form $\mathbb{AB}$ stand for $\mathbb{A} \cup \mathbb{B}$, the union of sets $\mathbb{A}$ and $\mathbb{B}$; thus $D_{\text{U}i} = D_{\text{U} \cup i}$, etc.

The relation between phenotypes and genotypes: This notation is sufficiently flexible to allow for any relation between phenotypes and genotypes. Let Z be the value of a phenotypic character in an individual. This value can be written in general as a function of the individual's genotype,

$$Z(\boldsymbol{X}) = \overline{Z} + \Sigma_{\mathbb{A} \subseteq \mathbb{G}} b_\mathbb{A}(\zeta_\mathbb{A} - D_\mathbb{A}) + e_Z, \tag{6}$$

where $\overline{Z}$ is the trait mean in the population and e_Z is a random environmental component that is independent of genotype and that has mean 0. The $\zeta_\mathbb{A}$ that appear on the right-hand side are calculated from the genotype vector $\boldsymbol{X}$ on the left using Equations 1 and 2. (Note that the term in the sum corresponding to the null set $\mathbb{A} = \varnothing$ makes no contribution because $\zeta_\varnothing = D_\varnothing = 1$.)

Equation 6 can describe any kind of genetic dominance, epistasis, sex differences in expression, genomic imprinting, etc. The relation between genotype and phenotype is determined by the choice of the coefficients $b_\mathbb{A}$. If there is gene-by-environment interaction, the b becomes a function of the state of the environment; it may be convenient to include that environmental state as a component of the context. Equation 6 also applies with multiple alleles, provided that the set $\mathbb{G}$ includes the appropriate number of repeated elements. For example, suppose that there are three alleles at locus i and two alleles at locus j. For haploid genotypes, the set $\mathbb{G}$ is then defined as $\{\mathbb{i}, \mathbb{i}, \mathbb{j}\}$. The coefficients b_i, b_j, b_{ii}, b_{ij}, b_{iij} are then required to account for the 5 d.f., and the set $\mathbb{A}$ in the summation of Equation 6 ranges over all five distinct subsets of $\mathbb{G}$.

The b coefficients can become quite numerous. For example, with just two biallelic loci each with four possible contexts (say, two sexes of carrier and two sexes of origin), there are $2^8 - 1 = 255$ possible b coefficients: 8 corresponding to single positions, 28 corresponding to pairs of positions, etc. The number of coefficients drops dramatically, however, in many cases. With no epistasis, genotype-phenotype relations can be fully described with only 8 distinct coefficients, while with a completely additive model (no dominance and no epistasis), only 4 distinct coefficients are needed.

EVOLUTION BY SELECTION AND TRANSMISSION

Here we use the notation proposed above and results from BT91 to find how the genetic composition of a population changes over the course of a generation. We first show how selection and transmission change a population and then end with some statistical bookkeeping.

Selection: Fitness is just another phenotypic charac-ter. Consequently, Equation 6 applies when the trait under consideration is fitness. BT91 showed that this insight is useful because the b coefficients then take on special significance: They can be used to calculate how the genetic state of the population changes.

We noted earlier that fitness (that is, relative reproductive output) can depend on just the genotype of the individual (as with viability selection), on the genotype of a mated pair (as with any form of nonrandom mating), or on the genotypes of a larger group of individuals (as with kin or group selection). To describe the effects of selection, we need to consider together the genomes of the *selection group,* by which we mean the set of individuals that interact to determine their mutual fitness. With simple viability selection and random mating, the selection group is a single individual. Often it is useful to define the selection group to be a male and female mated pair, which allows for nonrandom mating as well as viability selection. The selection group can be expanded to more than two individuals to accommodate group selection.

We denote the set of all positions in a selection group as $\mathbb{W}$. For example, take a selection group consisting of a male and female in a mated pair. With a single biallelic diploid locus i, the selection group is $\mathbb{W} = \{i_{\text{mm}}, i_{\text{mf}}, i_{\text{fm}}, i_{\text{ff}}\}$. With three alleles, the selection group is the same, but with each element appearing twice.

The absolute fitness of a selection group is defined to be the ratio of its frequency after selection to its frequency before. When a selection group consists of more than one individual, its "frequency" before selection is equal to the product of the frequencies of the genotypes of those individuals. Take, for example, a selection group consisting of a mated pair of male and female genotypes. The frequency before selection can often be taken as the product of the frequencies of the respective male and female genotypes, since the premating "groups" are equivalent to randomly chosen pairs of males and females. The frequency of the selection group after selection is the frequency with which those genotypes are found together among all mated pairs (weighting the pairs by their relative fecundities, if they differ). This representation of selection can account for viability selection within each sex as well as nonrandom mating and fecundity selection.

The genotype of the selection group is described by the vector $\mathbf{X}$, which includes the allelic state for every position in every individual in the group. Denoting the frequencies of the group's genotype before and after selection as $f(\mathbf{X})$ and $f'(\mathbf{X})$, the group's absolute fitness as $W(\mathbf{X})$, and the population's mean fitness as $\overline{W}$, we see from Equation 6 that we can always write the expected relative fitness of the genotype of a selection group in the form

$$w(\boldsymbol{X}) = \frac{W(\boldsymbol{X})}{\overline{W}} = \frac{f'(\boldsymbol{X})}{f(\boldsymbol{X})} = 1 + \Sigma_{\text{U} \subseteq \mathbb{W}} a_\text{U}(\zeta_\text{U} - D_\text{U}) \tag{7}$$

(see BT91, Equation 6).

The coefficients $a_\mathbb{A}$ defined by Equation 7 are called *selection coefficients*. The coefficient $a_\mathbb{A}$ represents the force of selection acting on the position in set $\mathbb{A}$. These coefficients can account for any form of selection within individuals (including dominance, epistasis, and genomic imprinting) and any form of nonrandom mating. Note that selection coefficients defined this way typically depend on allele frequencies, associations, and reference values, even if the fitnesses of genotypes are constant (BT91). If phenotypes include environmental (nongenetic) components, then the frequencies $f(\ )$ and $f'(\ )$ represent expectations averaging over those components (see APPLICATIONS). Note that the selection coefficients defined by (7) differ from those defined in BT91. Although similar in form, the fitness functions are not the same, and so selection coefficients from our system and that of BT91 cannot be interchanged.

Selection coefficients have simple interpretations. With biallelic loci, the coefficient $a_\mathbb{i}$ measures the force of direct selection acting on position $\mathbb{i}$ to increase the frequency of allele 1. Selection coefficients with multiple subscripts indicate that those positions have nonadditive effects on fitness. For example, dominance at a locus i in diploid males is measured by $a_{i_{\mathrm{mf}} i_{\mathrm{mm}}}$. This coefficient measures the force of selection favoring allele 1 at locus i when it appears in two copies, one inherited from a female (the individual's mother) and the other from a male (the father). Nonadditive fitness interactions between loci are represented by selection coefficients that have multiple positions with the same sex of carrier. The selection coefficient $a_{i_{\mathrm{ff}} j_{\mathrm{ff}}}$, for example, measures the departure from additivity for the alleles at loci i and j that are carried by females and were inherited from females. The effects of nonrandom mating appear in selection coefficients that include both male and female sexes of carrier. When there are more than two alleles per locus, there are selection coefficients that have the same position repeated. [The notation can accommodate a continuum-of-alleles model where there are an infinite number of alleles per locus, provided that fitness can be approximated by a polynomial function (see BT91). It may not be possible, however, to obtain a good approximation to a continuum-of-alleles model using a finite set of moments.]

Two points about Equation 7 are worth keeping in mind. If the phenotype contains an environmental component, the relative fitness $w(X)$ is understood to mean the relative fitness averaged over that environmental variation. Second, the selection coefficients depend on how the selection group is defined. For example, with random mating the selection group can be defined as a single individual, and no selection coefficients that include both sexes of carrier appear. But if the selection group is defined as a mated pair, the fitness function of Equation 7 generates selection coefficients with both sexes of carrier even under random mating. This discrepancy is not a problem, though, since the alternative

definitions of the selection group will produce the same results so long as the definition that is chosen is used consistently throughout the calculations.

Given any set of assumptions about how genotypes (or phenotypes) affect lifetime fitness, Equation 7 can be used to calculate the corresponding selection coefficients. APPENDIX A presents a simple example of two loci under epistatic viability selection. When several selection events occur over the course of a generation, the job is made easier by calculating coefficients for each event in isolation and then combining them. For example, suppose that fitness is the product of viability through two stages of the life cycle, each represented by Equation 7 but with coefficients $b_\mathbb{U}$ and $c_\mathbb{U}$. For biallelic loci, the overall selection coefficient is

$$a_\mathbb{U} = \Sigma_{\mathbb{V} \subseteq \mathbb{U}} b_\mathbb{V} c_{\mathbb{U} \backslash \mathbb{V}}. \tag{8}$$

[To prove Equation 8, write $w_a(\boldsymbol{X}) = w_b(\boldsymbol{X})\, w_c(\boldsymbol{X})$, expand each of the w's using Equation 7, use Equations 5 to eliminate products of ζ's, and finally match the corresponding coefficients of the ζ's on the right and left sides.] In the event of weak selection, the situation can be simplified further by approximation: If the selection coefficients b and c are of order s, then $a_\mathbb{U} = b_\mathbb{U} + c_\mathbb{U}$ to leading order in s.

Given the selection coefficients, we can determine the state of a population following the selection event. BT91 showed how the new allele frequencies and associations are given by

$$D'_\mathbb{A} = D_\mathbb{A} + \Sigma_{\mathbb{U} \subseteq \mathbb{W}} a_\mathbb{U}(D_{\mathbb{A}\mathbb{U}} - D_\mathbb{U} D_\mathbb{A}). \tag{9}$$

This is our main result for the effects of selection. We see that the change in the associations for positions in set $\mathbb{A}$, represented here by the second term on the right, is equal to a sum of all the selection coefficients acting on sets of positions in the population, weighted by the association between those positions and the ones in set $\mathbb{A}$.

Equation 9, which gives the new moments in terms of the old reference values, can be used to calculate changes in allele frequencies caused by selection. If we choose the reference values to be the allele frequencies before selection ($\wp_\mathbb{i} = p_\mathbb{i}$), then the change in allele frequency at position $\mathbb{i}$ is equal to $D'_\mathbb{i}$. With two alleles per locus, Equation 9 gives

$$\Delta p_\mathbb{i} = D'_\mathbb{i} = \Sigma_{\mathbb{U} \subseteq \mathbb{W}} a_\mathbb{U} D_{\mathbb{U}\mathbb{i}} = a_\mathbb{i} p q_\mathbb{i} + \Sigma_{\substack{\mathbb{U} \subseteq \mathbb{W} \\ \mathbb{U} \neq \mathbb{i}}} a_\mathbb{U} D_{\mathbb{U}\mathbb{i}}, \tag{10}$$

where

$$pq_\mathbb{U} \equiv \prod_{\mathbb{i} \in \mathbb{U}} p_\mathbb{i} q_\mathbb{i}. \tag{11}$$

On the right side of Equation 10, the first term represents selection acting directly on alleles at position $\mathbb{i}$. The second term represents the effects of indirect selection: the force of selection acting on other positions that is transmitted to position $\mathbb{i}$ through the associations.

Equation 10 gives an exact expression for the change in allele frequency at position $\mathbb{i}$ caused by selection. If all positions at locus i are equivalent, then this is equal

to the change in allele frequency at that locus. If not, the overall change at locus i is found by averaging Δp_i over all the positions at that locus. (Note, however, that the average allele frequency is not sufficient to fully describe the population.)

Transmission: "Transmission" refers to an event that changes the contexts of genes. Obvious examples are meiosis, where a gene carried by a diploid individual becomes a gene in a haploid individual (the gamete), and fertilization, where the reverse transition happens. Migration can also be considered as a form of transmission, since genes change their context as they move from one location to another. The effects of transmission on the state of the population are determined by the *transmission coefficients*. The transmission coefficient $t_{\mathbb{A} \leftarrow \mathbb{B}}$ is defined simply as the probability that the positions in set $\mathbb{A}$ were inherited from positions in set $\mathbb{B}$. (Note that this is generally not the same as the probability that the positions in set $\mathbb{B}$ are transmitted to set $\mathbb{A}$.)

To clarify the meaning of these coefficients, consider autosomal loci in a haploid population with two sexes. The transmission coefficient $t_{i_m \leftarrow i_f}$ is the probability that a gene at locus i in a male was inherited from a gene at locus i in a female and is therefore $\frac{1}{2}$. The transmission coefficient $t_{\{i_m, j_m\} \leftarrow \{i_f, j_f\}}$ is the probability that the genes at loci i and j in a male were both inherited from a female (the mother), which is $(1 - r_{ij})/2$, where r_{ij} is the recombination rate between loci i and j.

There are three constraints on transmission coefficients. First, transmission coefficients are zero unless each position in set $\mathbb{A}$ has a corresponding position in set $\mathbb{B}$ from which it descended. This implies that sets $\mathbb{A}$ and $\mathbb{B}$ must be equal when the context information is stripped from all of their positions; that is, $t_{\mathbb{A} \leftarrow \mathbb{B}} = 0$ if $A \neq B$. (For example, $t_{i_{ff} \leftarrow j_{fm}} = 0$ because $i \neq j$; a gene at locus i cannot be descended from a gene at locus j.) Second, the coefficients representing transmission to any given set $\mathbb{A}$ must sum to 1. (In the notation introduced below, $\Sigma_{\mathbb{B}: B=A} \, t_{\mathbb{A} \leftarrow \mathbb{B}} = 1$.) A third constraint on the coefficients applies when transmission represents recombination, segregation, and/or syngamy. Then the sex of origin for each position in set $\mathbb{A}$ must equal the sex of carrier for the corresponding position in set $\mathbb{B}$, since that is the sex of the parent from which a gene in set $\mathbb{A}$ descended.

Transmission coefficients often involve recombination between groups of more than two loci. We use r_A to denote the probability that recombination occurs somewhere in the set of loci A; that is, the alleles at those loci passed to a gamete are a mixture of those inherited from the individual's mother and father. Table 2 gives the transmission coefficients for several cases of interest, including autosomes, X-linked loci, and cytoplasmic factors.

Many models assume that there is no genetic variation for the rules of transmission. In that case, the effect of transmission on the moments (the allele frequencies

TABLE 2

Examples of transmission coefficients under meiosis and syngamy

Haploid autosomes

$$t_{i_f \leftarrow i_m} = \frac{1}{2}$$
$$t_{i_f j_f \leftarrow i_m j_f} = r_{ij}/2$$

Diploid autosomes

$$t_{i_{fm} \leftarrow i_{mm}} = \frac{1}{2}$$
$$t_{i_{fm} j_{fm} \leftarrow i_{mf} j_{mm}} = r_{ij}^m/2$$
$$t_{\mathbb{A}_{mf} \leftarrow \mathbb{A}_{fm}} = (1 - r_A^f)/2$$

Sex linkage in diploids: loci i and j X-linked, k autosomal

$$t_{i_m \leftarrow i_m} = 0$$
$$t_{i_{fm} j_{fm} \leftarrow i_{mf} j_{mf}} = 1$$
$$t_{i_{ff} j_{ff} \leftarrow i_{fm} j_{fm}} = (1 - r_{ij}^f)/2$$
$$t_{i_{fm} k_{fm} \leftarrow i_{mf} k_{mf}} = \frac{1}{2}$$

and associations) that describe a population is simple. Equation 3 implies that the moments after transmission, $D_{\mathbb{A}}''$, are then just a linear combination of the moments before, $D_{\mathbb{A}}'$. When the reference values are chosen to be equal for positions at each locus, the effect of transmission is particularly simple:

$$D_{\mathbb{A}}'' = \Sigma_{\mathbb{U}: U=A} t_{\mathbb{A} \leftarrow \mathbb{U}} D_{\mathbb{U}}'. \quad (12)$$

This is our main result for the effects of transmission. The summation is over all sets of positions $\mathbb{U}$ that could become set $\mathbb{A}$ following transmission. The notation "$\mathbb{U}: U = A$" means that $\mathbb{U}$ and $\mathbb{A}$ must be equal when the context information is stripped from them, that is, when $U = A$. (Taking the example of dioecious diploids, with $\mathbb{A} = i_{fm}$, the sum in Equation 12 has four terms at $\mathbb{U}$ takes the values i_{ff}, i_{fm}, i_{mf}, and i_{mm}.) This requirement follows from the first constraint on transmission described above. Equation 12 needs modification if different positions at the same locus have different reference values, as discussed in *Changing reference values* below. The two-locus example presented in APPENDIX A shows how transmission coefficients are used in calculating changes in allele frequencies.

Equation 12 can be easily generalized to allow different genotypes to follow different transmission rules. Examples include cases where there is meiotic drive or genetic variation in recombination rates. As in Equation 6, we write the transmission coefficients as a polynomial function of genotype,

$$t_{\mathbb{A} \leftarrow \mathbb{U}}(\boldsymbol{X}) = \bar{t}_{\mathbb{A} \leftarrow \mathbb{U}} + \Sigma_{\mathbb{V} \subseteq \mathbb{T}} \delta t_{\mathbb{A} \leftarrow \mathbb{U}|\mathbb{V}} (\zeta_{\mathbb{V}} - D_{\mathbb{V}}'), \quad (13)$$

where $\mathbb{T}$ is the set of all positions that influence transmission. The transmission coefficient $\bar{t}_{\mathbb{A} \leftarrow \mathbb{U}}$ is the mean probability that genes at positions $\mathbb{A}$ were inherited from positions $\mathbb{U}$, averaged over all genotypes. The coefficient $\delta t_{\mathbb{A} \leftarrow \mathbb{U}|\mathbb{V}}$ represents the effect of the set of positions

$\mathbb{V}$ on the transmission coefficient $t_{\mathbb{A} \leftarrow \mathbb{U}}$, in the same way that the selection coefficient $a_{\mathbb{V}}$ represents the effects of the set $\mathbb{V}$ on fitness. To find the effects of transmission on the associations, substitute (13) into (12) and then average over all genotypes:

$$D''_{\mathbb{A}} = E_X[\Sigma_{\mathbb{U}:U=A} t_{\mathbb{A} \leftarrow \mathbb{U}}(\boldsymbol{X})\zeta_{\mathbb{U}}]$$

$$= \Sigma_{\mathbb{U}:U=A}\bar{t}_{\mathbb{A} \leftarrow \mathbb{U}}D'_{\mathbb{U}} + \Sigma_{\mathbb{U}:U=A}\Sigma_{\mathbb{V} \subseteq \mathbb{T}}\delta t_{\mathbb{A} \leftarrow \mathbb{U}|\mathbb{V}}(D'_{\mathbb{U}\mathbb{V}} - D'_{\mathbb{U}}D'_{\mathbb{V}}) \tag{14}$$

(*cf.* BARTON 1995, Equation 3). This equation can be used to study the evolution of alleles that modify the transmission system, for example, by altering recombination rates or the breeding system. In fact, Equation 14 is similar to Equation 9, which describes the effect of selection. One can think of selection as a special form of genotype-dependent transmission, where the transmission is between corresponding positions at consecutive stages in the life cycle and where the transmission coefficients are just the relative fitnesses.

Changing reference values: As described earlier, our system of describing a population is defined relative to a set of reference values. The investigator is free to leave these fixed or to change them as often as desired. It is often convenient, however, to change the reference values once per generation. By updating the reference values to the current allele frequencies, the associations have simple interpretations, and we can calculate the per-generation changes in allele frequencies. Moreover, updating only once per generation avoids a proliferation of alegebra, involving reference values at intermediate stages that eventually cancel. Sometimes it is convenient to update the reference values at the zygote stage. Alternatively, it may be easiest to update them before transmission, since under normal meiosis (no meiotic drive, etc.) allele frequencies are unchanged and the change in associations caused by transmission often takes a simple form when the associations have already been centered. If we are interested only in finding the evolutionary equilibrium, changing reference values from one generation to the next is not an issue.

Changing the reference values changes the associations D, because the latter are defined in terms of the former. Denote the associations before and after the change as $D''_{\mathbb{A}}$ and $D'''_{\mathbb{A}}$, respectively, and the reference values before and after as $\wp''_i$ and $\wp'''_i$. (If the reference values have not been changed since the start of the generation, then $\wp''_i = \wp_i$.) The associations after the reference values change are found using Equations 1–3:

$$D'''_{\mathbb{A}} = E_X[\Pi_{i \in A}(X_i - \wp'''_i)] = E_X[\Pi_{i \in A}(\boldsymbol{X}_i - \wp''_i) + (\wp''_i - \wp'''_i)]$$

$$= \Sigma_{\mathbb{U} \subseteq \mathbb{A}}\Pi_{i \in U}(\wp''_i - \wp'''_i)E_X[\Pi_{j \in A \backslash U}(X_j - \wp''_j)]$$

$$= \Sigma_{\mathbb{U} \subseteq \mathbb{A}}[D''_{\mathbb{A} \backslash \mathbb{U}}\Pi_{i \in U}(\wp''_i - \wp'''_i)]. \tag{15}$$

Because the sum in (15) is not asterisked, it includes the term corresponding to $\mathbb{U} = \varnothing$, which is $D''_{\mathbb{A}}$. This

last expression is the main result for changing reference values.

The previous section notes that the transmission equation (12) does not hold when different positions at the same locus have different reference values (that is, $\wp_i \neq \wp_j$ for some $i = j$). In that case, the associations between positions before transmission must be adjusted to the reference values for the positions that those genes will occupy after transmission,

$$D''_{\mathbb{A}} = \Sigma_{\mathbb{U}:U=A} \; t_{\mathbb{A} \leftarrow \mathbb{U}}\Sigma_{\mathbb{U} \subseteq \mathbb{A}}[D'_{\mathbb{A} \backslash \mathbb{U}}\prod_{i \in U}(\wp_i - \wp_i*)], \tag{16}$$

where $\wp_i*$ is the reference value for the position in set $\mathbb{A}$ that corresponds to $\mathbb{i}$ in set $\mathbb{U}$.

With the exception of random drift, all of genetic evolution can be concisely represented by two equations: Equation 9 for the effects of selection and other deterministic forces and Equation 12 or 16 for the effects of transmission. These can be supplemented by Equation 15, which does the bookkeeping needed to ensure that the associations have a simple interpretation. Other deterministic forces, like mutation and migration, can also be described by these equations. Mutation can be represented as a form of frequency-dependent selection and migration as a form of transmission (since genes change their contexts when they move). It is easier to find the effects of mutation directly, however, which we do below. Before doing that, however, we develop an approximation that greatly simplifies the equations for selection and transmission.

THE QLE APPROXIMATION

The recursions derived above can be used to calculate the exact dynamics for a wide range of multilocus population genetic models. Although this approach may give more insight than directly following genotype frequencies, it will not necessarily be any more tractable. That is because exact results require following the dynamics of the same number of variables, regardless of whether they are genotype frequencies or moments (that is, allele frequencies and associations). One of the great appeals of the moment-based approach introduced by BT91 is that in some situations expressions for the associations can be greatly simplified by approximation. In this section, we derive approximate expressions for the associations and changes in allele frequencies when the population is in a state of QLE. The concept was introduced by KIMURA (1965) and greatly generalized by NAGYLAKI (1993) and NAGYLAKI *et al.* (1999); a concise summary of those results is given in BÜRGER (2000, p. 82).

The first fundamental assumption we must make is that all the associations D are of order a, by which we mean that they are not larger than a constant factor times the largest of the a's. BT91 shows that this condition is met when the forces that generate associations within a sex (epistasis, migration, etc.) are weak relative

to recombination and when nonrandom mating is not strong. An intuitive justification is that the associations are produced by evolutionary forces that are of order a (see Equation 9) and will not accumulate to values that are much larger than that if the forces breaking them down (recombination, segregation, and mutation) are sufficiently strong. The second assumption needed for the QLE approximation is that all the selection coefficients a are $\ll 1$. BT91 shows that when these two conditions hold, a population rapidly settles into a state where the allele frequencies are changing slowly, and the associations are close to the equilibrium values they would reach if the allele frequencies were in fact stationary (see also Nagylaki 1993). We can then neglect terms involving higher powers of the a's and also higher powers of the D's (because they are of order a). Furthermore, the effects of a series of events of selection, migration, and mutation can be added together, provided they are each of order a (Kirkpatrick and Servedio 1999).

Approximations for the associations: We assume that there are two alleles at each locus, which simplifies the analysis. The approach can be extended to multiple alleles following the leads of BT91. The main results developed below are illustrated with a simple two-locus example in appendix a.

Consider a life cycle in which we define the reference values to be the allele frequencies at the zygote stage. A series of selection events occur during the course of the generation. The generation ends with transmission, creating the zygotes for the next generation. We seek to derive an approximation for the dynamics of allele frequencies that is accurate up to (and including) terms of order a^2, which we denote $O(a^2)$. From Equation 10, we see that approximation requires in turn that we find an approximation for the associations D that is accurate to order a. To do that we find the values that give an equilibrium for the recursion equations for the D that are accurate to order a; those solutions are our QLE approximations for the associations. The results apply not just to selection but to other deterministic forces that generate associations (such as migration) so long as they are weak relative to recombination and segregation. To simplify the derivations, we assume that there is no genetic variation in the transmission coefficients, an assumption that could be relaxed (see Barton 1995). However, we must assume that the transmission coefficients are sufficiently large that forces of order a do not eventually generate strong associations. (This requires that the largest of absolute values of the selection coefficients a is much smaller than the smallest of the absolute values of the eigenvalues of the matrix of transmission coefficients $t_{\mathbb{A} \leftarrow \mathbb{U}}$.)

To begin deriving an approximate recursion for the D, Equation 9 gives the cumulative effect of selection and other deterministic forces on the associations between positions in set $\mathbb{A}$,

$$D'_{\mathbb{A}} = D_{\mathbb{A}} + \Sigma^*_{\mathbb{U} \subseteq \mathbb{W}} a_{\mathbb{U}} D_{\mathbb{U}\mathbb{A}} + O(a^2)$$
$$= D_{\mathbb{A}} + a_{\mathbb{A}} \, pq_{\mathbb{A}} + O(a^2), \qquad (17)$$

where $\mathbb{A}$ is a set of distinct positions and $pq_{\mathbb{A}}$ is defined by Equation 11. The asterisked Σ^* indicates that the sum does not include the term with $\mathbb{U} = \varnothing$; it has been separated out to give the first term, $D_{\mathbb{A}}$.

The first step of Equation 17 follows from Equation 9 because the D are of order a and therefore the term $a_{\mathbb{U}} D_{\mathbb{U}} D_{\mathbb{A}}$ in Equation 9 is of order a^3 and so can be neglected. The second step follows because the term $a_{\mathbb{U}} D_{\mathbb{U}\mathbb{A}}$ in the first line is of order a^2 except when $\mathbb{U} = \mathbb{A}$, in which event the reduction formula Equation 5 gives us $D_{\mathbb{U}\mathbb{A}} = D_{\mathbb{A}\mathbb{A}} = pq_{\mathbb{A}} + O(a)$.

The effect of changing reference values can also be simplified. Equation 10 shows that the change in the allele frequencies is of order a. If we define the reference values to be the allele frequencies, then the quantities $(\wp''_i - \wp'''_i)$ that appear in the product in Equation 15 are of order a. That equation therefore reduces simply to $D'''_{\mathbb{A}} = D''_{\mathbb{A}} + O(a^2)$, meaning that the effect of updating the reference values can be neglected. Assume that differences between positions at the same locus are $O(a)$, which holds under normal sexual inheritance. With help from Equations 12 and 17 we then get the full recursion for the associations over an entire generation:

$$D'''_{\mathbb{A}} = \Sigma_{\mathbb{U}: U = A} t_{\mathbb{A} \leftarrow \mathbb{U}} (D_{\mathbb{U}} + a_{\mathbb{U}} pq_{\mathbb{U}}). \qquad (18)$$

On setting $D'''_{\mathbb{A}} = D_{\mathbb{A}} \equiv \tilde{D}$, we get a QLE approximation for the associations that is accurate to order $O(a)$:

$$\hat{D}_{\mathbb{A}} = \Sigma_{\mathbb{U}: U = A} t_{\mathbb{A} \leftarrow \mathbb{U}} \tilde{D}_{\mathbb{U}} + \Sigma_{\mathbb{U}: U = A} t_{\mathbb{A} \leftarrow \mathbb{U}} a_{\mathbb{U}} pq_{\mathbb{U}}. \qquad (19)$$

The first sum on the right is zero if the positions in $\mathbb{A}$ include more than one sex of origin. That is because if $\mathbb{A}$ includes more than one sex of origin, then $\mathbb{U}$ in the first sum would have to include more than one sex of carrier. But $D_{\mathbb{U}}$ is of order a^2 if the positions in $\mathbb{U}$ include both sexes of carrier, since it represents associations between alleles in two randomly chosen zygotes.

Equation 19 is the main result of this section. It gives the solutions for the associations implicitly: The QLE value $\tilde{D}_{\mathbb{A}}$ on the left side depends on the QLE values for the other associations, which appear as $\tilde{D}_{\mathbb{U}}$ on the right side. The relationship is linear (because of the linear form of the transmission Equation 12), and so the solution can always be found using standard matrix algebra. Thus the $\tilde{D}$ can be calculated directly, using standard matrix methods, given a set of transmission rules that specify the t's, a set of allele frequencies from which we can calculate $pq_{\mathbb{U}}$, and a set of selection coefficients a. We have implicitly assumed here that the selection coefficients are constant in time, but the approach can be generalized to changing environments (see BT91, Appendix B; Barton 1995, Appendix 4). Briefly, the associations are determined by a time average of the

selection coefficients that are discounted by terms like $\exp(-r_A t)$. If, however, the environment changes on a time scale that is slow relative to the rate at which associations are changed by transmission, then the results given above apply.

The next two sections illustrate how to do this by carrying out the calculations for autosomal genes in dioecious haploids and for autosomal, sex-linked, and cytoplasmic genes in diploids.

Autosomal genes in haploids: The QLE approximation for autosomal inheritance in a haploid population with two sexes was found by BT91. This section rederives their result to illustrate the new notation and how to use Equation 19 to find a QLE approximation.

The context for each gene now contains only its sex of carrier. That is because an individual carries only one gene at each locus, rather than the two that must be distinguished in the case of diploids. The comments following Equation 19 imply that for this case the first sum on its right side reduces to $(t_{A \leftarrow A_f} \tilde{D}_{A_f} + t_{A \leftarrow A_m} \tilde{D}_{A_m})$, where A_f stands for set A with the sexes of carrier for all its positions converted to f and similarly for A_m. The associations among a set of autosomal loci are equal in male and female zygotes, so $\tilde{D}_{A_f} = \tilde{D}_{A_m}$, and further those quantities must be equal to $\tilde{D}_A$ on the left side of Equation 19 because all the positions in A must have the same sex of carrier. With no sex differences in recombination we have $t_{A \leftarrow A_f} = t_{A \leftarrow A_m} = (1 - r_A)/2$, where $(1 - r_A)$ is the probability that the loci in set A are not broken apart by recombination, and the factor of ½ accounts for the probability that genes in set A were inherited from a given parent.

Putting those facts together gives the QLE approximation

$$\tilde{D}_A = \frac{1}{r_A} \Sigma_{U:U=A} t_{A \leftarrow U} a_U p q_U + O(a^2). \tag{20}$$

This is equivalent to BT91's Equation 25. Some superficial differences are caused by three changes in notation. Their result is expressed in terms of recombination rates rather than transmission rates. Second, BT91 separately defined within-male, within-female, and between-sex (nonrandom mating) selection coefficients; all of these are included in the sum on the right side of Equation 20. Last, they counted separately selection coefficients with different permutations of the same set of positions, which generates the combinatorial terms in their expression.

Autosomal, sex-linked, and cytoplasmic genes in diploids: Now consider autosomal genes in a diploid population with two sexes. The context for a gene now includes both its sex of carrier and sex of origin. We allow for nonrandom mating and sex differences in selection and recombination. To simplify the calculation, however, we assume that there is no genetic variation in recombination rates and no genomic imprinting (that is, an

allele's sex of origin does not affect its expression). The approach outlined here can be directly extended to allow for more than two sexes, as might be appropriate to describe a population with partial selfing.

Careful consideration of Equation 19 shows that the associations fall into three cases. Case 1 are associations among a set of positions A that include both sexes of carrier. The QLE approximation for these associations is simply $\tilde{D}_A = 0$. That is because these represent associations between genes in two or more randomly chosen zygotes.

Case 2 are the associations between a set of positions that all have the same sex of carrier, but some have a male and others a female sex of origin. This kind of association exists for some sets of positions (for example, autosomal), but not others (for example, sets with only cytoplasmic loci). If they do exist, Equation 19 gives the QLE approximation

$$\tilde{D}_A = \Sigma_{U:U=A} t_{A \leftarrow U} a_U p q_U. \tag{21}$$

These associations come from nonrandom mating in the previous generation: Associations between genes with different sexes of origin within an individual appear when there are correlations between the genotypes of mating males and females in the previous generation. These associations, which include Hardy-Weinberg disequilibria (an excess or deficit of heterozygotes), are zero under random mating because then the selection coefficients for positions with both sexes of carrier are of order a^2. The transmission coefficient $t_{A \leftarrow U}$ can be translated into recombination rates according to the way that the genes in set A are inherited, as discussed above in the section on transmission.

Case 3, the last category of association, is when all positions in A have the same sex of carrier and all have the same sex of origin. Here D_A represents an association among genes within a single individual that were inherited from the same parent. The QLE approximations for this case depend on how the genes in set A are inherited. They can be calculated by first writing out Equation 19 for the associations that do exist, given the mode of inheritance, out of the four possible cases $\tilde{D}_{A_{ff}}$, $\tilde{D}_{A_{fm}}$, $\tilde{D}_{A_{mf}}$, and $\tilde{D}_{A_{mm}}$, where, for example, A_{fm} means that all positions in set A have a female sex of carrier and a male sex of origin. Inspection of the transmission coefficients reveals that these expressions do not depend on any associations that do not exist (*e.g.*, $\tilde{D}_{A_{fm}}$ does not depend on $\tilde{D}_{A_{mm}}$ when all genes in set A are X-linked, because $t_{A_{fm} \leftarrow A_{mm}} = 0$). Last, solve the resulting equations. That procedure leads to the following results for autosomal, X-linked, Y-linked, and cytoplasmic genes. The transmission coefficients used in the calculations are shown in Table 3.

When all the genes in set A are autosomal, all four of the possible case 3 associations exist. Solving Equation 19 then gives

<div style="column: left">

TABLE 3

Transmission coefficients for case 3 associations at QLE

All four contexts exist (Equation 22)

$$t_{A_{mm} \leftarrow A_{mm}} = t_{A_{mm} \leftarrow A_{mf}} = t_{A_{fm} \leftarrow A_{mm}} = t_{A_{fm} \leftarrow A_{mf}} = (1 - r_A^m)/2$$

$$t_{A_{mf} \leftarrow A_{fm}} = t_{A_{mf} \leftarrow A_{ff}} = t_{A_{ff} \leftarrow A_{fm}} = t_{A_{ff} \leftarrow A_{ff}} = (1 - r_A^f)/2$$

A_{mm} does not exist (Equation 24)

$$t_{A_{fm} \leftarrow A_{mm}} = 0$$

$$t_{A_{fm} \leftarrow A_{mf}} = (1 - r_A^m)$$

$$t_{A_{mf} \leftarrow A_{fm}} = t_{A_{mf} \leftarrow A_{ff}} = t_{A_{ff} \leftarrow A_{fm}} = t_{A_{ff} \leftarrow A_{ff}} = (1 - r_A^f)/2$$

Only A_{mf} and A_{ff} exist (Equation 25)

$$t_{A_{mf} \leftarrow A_{fm}} = t_{A_{ff} \leftarrow A_{fm}} = 0$$

$$t_{A_{mf} \leftarrow A_{ff}} = t_{A_{ff} \leftarrow A_{ff}} = 1 - r_A^f$$

Only A_{mm} exists (Equation 26)

$$t_{A_{mm} \leftarrow A_{mm}} = (1 - r_A^m)$$

$$t_{A_{mm} \leftarrow A_{mf}} = 0$$

These coefficients, which apply to diploids, are used to derive Equations 22–26. See the text for further details.

$$\tilde{D}_{A_{fx}} = \tilde{D}_{A_{mx}}$$

$$= \left(\frac{pq_A}{r_A^f + r_A^m} \right) \{ [a_{A_x} + F(A_{xx})](1 + r_A^{\sim x}) + [a_{A_{\sim x}} + F(A_{x \sim x})](1 - r_A^x) \},$$

(22)

where

$$F(A_{xy}) = \Sigma_{S+T=A}^* t_{A_{xy} \leftarrow \{S_y{}_f T_{ym}\}} a_{S_f T_m}.$$

(23)

Here r_A^x is the recombination rate for set A in sex x, where x can take the values m and f, and $\sim x$ stands for the opposite sex of x (for example, if $x = $ f then $\sim x = $ m). The term $F(\cdot)$ results from nonrandom mating, which creates associations between alleles inherited from different parents in the next generation. These alleles are brought together in single gametes by recombination, producing associations within the same gametic genome (*i.e.*, linkage disequilibria) two generations later. The summation in (23) is over all the different ways that the set of loci A can be partitioned into nonnull sets. With $A = \{i, j, k\}$, for example, the sum includes six terms: $S = \{i\}$ and $T = \{j, k\}$, $S = \{j, k\}$ and $T = \{i\}$, $S = \{j\}$ and $T = \{i, k\}$, etc. $F(\cdot)$, which appears in results below, vanishes under random mating. The selection coefficients a_{A_f} and a_{A_m} reflect epistatic selection within each sex and vanish when there is no epistasis.

When A is a set of either all X-linked genes or a mixture of X-linked and autosomal genes, $\tilde{D}_{A_{mm}}$ does not exist. Solving Equation 19 for the remaining three kinds of associations in case 3 gives

$$\tilde{D}_{A_{ff}} = \tilde{D}_{A_{mf}} = \frac{pq_A}{r_A^f(2 - r_A^m) + r_A^m}$$

$$\times \{2[a_{A_f} + F(A_{ff})] + [a_{A_m} + F(A_{fm})](1 - r_A^f)\},$$

</div>

<div style="column: right">

$$\tilde{D}_{A_{fm}} = \frac{pq_A}{r_A^f(2 - r_A^m) + r_A^m}$$

$$\times \{2[a_{A_f} + F(A_{ff})](1 - r_A^m) + [a_{A_m} + F(A_{fm})](1 + r_A^f)\}.$$

(24)

When A is a mixture of cytoplasmic and nuclear genes or only a set of cytoplasmic genes, then $\tilde{D}_{A_{fm}}$ and $\tilde{D}_{A_{mm}}$ do not exist. Using the transmission coefficients given in Table 3, we find the remaining two kinds of case 3 associations are

$$\tilde{D}_{A_{ff}} = \tilde{D}_{A_{mf}} = \frac{pq_A}{r_A^f} [a_{A_f} + F(A_{ff})].$$

(25)

If A has only cytoplasmically inherited genes, then $r_A^f = 0$ and that expression becomes undefined. In this situation, there is no recombination to break down associations generated by selection, so they become large and the QLE approximation fails.

A similar situation occurs with Y-linkage. When A is a set of Y-linked genes or a mixture of Y-linked and autosomal genes, the only kind of case 3 association that exists is

$$\tilde{D}_{A_{mm}} = \frac{pq_A}{r_A^m} [a_{A_m} + F(A_{mm})].$$

(26)

If A includes only Y-linked loci, then $r_A^m = 0$ and again there is no QLE approximation for $\tilde{D}_{A_{mm}}$.

We stop our inventory of the case 3 associations at this point. There are modes of transmission not discussed above, as, for example, when male, female, and hermaphroditic individuals occur in the population. Associations for those cases can be calculated, however, from Equation 19 using the same method.

Changes in allele frequencies at QLE: When a population is in quasi-linkage equilibrium, changes in allele frequencies can be approximated by simple expressions. The exact expression for allele frequency change is given by Equation 10. The QLE approximation for Δp_i is found by substituting into that equation the QLE approximations for the associations. We saw in the previous section that the approximations for the D depend on the specifics of the genetic system that is being modeled. When the model consists of diploid autosomal loci, for example, the associations are given by Equations 21 and 22.

MUTATION AND MIGRATION

Mutation and migration are two other deterministic forces that change the genetic composition of a population. This section shows how they change allele frequencies and the associations between loci.

Mutation: While the effects of mutation on allele frequencies have been understood since Haldane (1927), its effects on associations among loci have not been fully worked out. Indeed, it seems to us that the general case,

</div>

in which there is an arbitrary matrix of mutation rates between alternative alleles and an arbitrary representation of allelic state, does not lead to simple expressions. BÜRGER (2000, p. 190) gives expressions for the effects of "random walk" mutation, in which the change in allelic effect X_i has a distribution independent of its current value, and for "house of cards" mutation, in which the absolute value of the new mutation has a constant distribution. With multiple alleles, it is natural to represent the allelic state by a vector of length equal to the number of alleles (*e.g.*, $X_i = \{0, 0, 1, 0\}$ represents a position $\mathbb{i}$ carrying the third of four alleles). Then, the change in moments due to mutation at position $\mathbb{i}$ is just given by multiplying the moments by the matrix of mutation rates. BAAKE (2001) uses this representation to derive an explicit expression for the change in multilocus associations due to mutation and weak recombination.

Here, we give the general result for two alleles at each locus, with allelic state taking the values 0 and 1; see APPENDIX B for details. Denote the mutation rate at position $\mathbb{i}$ from allele 0 to allele 1 as u_i and the reverse mutation rate from 1 to 0 as v_i. The change in frequency of allele 1 at position $\mathbb{i}$ caused by mutation is

$$\Delta p_i = v_i q_i - u_i p_i \tag{27}$$

(HALDANE 1927). The associations after mutation are

$$D_{\mathbb{A}}'' = D_{\mathbb{A}} \Pi_{i \in \mathbb{A}} (1 - u_i - v_i), \tag{28}$$

where the reference values are equal to the current (postmutation) allele frequencies. This result shows that mutation erodes the associations. A striking fact is that the rate at which they decline depends only on the mutation rates at the positions involved and not on the allele frequencies or other properties of the genetic state of the population.

Migration: The effects of migration on single loci were found by WRIGHT (1931), and later workers have understood that migration can generate associations between loci. General results for the effects of migration on associations, however, have apparently not been worked out previously.

The change in the frequency of allele 1 at position $\mathbb{i}$ caused by migration is

$$\Delta p_i = m(p_i^M - p_i^R), \tag{29}$$

where p_i^M and p_i^R are the allele frequency among the migrants and residents, respectively (WRIGHT 1931). The exact values for the centered associations after migration are derived in APPENDIX C as

$$D_{\mathbb{A}}'' = \Sigma_{\mathbb{U} \subseteq \mathbb{A}} (-m)^{|\mathbb{U}|} d_{\mathbb{U}} \{(1 - m) D_{\mathbb{A} \setminus \mathbb{U}}^R + m \Sigma_{\mathbb{V} \subseteq \mathbb{A} \setminus \mathbb{U}} d_{\mathbb{V}} D_{(\mathbb{A} \setminus \mathbb{U}) \setminus \mathbb{V}}^M\}, \tag{30}$$

where

$$d_{\mathbb{U}} = \Pi_{i \in \mathbb{U}} (p_i^M - p_i^R) \tag{31}$$

and $d_{\varnothing} \equiv 1$. The $D_{\mathbb{A}}^M$ are the centered associations in

the migrant population and the $D_{\mathbb{A}}^R$ are the centered associations among the residents before migration. (That is, the reference values for $D_{\mathbb{A}}^M$ are the allele frequencies among the migrants, and the reference values for $D_{\mathbb{A}}^R$ are the allele frequencies in the residents before migration.) The first term in Equation 30 is produced when $\mathbb{U} = \varnothing$ and $\mathbb{V} = \mathbb{A}$, giving $(1 - m) D_{\mathbb{A}}^R + m d_{\mathbb{A}}$.

We illustrate the use of this result with two special cases that may be of general interest. The first is the association between pairs of positions generated by migration. The association between positions $\mathbb{i}$ and $\mathbb{j}$ after migration is in general

$$D_{ij}'' = (1 - m) D_{ij}^R + m D_{ij}^M + m(1 - m)(p_i^M - p_i^R)(p_j^M - p_j^R). \tag{32}$$

The associations following migration are a weighted average of the associations in the contributing populations (the first two terms) and a component caused by differences in allele frequencies in the two populations (the third term).

A second situation that may also be of general interest is when the associations among the residents and migrants are initially zero. Then Equation 30 gives

$$D_{\mathbb{A}}'' = d_{\mathbb{A}} \{(-m)^{|\mathbb{A}|} (1 - m) + m(1 - m)^{|\mathbb{A}|}\}. \tag{33}$$

Thus the associations generated in this case are simply proportional to the differences in allele frequencies between the contributing populations. More generally, Equation 30 shows that if a pair of demes is initially at linkage equilibrium, associations will evolve such that $D_{\mathbb{A}}$ remains proportional to the product of allele frequency differences at the loci involved, $d_{\mathbb{A}}$ (BARTON 2000).

The exact results can be used to find simple approximations for the effects of migration on the associations. Equation 30 shows that the change caused by migration is

$$\Delta_m D_{\mathbb{A}} = m d_{\mathbb{A}} + O(mD, m^2). \tag{34}$$

To find an approximation for $D_{\mathbb{A}}$ at a migration-recombination equilibrium, we need the change caused by transmission. Taking the example of a set of autosomal positions with the same sex of origin in a random mating population, that change is $\Delta_t D_{\mathbb{A}} = -r_{\mathbb{A}} + O(D^2)$. Setting the net change to zero gives a leading-order approximation for the associations at a migration-recombination balance:

$$\hat{D}_{\mathbb{A}} = \frac{m d_{\mathbb{A}}}{r_{\mathbb{A}}} + O(mD, m^2, D^2). \tag{35}$$

That result was derived by KIRKPATRICK and SERVEDIO (1999) by treating migration as a form of frequency-dependent selection. Our new approach extends that earlier result in two ways. It can be applied to genes other than autosomes by using the appropriate transmission probabilities. Further, the approximation can be

expanded to include higher-order terms, for example, those involving m^2 and mD.

APPLICATIONS

This section uses the machinery described above to develop results for the effects of natural and sexual selection. The aim is both to illustrate how these methods work and to develop some results that are biologically interesting in their own right.

In the first application, we find the selection coefficients generated by natural selection acting on an additive polygenic trait and use those results to study how it evolves under autosomal inheritance. Next, we find approximations for the genetic correlation between a female mating preference and a male display trait produced by sexual selection. Then, we see how the mode of inheritance affects this correlation by deriving results for haploid autosomal, diploid autosomal, and diploid X-linked genes.

Quadratic stabilizing selection on an additive polygenic trait: Many problems in evolutionary biology involve evolution of traits controlled by multiple genes of approximately additive effect. In this section we derive exact expressions for the selection coefficients on single positions and sets of positions that result from quadratic stabilizing selection. These can be used to calculate the evolutionary changes in the mean, variance, and higher moments of the trait. The same methods can be used to calculate selection coefficients more generally, for an arbitrary form of selection acting on any number of additive genes.

Consider a trait controlled by a set of genes with additive effects under stabilizing selection. The calculations illustrate a general strategy for calculating selection coefficients: Write an explicit model for the phenotype, write the fitness function as a polynomial, substitute the expression for the phenotype into that fitness function, equate the result with Equation 7, and finally pick out the coefficients of the fitness function that correspond to the a's. This example is very similar to one in BT91 (p. 244). It introduces readers who are not familiar with that article to the approach and shows those who are how the new notation works.

The model for the phenotype of an individual comes from Equation 6, which simplifies under our assumption that genes have additive effects,

$$Z = \overline{Z} + \Sigma_{i \in \mathbb{W}} b_i \zeta_i + e_Z, \qquad (36)$$

where e_Z, the random environmental component, has variance V_e. The sum is over all positions i that affect expression of the trait and hence fitness; for example, with autosomal genes in a diploid male, the set $\mathbb{W}$ includes positions inherited from both males and females: $\{i_{mm}, i_{mf} \ldots \}$.

Our model for fitness is the quadratic function

$$W(Z) = 1 - \frac{Z^2}{2\omega^2}, \qquad (37)$$

where ω^2 is the width of the fitness function and is inversely related to the strength of stabilizing selection. The trait has been scaled so that the fitness optimum is at $Z = 0$. Clearly, selection must be weak enough that fitness remains positive [$\mathrm{Var}(Z) \ll \omega^2$]. This quadratic model is a weak-selection approximation to other forms of stabilizing selection, such as the Gaussian $W(Z) \propto \exp(-Z^2/2\omega^2)$ (see BT91).

Substituting Equation 36 into Equation 37 and averaging over the environmental variation gives

$$E_e[W(X)] = 1 - \frac{V_e}{2\omega^2} - \frac{1}{2\omega^2}(\Sigma_i b_i \zeta_i + \overline{Z})^2$$

$$= 1 - \frac{V_e}{2\omega^2} - \frac{\overline{Z}^2}{2\omega^2} - \frac{\overline{Z}}{\omega^2}\Sigma_{i \in \mathbb{W}} b_i \zeta_i - \frac{1}{2\omega^2}\Sigma_{i,j \in \mathbb{W}} b_i b_j \zeta_i \zeta_j, \qquad (38)$$

where E_e is the expectation over the distribution of e_Z. To find the selection coefficients a_U, the fitness function must be put in the form of Equation 7. We do that by rewriting the last sum in Equation 38 to separate out the terms with repeated positions and then using Equation 5 to reduce those terms:

$$E_e[W(X)] = 1 - \frac{V_e}{2\omega^2} - \frac{\overline{Z}^2}{2\omega^2} - \frac{\overline{Z}}{\omega^2}\Sigma_{i \in \mathbb{W}} b_i \zeta_i$$

$$- \frac{1}{2\omega^2}\Sigma_{i \in \mathbb{W}} b_i^2 [pq_i - \zeta_i(p_i - q_i)] - \frac{1}{\omega^2}\Sigma_{i<j} b_i b_j \zeta_i \zeta_j. \qquad (39)$$

The selection coefficients are found from Equation 39 by dividing through by the mean fitness $\overline{W}$ and identifying the coefficients of ζ_i and $\zeta_i \zeta_j$. All coefficients a_U involving more than two positions ($|\mathbb{U}| > 2$) are zero. The selection acting on the single position i is

$$a_i = -b_i \frac{\overline{Z}}{\overline{W}\omega^2} + b_i^2 \frac{1}{2\overline{W}\omega^2}(p_i - q_i). \qquad (40)$$

The first term represents the effect of directional selection, which pushes the mean of the trait toward the optimum at $Z = 0$. It is proportional to b_i, which is the effect that allele 1 at position i has on the phenotype. The second term represents the effect of stabilizing selection, which favors fixation by driving the allele frequency toward 0 if $p_i < \frac{1}{2}$ and toward 1 if $p_i > \frac{1}{2}$ (WRIGHT 1935). This effect will be weak when the effect of the locus on the phenotype is small relative to the width of the fitness function ($b_i \ll \omega$). [Note that if allele frequencies differ in males and females ($p_{imx} \neq p_{ifx}$) or in genes inherited from males and females ($p_{ixm} \neq p_{ixf}$), then the different positions at a single locus can have different selection coefficients (e.g., $a_{i_{xm}} \neq a_{i_{xf}}$) even if selection acts identically on males and females.]

The selection coefficient acting jointly on positions i

and $\mathbb{j}$ is given by the coefficient of $\zeta_i\zeta_j$ in the fitness function Equation 39:

$$a_{ij} = -\frac{1}{\overline{W}\omega^2}b_i b_j \quad \text{for } \mathbb{i} \neq \mathbb{j}. \tag{41}$$

To see the meaning of this coefficient, adopt the convention of naming the alleles such that allele 1 produces a larger phenotype than allele 0 at each locus; then b_i and b_j are positive. With stabilizing selection, a_{ij} is negative, indicating that selection favors the combination of allele 0 at one locus with allele 1 at the other. This is simply the well-known fact that stabilizing selection produces negative associations. Disruptive selection can be modeled by taking ω^2 as negative. In that event, a_{ij} is positive: Selection favors positive associations between alleles that increase the trait and also associations between alleles that decrease the trait.

To complete the analysis of this model we find how the population evolves. To determine the state of the population in the following generation, we need to make some assumptions about inheritance, that is, the rules of transmission. To keep things simple, we assume that $\mathbb{i}$ and $\mathbb{j}$ are autosomally inherited diploid genes, that mating is random, and that there is no meiotic drive. We define the reference values for each position to be the frequency of allele 1 there.

The change in allele frequencies caused by selection is found using Equation 10 with some help from Equation 5,

$$\Delta p_i = \Sigma_{U \subseteq W} a_U D_{iU}$$
$$= a_i p_i q_i + \Sigma_{j \neq i} a_j D_{ij} + \Sigma_{j \neq i} a_{ij} D_{iij} + \Sigma_{\substack{j,k \neq i \\ j < k}} a_{jk} D_{ijk}$$
$$= a_i p q_i + \Sigma_{j \neq i}[a_j - a_{ij}(p_i - q_i)]D_{ij} + \Sigma_{\substack{j,k \neq i \\ j < k}} a_{jk} D_{ijk}, \tag{42}$$

where the special cases such as $\mathbb{i} = \mathbb{j}$ have been separated out. Note that the sums are over all *positions* and so include contributions from genes on both maternal and paternal genomes.

Since transmission does not change allele frequencies, the overall change in the mean of the trait from the start of one generation to the next can be found by summing Δp_i and substituting for the a_i, a_{ij} from Equations 40 and 41. This leads to the simple expression

$$\Delta \overline{Z} = \Sigma_i b_i \Delta p_i = -\frac{\overline{Z}}{\overline{W}\omega^2}\Sigma_{i<j} b_i b_j D_{ij} - \frac{1}{2\overline{W}\omega^2}\Sigma_{\substack{j,k \neq i \\ j<k}} b_i b_j b_k D_{ijk}. \tag{43}$$

The result is a sum of two terms. The first is the standard equation for the selection response of a quantitative trait under directional selection: It is the product of the directional selection gradient, which in this case is $-\overline{Z}/\omega^2$, and the additive genetic variance for the trait, which appears as the first summation on the right side. The second term is the result of stabilizing selection

acting on the genetic variance. That form of selection also changes allele frequencies, and hence the trait mean, if the trait distribution is skewed: The second sum is just the skew of the distribution of breeding values of Z (Bürger 1991; Turelli and Barton 1994).

Now consider the change in the association D_{ij}. Using Equations 5 and 9, we find that the (uncentered) value after selection is

$$D'_{ij} = \Sigma_{U \subseteq W} a_U(D_{ijU} - D_{ij}D_U)$$
$$= D_{ij} + (a_i D_{iij} + a_j D_{ijj}) + \Sigma_{k \neq i,j} a_k D_{ijk}$$
$$\quad + a_{ij}(D_{ijij} - D_{ij}^2) + \Sigma_{k \neq i,j} a_{ik}(D_{iijk} - D_{ij}D_{ik})$$
$$\quad + \Sigma_{k \neq i,j} a_{jk}(D_{ijjk} - D_{ij}D_{jk}) + \Sigma_{\substack{k<l \\ k,l \neq i,j}} a_{kl}(D_{ijkl} - D_{ij}D_{kl}), \tag{44}$$

where again, the various special cases have been separated out. The expression can be further simplified using Equation 5 to reduce associations with repeated indices, as above. It is convenient to change the reference values at this point from the old allele frequencies to the new. Using Equation 15,

$$D''_{ij} = D'_{ij} - \Delta p_i \Delta p_j - \Delta p_i \Delta p_j + \Delta p_i \Delta p_j = D'_{ij} - \Delta p_i \Delta p_j. \tag{45}$$

Finally, the association at the start of the next generation is given by $D'''_{ij} = (1 - r_{ij})D''_{ij}$ for positions $\mathbb{i}$, $\mathbb{j}$ inherited from the same parent and 0 for positions inherited from different parents. Just as for the trait mean, this procedure again leads to a relatively simple expression for the genetic variance at the start of the next generation (*cf.* Equations 53 and 54 of BT91). However, this expression involves third- and fourth-order associations, and so approximations are required to obtain a closed set of equations.

Sexual selection by female choice: We mentioned earlier that the multilocus machinery developed here can be used to study the genetic consequences of nonrandom mating. This section shows how to calculate a QLE approximation for the genetic correlation (or covariance) between a female mating preference and male display trait that is generated by sexual selection. This quantity is important to many theories about sexual selection (Kirkpatrick and Ryan 1991). Previous work has calculated the covariance expected under autosomal inheritance in diploids (Lande 1981; Barton and Turelli 1991) and haploids (Kirkpatrick 1982; Kirkpatrick and Barton 1997). Here we extend the earlier results by finding the covariance when some genes are sex linked. In addition to the biological interest in the result, the derivations illustrate how nonrandom mating and nonautosomal inheritance are modeled in our framework.

Consider a pair of characters, one expressed in females and the other in males, that together affect the probability that a male and a female will mate. We refer to the first as a "preference" and the second as the "male

trait." In fact, the preference need not be a behavioral phenotype: It can be any character that affects mating probabilities. The value of a female's preference phenotype is denoted P and that of the male trait is T. A set of genes $\mathbb{P}$ affects the preference and a set $\mathbb{T}$ affects the male trait. We assume that these sets are disjoint (that is, no loci have pleiotropic effects on the preference and male trait) and that both sets act additively, so that the preference and male trait phenotype of an individual can be written in the form of Equation 36. We further assume that the parent from which a gene was inherited (its sex of origin) does not affect the gene's expression. Carrying allele 1 at position $\mathbb{i}$ rather than allele 0 increases the preference phenotype by b_i^P; the corresponding effect of allele 1 at position $\mathbb{j}$ on the male trait is b_j^T.

We begin by calculating the selection coefficients, which are independent of the inheritance rules. We then use them to find QLE approximations for the preference-trait covariance under three types of inheritance: haploid autosomal, diploid autosomal, and diploid X-linked.

Selection coefficients: Derivation of the selection coefficients follows Kirkpatrick and Barton (1997). The phenotypic distribution of the preference among females at birth, $f_P(\cdot)$, has mean $\overline{P}$ and variance σ_P^2; the corresponding distribution $f_T(\cdot)$ for the trait among males at birth has mean $\overline{T}$ and variance σ_T^2. The frequency of matings between a female with preference phenotype P and a male with trait phenotype T, denoted $M(P, T)$, has means $\overline{P}^*$ and $\overline{T}^*$, variances σ_P^{*2} and σ_T^{*2}, and correlation ρ_{PT}. We define the selection group as a mated pair of individuals. The relative fitness of a selection group is defined as its frequency after selection divided by its frequency before. The frequency after selection is given by M, while the "frequency" of a mated pair before pairing occurs is simply the frequency with which it would occur under random mating.

We now make two kinds of approximations. Assume that the preference and trait are not evolving rapidly, so that $\overline{P}^* \approx \overline{P}$, $\overline{T}^* \approx \overline{T}$, $\sigma_P^{*2} \approx \sigma_P^2$, and $\sigma_T^{*2} \approx \sigma_T^2$. Next, approximate f_P, f_T, and M with Gaussian densities. The relative fitness of a mated pair in which the female has preference P and the male has trait T is

$$w(P, T) = \frac{M(P, T)}{f_P(P)f_T(T)}$$

$$\approx \frac{1}{\sqrt{1 - \rho^2}}\exp\{[-\rho((P - \overline{P})^2\sigma_T^2\rho_{PT} - 2(P - \overline{P})(T - \overline{T})\sigma_t\sigma_T$$

$$+ (T - \overline{T})^2\sigma_P^2\rho_{PT})]/(2(1 - \rho_{PT}^2)\sigma_P^2\sigma_T^2)\}. \tag{46}$$

To calculate the selection coefficients, the fitness function must be expressed as a polynomial. Taking the first-order Taylor series of $w(P, T)$ around $\overline{P}$ and $\overline{T}$ gives

$$w(P, T) \approx \frac{1}{\sqrt{1 - \rho_{PT}^2}} + \frac{\rho_{PT}(P - \overline{P})(T - \overline{T})}{(1 - \rho_{PT}^2)^{3/2}\sigma_P\sigma_T}$$

$$\approx 1 + \rho_{PT}\frac{(P - \overline{P})(T - \overline{T})}{\sigma_P\sigma_T}. \tag{47}$$

The last step, which is not required for the QLE approximation, linearizes the fitness function in ρ_{PT}. The result is quite accurate for $\rho_{PT} < 0.4$.

Now substitute the expressions for P and T written in the form of Equation 36 into Equation 47. The selection coefficient for a set of loci $\mathbb{A}$ is then again given by the coefficient of $\zeta_{\mathbb{A}}$ in the fitness function. We find that the selection coefficient for a preference position $\mathbb{i}$ in a female and a trait position $\mathbb{j}$ in a male is

$$a_{i_f j_m} \approx \rho_{PT}\frac{b_{i_f}^P b_{j_m}^T}{\sigma_P\sigma_T}. \tag{48}$$

The context for each locus has only the sex of carrier (female for the preference locus i, male for the trait locus j). That is because of our assumption that the parent from which a gene was inherited (its sex of origin) does not affect its expression.

This result shows that the force of sexual selection that unites a female preference gene with a male trait gene is simply proportional to ρ_{PT}, the phenotypic correlation between the preference and trait among mated pairs. It is also proportional to the size of each gene's effect on the phenotype relative to the character's phenotypic standard deviation. Selection coefficients for all other sets of positions are 0.

These selection coefficients are valid regardless of how the genes affecting the preference and the male trait are inherited. In the following two sections we find the genetic correlation generated by this type of selection when the loci are haploid autosomal, diploid autosomal, and diploid X-linked. The examples show how the QLE approximation accommodates different modes of inheritance.

Haploid autosomal inheritance: We begin by calculating the genetic correlation in males between the female preference and male trait. The definition of the additive genetic correlation in zygotes is

$$r_{PT} = \frac{G_{PT}}{\sqrt{G_P G_T}}, \tag{49}$$

where G_{PT} is the genetic covariance between the preference and male trait, G_P is the additive genetic variance of the preference in females, and G_T the genetic variance of the trait in males. The genetic variances are

$$G_P = \Sigma_{i,j\in\mathbb{P}}b_i^P b_j^P D_{ij} = \Sigma_{i\in\mathbb{P}}(b_i^P)^2pq_i + O(a)$$

$$G_T = \Sigma_{i,j\in\mathbb{T}}b_i^T b_j^T D_{ij} = \Sigma_{i\in\mathbb{T}}(b_i^T)^2pq_i + O(a). \tag{50}$$

In haploids, a context needs to keep track of a gene's sex of carrier only, so the covariance is

$$G_{PT} = \Sigma_{i\in\mathbb{P}}\Sigma_{j\in\mathbb{T}}b_{i_f}^P b_{j_m}^T D_{i_m j_m}. \tag{51}$$

When the preference and trait loci are autosomal, the

covariance is the same in male and female zygotes. Now substitute the selection coefficients we just calculated (Equation 48) into the expression for the associations at QLE (Equation 20), which gives

$$\tilde{D}_{i_m j_m} = \left(\frac{t_{i_m j_m \leftarrow i_f j_m}}{r_{ij}} \right) a_{i_f j_m} pq_{ij} = \frac{1}{2} \rho_{\mathrm{PT}} \frac{b_{i_f} b_{j_m}}{\sigma_P \sigma_T} pq_{ij}. \quad (52)$$

Assembling these results shows that the genetic covariance and genetic correlation between a female mating preference and a male display trait in haploids,

$$G_{\mathrm{PT}} \approx \frac{1}{2} \rho_{\mathrm{PT}} h_{\mathrm{P}} h_{\mathrm{T}} \sqrt{G_{\mathrm{P}} G_{\mathrm{T}}}, \quad r_{\mathrm{PT}} \approx \frac{1}{2} \rho_{\mathrm{PT}} h_{\mathrm{P}} h_{\mathrm{T}}, \quad (53)$$

where h_{P} and h_{T} are the square roots of the heritabilities of the preference in females and the trait in males. This result agrees with KIRKPATRICK and BARTON (1997), who used the BT91 framework for their calculation.

Diploid autosomal inheritance: In diploids, there are two positions at the preference locus and two at the trait locus, corresponding to copies of those genes inherited from mothers and fathers. The genetic covariance between preference and trait in males is therefore

$$G_{\mathrm{PT}}^{\mathrm{m}} = \Sigma_{i \in \mathbb{P}} \Sigma_{j \in \mathbb{T}} b_{i_f}^{\mathrm{P}} b_{j_m}^{\mathrm{T}} (D_{i_{mf} j_{mf}} + D_{i_{mf} j_{mm}} + D_{i_{mm} j_{mf}} + D_{i_{mm} j_{mm}}). \quad (54)$$

The genetic variances are twice the values given by Equation 50 because diploid zygotes carry two haploid genomes. The covariance in female zygotes has the same value because the loci are autosomal.

To find those associations, we use Equations 21 and 22. The function $F(\cdot)$ that appears there is calculated using (23):

$$F(ij_{\mathrm{ff}}) = F(ij_{\mathrm{mf}}) = t_{i_{ff} j_{ff} \leftarrow i_f j_m} a_{i_f j_m} = \frac{1}{2} r_{ij}^{\mathrm{f}} a_{i_f j_m},$$

$$F(ij_{\mathrm{fm}}) = F(ij_{\mathrm{mm}}) = t_{i_{fm} j_m \leftarrow i_{mf} j_{mm}} a_{i_f j_m} = \frac{1}{2} r_{ij}^{\mathrm{m}} a_{i_f j_m}. \quad (55)$$

The QLE approximations for the associations are found by substituting the selection coefficient from Equation 48 into those expressions and then those results into (21) and (22), giving

$$\tilde{D}_{i_{mf} j_{mf}} = \tilde{D}_{i_{mm} j_{mm}} = \frac{1}{2} \rho_{\mathrm{PT}} \frac{b_{i_f}^{\mathrm{P}} b_{j_m}^{\mathrm{T}}}{\sigma_P \sigma_T} pq_{ij},$$

$$\tilde{D}_{i_{mf} j_{mm}} = \rho_{\mathrm{PT}} \frac{b_{i_f}^{\mathrm{P}} b_{j_m}^{\mathrm{T}}}{\sigma_P \sigma_T} pq_{ij}, \quad \tilde{D}_{i_{mm} j_{mf}} = 0. \quad (56)$$

Substituting these into (54) shows that the genetic covariance and genetic correlation are the same as those we found for the haploid case, Equations 53. This is a useful result, as it shows that several basic conclusions based on haploid models regarding the evolution of mating preferences (KIRKPATRICK and BARTON 1997) and reinforcement (KIRKPATRICK and SERVEDIO 1999) carry over to diploids. An interesting conclusion is that the

genetic correlation is independent of the recombination rates, even when they differ between the sexes.

X-linked inheritance: To illustrate how our methods extend to other forms of inheritance, consider next a case in which genetic variation in a female mating preference is X-linked while variation in the male trait is autosomal. We see that the genetic covariance and correlation in males are different than when both characters are autosomally inherited.

The preference-trait covariance in males is

$$G_{\mathrm{PT}}^{\mathrm{m}} = \Sigma_{i \in \mathbb{P}} \Sigma_{j \in \mathbb{T}} b_{i_f}^{\mathrm{P}} b_{j_m}^{\mathrm{T}} (D_{i_{mf} j_{mf}} + D_{i_{mf} j_{mm}}). \quad (57)$$

Because males do not inherit an X chromosome from their mothers, the terms $D_{i_{mm} j_{mf}}$ and $D_{i_{mm} j_{mm}}$ that appear in the diploid case (see Equation 54) do not exist. The genetic variance for the trait is given by twice the expressions in Equation 50 for haploids, because those loci are diploid and autosomal, while the variance for the male trait is given by (50), because males have only one haploid genome for X-linked loci.

The associations that appear in Equation 57 are given by Equations 21 and 24. The values for the function $F(\cdot)$ that appears in (24) are now

$$F(ij_{\mathrm{ff}}) = F(ij_{\mathrm{mf}}) = t_{i_{ff} j_{ff} \leftarrow i_{ff} j_m} a_{i_f j_m} = \frac{1}{4} a_{i_f j_m},$$

$$F(ij_{\mathrm{fm}}) = t_{i_{fm} j_{fm} \leftarrow i_{mf} j_{mm}} a_{i_f j_m} = \frac{1}{2} a_{i_f j_m},$$

$$F(ij_{\mathrm{mm}}) = t_{i_{mm} j_{mm} \leftarrow i_{mf} j_{mm}} a_{i_f j_m} = 0. \quad (58)$$

The QLE approximations for the associations in males are therefore

$$\tilde{D}_{i_{mf} j_{mf}} = \frac{3}{5} a_{i_f j_m} pq_{ij} = \frac{3}{5} \rho_{\mathrm{PT}} \frac{b_{i_f}^{\mathrm{P}} b_{j_f}^{\mathrm{T}}}{\sigma_P \sigma_T} pq_{ij},$$

$$\tilde{D}_{i_{mf} j_{mm}} = a_{i_f j_m} pq_{ij} = \rho_{\mathrm{PT}} \frac{b_{i_f}^{\mathrm{P}} b_{j_m}^{\mathrm{T}}}{\sigma_P \sigma_T} pq_{ij}. \quad (59)$$

Putting these facts together shows that in male zygotes the genetic covariance and correlation between the mating preference and male trait are

$$G_{\mathrm{PT}}^{\mathrm{m}} \approx \frac{4}{5} \rho_{\mathrm{PT}} h_{\mathrm{P}} h_{\mathrm{T}} \sqrt{G_{\mathrm{P}} G_{\mathrm{T}}}, \quad r_{\mathrm{PT}}^{\mathrm{m}} \approx \frac{4}{5} \rho_{\mathrm{PT}} h_{\mathrm{P}} h_{\mathrm{T}}. \quad (60)$$

An interesting conclusion is that the genetic covariance and correlation are 60% larger here than they are when the preference and male trait are both autosomally inherited (given by Equation 53). Thus the impact of indirect selection on a female mating preference depends on how the preference and trait are inherited.

DISCUSSION

We have set out a general notation that describes arbitrary modes of selection and genetic transmission. The key components are the representation of genotype

frequencies in terms of means and higher moments of the distribution of allelic states, of selection as a polynomial function of genotype, and of genetic transmission as the movement of genes between different contexts. The first two components are already well developed, particularly in models of additive quantitative traits. The main contribution of this article is to combine them with a generalized representation of transmission. The calculations can be automated, as described in APPENDIX D.

How does a general multilocus notation help us to better understand evolution? A set of equations for changes in genotype frequencies can be derived automatically for arbitrary models, but will in all but the simplest cases be impenetrably complicated. The value of an algebraic expression written in terms of multilocus moments or cumulants is that it allows one to identify and interpret the key processes responsible for evolutionary change. For example, Equations 30–35 show that migration builds up associations among loci in proportion to the product of the allele frequency differences. A second example comes from the analysis of sexual selection. The analysis shows that the genetic correlation between a female preference and a male trait is directly proportional to the phenotypic correlation between the preference and trait in mating pairs. Further, the genetic correlation depends on the way in which the preference and trait genes are inherited (Equations 53 and 60).

Defining a model in a standard notation can reveal similarities among apparently different mechanisms. For example, if the indirect selection on a modifier of recombination is expressed in terms of selection coefficients (a_U), it can be shown to depend on the effects of recombination on the mean and variance of log(fitness), regardless of the causes of fitness variation (BARTON 1995). An unambiguous notation may also clarify conceptual issues. We believe that models of group selection using our notation may clarify definitions of fitness and of "levels of selection."

Models of multiple loci are most fruitful when combined with appropriate approximations. The best developed is the QLE approximation, which assumes that processes such as epistasis and migration that generate associations among loci are weak, relative to those that break them down, such as recombination, segregation, and mutation. This leads to simple expressions for associations of all orders and is likely to be accurate for most sexually reproducing populations. Several workers have explored this approach, using different measures of association (linkage disequilibrium). Table 4 summarizes these measures and their corresponding versions of the QLE approximation. (For a more detailed treatment, see BÜRGER 2000, pp. 82 and 183–190.)

The different measures can be divided into two classes. Most measures are defined for each genotype, usually as a difference between its actual frequency and

TABLE 4

Alternative measures of association (linkage disequilibrium)

Symbol	Definition	Comments	Reference
D_U	$E_{\mathbf{X}}[\prod_{i \in U}(X_i - \wp_i)]$	Defined in this article	This article
$\Delta_U^{(S)}$	Equivalent to D_U	Defined by SLATKIN (1972)	SLATKIN (1972); NAGYLAKI (1993)
D_U or $\Delta \mathbf{X}_U^{(U)}$	$\sum_{V \subseteq U}(-1)^{|UV|}p_{U\backslash V}E[X_V]$	Two alleles at each locus; $X_i = 0$ or 1; $E[X_V]$ is the marginal frequency of gametes with "1" alleles at loci in V^a	SLATKIN (1972); NAGYLAKI (1993); CHRISTIANSEN (1999, Eq. 2.4.1)
$\theta_{\mathbf{X}}$	$f(\mathbf{X}) - \tilde{f}(\mathbf{X})$	Deviation of gamete frequency $f(\mathbf{X})$ from that at LE, $\tilde{f}(\mathbf{X})^b$	BÜRGER (2000, p. 83)
$\theta_{U}^{(U)}$	$f_U(\mathbf{X}_U) - \tilde{f}_U(\mathbf{X}_U)$	Deviation of marginal frequencies, restricted to the set of loci $U^{a,b}$	BÜRGER (2000, p. 82)
$\Theta_{\mathbf{X}}$	$\overline{1/W}\sum_{\mathbf{Y}}W(\mathbf{X},\mathbf{Y})\sum_{G \subseteq G'F_1}(f(\mathbf{X})\cdot f(\mathbf{Y}) - f(\mathbf{X}\mathbf{Y})_{G\setminus U}/f(\mathbf{X}_G,\mathbf{N}_1)))\}$	Here $W(\mathbf{X}, \mathbf{Y})$ is the fitness of a diploid with genomes $\{\mathbf{X}, \mathbf{Y}\}$ and G is the set of all loci; assumes random mating and no position effects; under selection and recombination $f(\mathbf{X}) = f(\mathbf{X})^{W(\mathbf{X})}/\overline{W} - \Theta_{\mathbf{X}}$	NAGYLAKI (1993); BÜRGER (2000, p. 56)
$\Theta_{\mathbf{X}_U}^{(U)}$	$\sum_{\mathbf{X}_{G\setminus U}}\Theta_{\mathbf{X}}$	Marginal version of $\Theta_{\mathbf{X}}$	BÜRGER (2000, p. 71)
$L(\mathbf{X}_U)$	$L'(\mathbf{X}_U) = L(\mathbf{X}_U)(1 - r_U)$	Principal component; decays geometrically for every set of loci	BENNETT (1954); DAWSON (2000)
$z_{[i,j,k,l]}$	$\mathcal{D}_{[i,j,k,l]} - \mathcal{D}_{[i,j]}\mathcal{D}_{[k,l]}$	Decays geometrically under mutation and weak recombination; depends on gene order	BAAKE (2001)
c_U or κ_U		Cumulant measure; corresponds to moment (association) D_U	BÜRGER (2000); TURELLI and BARTON (1994)

[a] $\mathbf{X}_S$ indicates the subset of allelic states for the set of loci S; corresponds to BÜRGER's (2000) multiindex i_S.
[b] Gametes are denoted here by the vector $\mathbf{X} = \{X_i, \ldots\}$, rather than by BÜRGER's (2000) "multiindex" of integers, $i = \{0, 1, 2, \ldots\}$, indicating which allele is present at each locus. Introduced as d by NAGYLAKI (1993).

the frequency expected at linkage equilibrium. In contrast, we define associations as moments of the distribution of allelic states; these moments are defined for each set of positions (D_i, D_{ijk}, etc.), and correspond in a natural way to a polynomial representation of selection. Moreover, they lead to a simple representation of transmission (Equation 12). When selection acts on an additive quantitative trait, the effects of selection are more elegantly represented in terms of the cumulants of the distribution of allelic effects, rather than the moments (Bürger 1991, 2000; Turelli and Barton 1994). (This is because the cumulants of the trait distribution are then simply sums over multilocus cumulants of the same order.) Since we do not deal only with additive traits in this article, we have used moments throughout. Moments and cumulants are equivalent to leading order when the moments are small, which holds under QLE. In any event, it is simple to transform between the two representations as necessary (*e.g.*, using the *Mathematica* packages).

The multilocus notations used by Christiansen (1999) and Bürger (2000) are closest to that used here. The main difference is that we deal with sets of genes in context, or *positions*, which allows us to avoid restrictive assumptions such as autosomal diploid inheritance, random mating, and equal transmission rates in males and females. The relation between the notations can be illustrated by comparing expressions for associations among loci at QLE. Christiansen's (1999) Equation 7.19 for the associations among a set of loci M in a gamete is

$$\tilde{D}_M \approx \frac{[\varepsilon \hat{C}_M] [\hat{\pi}_M^R(\varnothing) \ \hat{\pi}_M^R(M)]}{1 - 2R_M(0)}. \tag{61}$$

Here, $[\varepsilon \hat{C}_M]$ is a measure of epistasis among the loci in set M; under his assumption of no sex differences in selection, it is equivalent to our selection coefficients $a_{M_{ff}}$, $a_{M_{fm}}$, $a_{M_{mf}}$, and $a_{M_{mm}}$. The quantity $[\hat{\pi}_M^R(\varnothing) \ \hat{\pi}_M^R(M)]$ is the frequency of M gametes at linkage equilibrium, which is equal to our product of allele frequencies pq_M. Last, $R_M(\varnothing)$ is the chance that at meiosis, a gamete derives all the genes in the set M from the maternal genome, which is $(1 - r_M)/2$ in our notation. Under Christiansen's assumptions of autosomal inheritance, random mating, equal selection in males and females, and equal recombination rates in males and females, our Equation 22 gives

$$\tilde{D}_{M_{fx}} = \tilde{D}_{M_{mx}} = \frac{a_{M_{xy}} pq_M}{r_M}. \tag{62}$$

Thus the two approaches are consistent.

Bürger's (2000, p. 188) expression for cumulant associations at QLE in generation t is

$$c_{\mathbf{m}}(t) = \frac{\tilde{h}_{\mathbf{m},LE}(C, t)}{r_{\mathbf{m}}} + O(s^2). \tag{63}$$

Here, $\mathbf{m}$ is a vector containing the number of times that each allele is included in the association; for example, with positions $\{\mathbb{i}, \mathbb{j}, \mathbb{k}\}$, Bürger's $c_{\{1, 0, 2\}}$ corresponds to our D_{ikk}. (Note that Christiansen's 1999 notation is restricted to two alleles per locus, whereas Bürger's 2000 and ours allow multiple alleles.) The term $\tilde{h}_{\mathbf{m},LE}(C, t)$ is a composite measure of epistasis and allele frequencies. It is equal to the change due to selection at linkage equilibrium, which is $a_M pq_M$ in our notation. Because the corresponding cumulants and moments are equal to leading order when QLE applies, Equations 62 and 63 agree. Thus Bürger's and our approaches are consistent.

Although the QLE results developed in this article are consistent internally and with independent derivations, we have not rigorously shown that this quasi-equilibrium is unique or that the population will always converge to it when the assumptions are met. Nagylaki (1993; see also Nagylaki *et al.* 1999) showed that autosomal loci in a random mating diploid population under weak selection converge to a QLE. In this article we have relaxed his assumptions to allow for nonrandom mating, other forms of inheritance, and other evolutionary forces (migration and mutation). Since the effects of migration and mutation are equivalent to forms of frequency-dependent selection, Nagylaki's results should also apply when they act so long as the effective selection coefficients that they generate are small. The consequences of nonrandom mating and variations in inheritance are more difficult to account for. We expect convergence to our QLE values whenever all the a's are sufficiently small and all eigenvalues of the matrix of transmission coefficients $t_{A \leftarrow U}$ are of order 1, but that conjecture remains to be proven.

The QLE approximation developed here can be used to study a variety of interesting models. Selection on the genetic system (recombination, selfing, and mutation rate, for example) can be studied by assuming a modifier allele of small effect. Even if the system as a whole is under strong selection, associations involving the modifier will be weak and can therefore be modeled by a set of linear equations (Barton 1995). The infinitesimal model is based on a different kind of approximation, in which the limit of a large number of loci, each with infinitesimal effect, is taken. Turelli and Barton (1994) give a heuristic argument that extends this model to allow for linkage and epistasis. However, as Bürger (2000, p. 189) points out, this extension remains to be proven rigorously.

The biggest lacuna in the framework presented here is the lack of a model for random genetic drift. Its effects can be included by accounting for the variation in allele frequencies and associations caused by random sampling. The effects of drift on allele frequencies and pairwise associations have been well studied (Ewens 1979), but exact results and approximations for the joint probability distribution of the higher-order associations

remain to be worked out. A basic implementation of random drift is included in the *Mathematica* packages.

For the future, there is considerable scope for applying the methods set out in this article to bring together analyses of particular models and to explore better ways to approximate general multilocus systems.

LITERATURE CITED

Baake, E., 2001   Mutation and recombination with tight linkage. J. Math. Biol. **42:** 455–488.

Baird, S. J. E., 1995   The mixing of genotypes in hybrid zones: a simulation study of multilocus clines. Evolution **49:** 1038–1045.

Barton, N. H., 1983   Multilocus clines. Evolution **37:** 454–471.

Barton, N. H., 1986   The effects of linkage and density-dependent regulation on gene flow. Heredity **57:** 415–426.

Barton, N. H., 1992   On the spread of new gene combinations in the third phase of Wright's shifting balance. Evolution **46:** 551–557.

Barton, N. H., 1995   A general model for the evolution of recombination. Genet. Res. **65:** 123–144.

Barton, N. H., 2000   Estimating multilocus linkage disequilibria. Heredity **84:** 373–389.

Barton, N. H., and M. Shpak, 2000   The stability of symmetrical solutions to polygenic models. Theor. Popul. Biol. **57:** 249–264.

Barton, N. H., and M. Turelli, 1987   Adaptive landscapes, genetic distance, and the evolution of quantitative characters. Genet. Res. **49:** 157–174.

Barton, N. H., and M. Turelli, 1991   Natural and sexual selection on many loci. Genetics **127:** 229–255.

Bennett, J. H., 1954   On the theory of random mating. Ann. Eugen. **18:** 311–317.

Bulmer, M. G., 1980   *The Mathematical Theory of Quantitative Genetics.* Oxford University Press, Oxford.

Burger, R., 1991   Moments, cumulants and polygenic dynamics. J. Math. Biol. **30:** 199–213.

Burger, R., 2000   *The Mathematical Theory of Selection, Recombination and Mutation.* Wiley, Chichester, UK.

Christiansen, F. B., 1987   The deviation from linkage equilibrium with multiple loci varying in a stepping-stone cline. J. Genet. **40:** 45–67.

Christiansen, F. B., 1999   *Population Genetics of Multiple Loci.* John Wiley & Sons, New York.

Dawson, K. J., 1999   The dynamics of infinitesimally rare alleles, applied to the evolution of mutation rates and the expression of deleterious mutations. Theor. Popul. Biol. **55:** 1–22.

Dawson, K. J., 2000   The decay of linkage disequilibrium under random union of gametes: how to calculate Bennett's principal components. Theor. Popul. Biol. **58:** 1–20.

Doebeli, M., 1996   A quantitative genetic model for sympatric speciation. J. Evol. Biol. **9:** 893–909.

Ewens, W. J., 1979   *Mathematical Population Genetics.* Springer Verlag, Berlin.

Fisher, R. A., 1918   The correlation between relatives on the supposition of Mendelian inheritance. Proc. R. Soc. Edinb. **52:** 399–433.

Fisher, R. A., 1952   *The Genetical Theory of Natural Selection*, Ed. 2. Dover, New York.

Fisher, R. A., 1953   A fuller theory of 'junctions' in inbreeding. Heredity **8:** 187–198.

Gillespie, J. H., 1981   Mutation modification in a random environment. Evolution **35:** 468–476.

Haldane, J. B. S., 1927   A mathematical theory of natural and artificial selection. V. Selection and mutation. Proc. Camb. Philos. Soc. **26:** 220–230.

Hill, W. G., and A. Robertson, 1966   The effect of linkage on limits to artificial selection. Genet. Res. **8:** 269–294.

Hudson, R., 1990   Gene genealogies and the coalescent process. Oxf. Surv. Evol. Biol. **7:** 1–44.

Kimura, M., 1965   Attainment of quasilinkage equilibrium when gene frequencies are changing by natural selection. Genetics **52:** 875–890.

Kirkpatrick, M., 1982   Sexual selection and the evolution of female choice. Evolution **36:** 1–12.

Kirkpatrick, M., and N. H. Barton, 1997   The strength of indirect selection on female mating preferences. Proc. Natl. Acad. Sci. USA **94:** 1282–1286.

Kirkpatrick, M., and M. J. Ryan, 1991   The evolution of mating preferences and the paradox of the lek. Nature **350:** 33–38.

Kirkpatrick, M., and M. Servedio, 1999   The reinforcement of mating preferences on an island. Genetics **151:** 865–884.

Kondrashov, A. S., 1984   On the intensity of selection for reproductive isolation at the beginnings of sympatric speciation. Genetika **20:** 408–415.

Lande, R., 1981   Models of speciation by sexual selection on polygenic traits. Proc. Natl. Acad. Sci. USA **78:** 3721–3725.

Nagylaki, T., 1993   The evolution of multilocus systems under weak selection. Genetics **134:** 627–647.

Nagylaki, T., J. Hoffbauer and P. Brunovsky, 1999   Convergence of multilocus systems under weak epistasis or weak selection. J. Math. Biol. **38:** 103–133.

Otto, S. P., and Y. Michalakis, 1998   The evolution of recombination in changing environments. Trends Ecol. Evol. **13:** 145–151.

Price, G. R., 1970   Selection and covariance. Nature **227:** 520–521.

Shapiro, J. L., A. Prügel-Bennett and M. Rattray, 1994   A statistical mechanics formulation of the dynamics of genetic algorithms. Lect. Notes Comput. Sci. **865:** 17–27.

Shpak, M., and A. S. Kondrashov, 1999   Applicability of the hypergeometric phenotypic model to haploid and diploid populations. Evolution **53:** 600–604.

Slatkin, M., 1972   On treating the chromosome as the unit of selection. Genetics **72:** 157–168.

Sniegowski, P., P. J. Gerrish and T. Johnson, 2000   Evolution of mutation rates: separating causes from consequences. BioEssays **22:** 1057–1066.

Turelli, M., and N. H. Barton, 1990   Dynamics of polygenic characters under selection. Theor. Popul. Biol. **38:** 1–57.

Turelli, M., and N. H. Barton, 1994   Genetic and statistical analyses of strong selection on polygenic traits: What, me normal? Genetics **138:** 913–941.

Wolfram, S., 1999   *The Mathematica Book*, Ed. 4. Wolfram Media/ Cambridge University Press, Cambridge, UK.

Wright, S., 1931   Evolution in Mendelian populations. Genetics **16:** 97–159.

Wright, S., 1935   Evolution in populations in approximate equilibrium. J. Genet. **30:** 257–266.

Communicating editor: J. B. Walsh

APPENDIX A: A TWO-LOCUS EXAMPLE

The goal is to develop a simple example that illustrates the notation and calculations involved in selection, transmission, and the QLE approximation. We look at the effects of viability selection acting on two diploid autosomal loci in a random mating hermaphroditic population.

Consider a simple selection scheme in which allele 1 at locus i changes relative fitness by s_i and allele 1 at locus j by s_j. Interactions between the loci also affect fitness. More specifically, assume that each interaction between an allele 1 at locus i and an allele 1 at locus j changes relative fitness by e_{ij}. These assumptions lead to the fitnesses for diploid genotypes shown in Table A1.

Because mating is random, we can define the selection group as a single individual. Since the population is

TABLE A1

Fitnesses for the two-locus example

Paternal haplotype	Maternal haplotype			
	0 0	0 1	1 0	1 1
0 0	1	$1 + s_j$	$1 + s_i$	$1 + s_i + s_j + e_{ij}$
0 1	$1 + s_j$	$1 + 2s_j$	$1 + s_i + s_j + e_{ij}$	$1 + s_i + 2s_j + 2e_{ij}$
1 0	$1 + s_i$	$1 + s_i + s_j + e_{ij}$	$1 + 2s_i$	$1 + 2s_i + s_j + 2e_{ij}$
1 1	$1 + s_i + s_j + e_{ij}$	$1 + s_i + 2s_j + 2e_{ij}$	$1 + 2s_i + s_j + 2e_{ij}$	$1 + 2s_i + 2s_j + 4e_{ij}$

The haplotypes indicate the alleles carried at loci i and j.

hermaphrodite, the contexts for the two loci need only to give the sex of origin for each gene. For example, i_m stands for a gene that was inherited via a sperm.

To calculate the selection coefficients (the a's), first write the fitnesses that appear in Table A1 in the form of a polynomial in the X's:

$$W(X_{i_\mathrm{f}}, X_{j_\mathrm{f}}, X_{i_\mathrm{m}}, X_{j_\mathrm{m}}) = 1 + s_i(X_{i_\mathrm{f}} + X_{i_\mathrm{m}}) + s_j(X_{j_\mathrm{f}} + X_{j_\mathrm{m}})$$
$$+ e_{ij}(X_{i_\mathrm{f}}X_{j_\mathrm{f}} + X_{i_\mathrm{f}}X_{j_\mathrm{m}} + X_{i_\mathrm{m}}X_{j_\mathrm{f}} + X_{i_\mathrm{m}}X_{j_\mathrm{m}}).$$
(A1)

Next, write out the fitness function that defines the selection coefficients, Equation 7:

$$W(X_{i_\mathrm{f}}, X_{j_\mathrm{f}}, X_{i_\mathrm{m}}, X_{j_\mathrm{m}}) = \overline{W}\{1 + a_{i_\mathrm{f}}[(X_{i_\mathrm{f}} - p_i) - D_{i_\mathrm{f}}] + a_{i_\mathrm{m}}[(X_{i_\mathrm{m}} - p_i) - D_{i_\mathrm{m}}]$$

$$+ a_{j_\mathrm{f}}[(X_{j_\mathrm{f}} - p_j) - D_{j_\mathrm{f}}] + a_{j_\mathrm{m}}[(X_{j_\mathrm{m}} - p_j) - D_{j_\mathrm{m}}]$$

$$+ a_{i_\mathrm{f}j_\mathrm{f}}[(X_{i_\mathrm{f}} - p_i)(X_{j_\mathrm{f}} - p_j) - D_{i_\mathrm{f}j_\mathrm{f}}]$$

$$+ a_{i_\mathrm{f}j_\mathrm{m}}[(X_{i_\mathrm{f}} - p_i)(X_{j_\mathrm{m}} - p_j) - D_{i_\mathrm{f}j_\mathrm{m}}]$$

$$+ a_{i_\mathrm{m}j_\mathrm{f}}[(X_{i_\mathrm{m}} - p_i)(X_{j_\mathrm{f}} - p_j) - D_{i_\mathrm{m}j_\mathrm{f}}]$$

$$+ a_{i_\mathrm{m}j_\mathrm{m}}[(X_{i_\mathrm{m}} - p_i)(X_{j_\mathrm{m}} - p_j) - D_{i_\mathrm{m}j_\mathrm{m}}]$$

$$+ a_{i_\mathrm{f}i_\mathrm{m}}[(X_{i_\mathrm{f}} - p_i)(X_{i_\mathrm{m}} - p_i) - D_{i_\mathrm{f}i_\mathrm{m}}]$$

$$+ a_{j_\mathrm{f}j_\mathrm{m}}[(X_{j_\mathrm{f}} - p_j)(X_{j_\mathrm{m}} - p_j) - D_{j_\mathrm{f}j_\mathrm{m}}]$$

$$+ a_{i_\mathrm{f}i_\mathrm{m}j_\mathrm{f}}[(X_{i_\mathrm{f}} - p_i)(X_{i_\mathrm{m}} - p_i)(X_{j_\mathrm{f}} - p_j) - D_{i_\mathrm{f}i_\mathrm{m}j_\mathrm{f}}]$$

$$+ a_{i_\mathrm{f}i_\mathrm{m}j_\mathrm{m}}[(X_{i_\mathrm{f}} - p_i)(X_{i_\mathrm{m}} - p_i)(X_{j_\mathrm{m}} - p_j) - D_{i_\mathrm{f}i_\mathrm{m}j_\mathrm{m}}]$$

$$+ a_{i_\mathrm{f}j_\mathrm{f}j_\mathrm{m}}[(X_{i_\mathrm{f}} - p_i)(X_{j_\mathrm{f}} - p_j)(X_{j_\mathrm{m}} - p_j) - D_{i_\mathrm{f}j_\mathrm{f}j_\mathrm{m}}]$$

$$+ a_{i_\mathrm{m}j_\mathrm{f}j_\mathrm{m}}[(X_{i_\mathrm{m}} - p_i)(X_{j_\mathrm{f}} - p_j)(X_{j_\mathrm{m}} - p_j) - D_{i_\mathrm{m}j_\mathrm{f}j_\mathrm{m}}]$$

$$+ a_{i_\mathrm{f}j_\mathrm{m}j_\mathrm{f}j_\mathrm{m}}[(X_{i_\mathrm{f}} - p_i)(X_{i_\mathrm{m}} - p_i)(X_{j_\mathrm{f}} - p_j)$$

$$\times (X_{j_\mathrm{m}} - p_j) - D_{i_\mathrm{f}i_\mathrm{m}j_\mathrm{f}j_\mathrm{m}}]\}.$$
(A2)

There are 15 selection coefficients, as required with four positions with two alleles each ($15 = 2^4 - 1$). Equating (A1) and (A2) and identifying coefficients with the same combinations of X's gives the selection coefficients

$$a_{i_\mathrm{f}} = a_{i_\mathrm{m}} = \frac{s_i}{\overline{W}},$$

$$a_{j_\mathrm{f}} = a_{j_\mathrm{m}} = \frac{s_j}{\overline{W}},$$

$$a_{i_\mathrm{f}j_\mathrm{f}} = a_{i_\mathrm{m}j_\mathrm{m}} = a_{i_\mathrm{f}j_\mathrm{m}} = a_{i_\mathrm{m}j_\mathrm{f}} = \frac{e_{ij}}{\overline{W}}.$$
(A3)

The other 8 selection coefficients that appear in (A2) are 0.

The change in the frequency of allele 1 at locus i is calculated using Equation 10. Because the population is hermaphroditic, allele frequencies and associations in genes inherited via sperm will be equal to those inherited via eggs: $p_{i_\mathrm{f}} = p_{i_\mathrm{m}} \equiv p_i$ and $D_{i_\mathrm{f}j_\mathrm{f}} = D_{i_\mathrm{m}j_\mathrm{m}} \equiv D_{ij}$. After one generation of random mating, associations involving more than one sex of origin (*e.g.*, $D_{i_\mathrm{f}j_\mathrm{m}}$) are zero. The exact change in allele frequency will then be

$$\Delta p_i = \Delta p_{i_\mathrm{f}}$$

$$= t_{i_\mathrm{f} \leftarrow i_\mathrm{f}}\{a_{i_\mathrm{f}}pq_i + a_{j_\mathrm{f}}D_{ij} + a_{i_\mathrm{f}j_\mathrm{f}}D_{iij}\}$$

$$+ t_{i_\mathrm{f} \leftarrow i_\mathrm{m}}\{a_{i_\mathrm{m}}pq_i + a_{j_\mathrm{m}}D_{ij} + a_{i_\mathrm{m}j_\mathrm{m}}D_{iij}\}$$

$$= \frac{1}{2}\left\{\frac{s_i}{\overline{W}}pq_i + \frac{s_j}{\overline{W}}D_{ij} + \frac{e_{ij}}{\overline{W}}D_{iij}\right\}$$

$$+ \frac{1}{2}\left\{\frac{s_i}{\overline{W}}pq_i + \frac{s_j}{\overline{W}}D_{ij} + \frac{e_{ij}}{\overline{W}}D_{iij}\right\}$$

$$= \frac{1}{\overline{W}}[s_i pq_i + s_j D_{ij} + e_{ij}(1 - 2p_i)D_{ij}].$$
(A4)

That result can be written entirely in terms of the selection coefficients and allele frequencies once the population reaches QLE. At that point, Equation 22 shows that the association between the loci is

$$D_{ij} = \left(\frac{pq_{ij}}{2r_{ij}}\right)[a_{i_\mathrm{f}j_\mathrm{f}}(1 + r_{ij}) + a_{i_\mathrm{m}j_\mathrm{m}}(1 + r_{ij})] + O(a^2)$$

$$= \frac{e_{ij}pq_{ij}}{r_{ij}} + O(a^2),$$
(A5)

where by $O(a^2)$ we mean terms that are no larger than a constant of order 1 times the largest of the a's given above in Equation A3. At QLE, the mean fitness is to

$O(a)$ equal to the mean fitness of a population with no associations (at linkage equilibrium):

$$\overline{W} = 1 + 2s_i p_i + 2s_j p_j + 4e_{ij} p_i p_j + O(a^2). \quad \text{(A6)}$$

Putting together these last three results shows that the change in frequency of allele 1 at locus i in a population at QLE is

$$\Delta p_i = \frac{1}{\overline{W}}\left[s_i p q_i + s_j \frac{e_{ij} p q_{ij}}{r_{ij}} + (1 - 2p_i)\frac{e_{ij}^2 p q_{ij}}{r_{ij}} \right] + O(a^3)$$

$$= s_i p q_i (1 - 2s_i p_i - 2s_j p_j - 4e_{ij} p_i p_j)$$

$$+ \frac{e_{ij} p q_{ij}}{r_{ij}}[s_j + (1 - 2p_i)e_{ij}] + O(a^3). \quad \text{(A7)}$$

APPENDIX B: MUTATION

The centered associations following mutation can be written as

$$D''_{\mathbb{A}} = E_{X,M}[\Pi_{i\in\mathbb{A}}((X_i + \delta X_i) - (p_i + \Delta_\mu p_i))], \quad \text{(B1)}$$

where δX_i is the random change to allelic state at position i caused by mutation. The expectation $E_{X,M}[\ ]$ is over random mutational events, M, as well as over allelic states, X. $\Delta_\mu p_i$ is the change in the frequency of allele 1 at position i. Denoting the mutation rate at position i from allele 0 to allele 1 as u_i and the reverse mutation rate from 1 to 0 as v_i, that change is

$$\Delta_\mu p_i = u_i q_i - v_i p_i. \quad \text{(B2)}$$

Equation B1 can be simplified by defining the random variable $R[x]$ as having value 1 with probability x and zero otherwise. Let $\hat{R}[x] \equiv R[x] - x$. If $X_i = 0$, then $\delta X_i = +R[u_i]$, and if $X_i = 1$, then $\delta X_i = -R[v_i]$. Hence, $\delta X_i = (1 - X_i)R[u_i] - X_i R[v_i]$. Rewriting this in terms of $\hat{R}$ and $\zeta_i = X_i - p_i$,

$$\delta X_i = (q_i - \zeta_i)(u_i + \hat{R}(u_i)) - (p_i + \zeta_i)(v_i + \hat{R}(v_i)). \quad \text{(B3)}$$

Substituting into Equation B1,

$$D''_{\mathbb{A}} = E_{X,M}[\Pi_{i\in\mathbb{A}}(p_i + \zeta_i + (q_i - \zeta_i)(u_i + \hat{R}(u_i))$$

$$- (p_i + \zeta_i)(v_i + \hat{R}(v_i)) - p_i - \Delta_\mu p_i]$$

$$= E_{X,M}[\Pi_{i\in\mathbb{A}}(q_i \hat{R}(u_i) - p_i \hat{R}(v_i)$$

$$+ \zeta_i(1 - u_i - \hat{R}(u_i) - v_i - \hat{R}(v_i)))]. \quad \text{(B4)}$$

Now, the $\hat{R}$ are independent across loci and are independent of allelic state and so vanish when the expectation over mutational events is taken. This leads immediately to Equation 28.

APPENDIX C: MIGRATION

We calculate the centered associations following a migration event. The result is most easy to interpret when it is expressed in terms of the centered association among the residents and migrants before the event. We therefore define the reference values in the resident population, $\wp_i^R$, to equal the allele frequencies in that population before migration, p_i^R.

Just after migration, the (uncentered) associations are given by Equation 3,

$$D'_{\mathbb{A}} = E_X[\Pi_{i\in\mathbb{A}}(X'_i - p_i^R)], \quad \text{(C1)}$$

where X'_i is the allelic state of an individual in the population after migration. That expression can be rewritten in terms of X_i^M, the allelic states of the migrants, and X_i^R, the allelic states of the residents, as

$$D'_{\mathbb{A}} = (1 - m)E_X[\Pi_{i\in\mathbb{A}}(X_i^R - p_i^R)] + mE_X[\Pi_{i\in\mathbb{A}}(X_i^M - p_i^R)]$$

$$= (1 - m)D_{\mathbb{A}}^R + mD_{\mathbb{A}}^{MR}, \quad \text{(C2)}$$

where m is the migration rate. The $D_{\mathbb{A}}^R$ are the (centered) associations among the residents before migration, while $D_{\mathbb{A}}^{MR}$ are the uncentered associations of the migrants measured relative to the resident reference values, $\wp_i^R = p_i^R$.

The next step is to express the $D_{\mathbb{A}}^{MR}$ in terms of the centered associations in the migrant population, $D_{\mathbb{A}}^M$. To do that we need to rescale those associations in terms of the reference values for the migrant population, $\wp_i^M$. Using Equation 15 for changing reference values and some algebra gives

$$D_{\mathbb{A}}^{MR} = \Sigma_{\mathbb{U}\subseteq\mathbb{A}} d_{\mathbb{U}} D_{\mathbb{A}\setminus\mathbb{U}}^M, \quad \text{(C3)}$$

where $d_{\mathbb{U}}$ was defined earlier in Equation 31. The sum includes the term in which $\mathbb{U}$ equals the empty set $\varnothing$, which generates the term $D_{\mathbb{A}}^M$. Substituting this expression for $D_{\mathbb{A}}^{MR}$ in Equation C2 produces

$$D'_{\mathbb{A}} = (1 - m)D_{\mathbb{A}}^R + m\Sigma_{\mathbb{U}\subseteq\mathbb{A}} d_{\mathbb{U}} D_{\mathbb{A}\setminus\mathbb{U}}^M. \quad \text{(C4)}$$

These associations $D'_{\mathbb{A}}$ are based on the reference values for the residents before migration, $\wp_i^R = p_i^R$.

To find the central moments after migration, we set the new reference points to the allele frequencies after migration: $\wp_i'' = (1 - m)p_i^R + mp_i^M$. Using Equation 15 as before gives Equation 30, above.

APPENDIX D: *MATHEMATICA* EXAMPLES

Here, we give some examples that outline how *Mathematica* (WOLFRAM 1999) can be used to find algebraic expressions for the changes in associations due to various evolutionary processes. Software that extends *Mathematica* to do the calculations described in this article is available on the Web at http://helios.bto.ed.ac.uk/evolgen. The notation used by that software is essentially the same as that in the text, but there are a few differences that need to be explained. The examples below show the user's input to *Mathematica* in boldface type and the program's output in regular type.

Stages in the life cycle: Different stages in the life

cycle are denoted by the first element in the context, rather than by primes. Thus, the association between two positions in male gametes would be written $\mathfrak{D}_{\{i\{\mathfrak{G},\mathfrak{m}\},j\{\mathfrak{G},\mathfrak{m}\}\}}$, where $\mathfrak{G}$ denotes the gamete stage. (Special German font is used to denote contexts to avoid clashes with other symbols.) Similarly, the association between two positions in a male diploid zygote, one inherited from the mother and one from the father, would be written $\mathfrak{D}_{\{i\{\mathfrak{Z},\mathfrak{m},\mathfrak{m}\},j\{\mathfrak{Z},\mathfrak{m},\mathfrak{f}\}\}}$.

The first step is to define the contexts for each stage. For example, this defines gamete ($\mathfrak{G}$), zygote ($\mathfrak{Z}$), adult ($\mathfrak{A}$), and new gamete ($\mathfrak{G}*$) stages:

DefineContext [$\mathfrak{G}$, {Sex}];
DefineContext [$\mathfrak{Z}$, {Sex, SexOfOrigin}];
DefineContext [$\mathfrak{A}$, {SexOfOrigin}];
DefineContext [$\mathfrak{G}*$, {Sex}];

Recombination: Expressions for the mean and for the associations at one stage are expressed in terms of variables at the previous stage by applying various rules. For example, the mean contribution of a gene in the new gamete pool is

$$\mathfrak{m}_{j\{\mathfrak{G}*,\mathfrak{m}\}}\,/\!\!/.\text{Recombination[ ]}$$

$$\mathfrak{m}_{j\{\mathfrak{A},\mathfrak{m},\mathfrak{m}\}}\,\mathfrak{r}_{\{\},\{j\}} + \mathfrak{m}_{j\{\mathfrak{A},\mathfrak{m},\mathfrak{f}\}}\,\mathfrak{r}_{\{j\},\{\}}.$$

Here, $/\!\!/.$ denotes the application of a transformation rule, and $\mathfrak{r}_{\{\},\{j\}}$ is a generalized recombination rate, which gives the chance that the gene j was inherited from the father at meiosis. Further rules apply the usual assumptions of Mendelian inheritance,

$$\mathfrak{m}_{j\{\mathfrak{G}*,\mathfrak{m}\}}\,/\!\!/.\text{Recombination[ ]} \vee \text{SymmetricRecombination[ ]}$$

$$\frac{1}{2}\mathfrak{m}_{j\{\mathfrak{A},\mathfrak{m},\mathfrak{f}\}} + \frac{1}{2}\mathfrak{m}_{j\{\mathfrak{A},\mathfrak{m},\mathfrak{m}\}},$$

where $\text{rule}_1 \vee \text{rule}_2$ denotes a collection of rules. When applied to an association, the result depends on the change in the reference values, $\wp$, defined for each stage:

$$\mathfrak{D}_{\{i\{\mathfrak{G}*,\mathfrak{m}\},j\{\mathfrak{G}*,\mathfrak{m}\}\}}\,/\!\!/.\text{Recombination[ ]} \vee \text{SymmetricRecombination[ ]}$$

$$/\!\!/\text{Simplify}$$

$$\mathfrak{r}_{\{\},\{i,j\}}(\mathfrak{D}_{\{i\{\mathfrak{A},\mathfrak{m},\mathfrak{f}\},j\{\mathfrak{A},\mathfrak{m},\mathfrak{f}\}\}} + \mathfrak{D}_{\{j\{\mathfrak{A},\mathfrak{m},\mathfrak{f}\}\}}(-\mathfrak{P}_{i\{\mathfrak{G}*,\mathfrak{m}\}} + \mathfrak{P}_{i\{\mathfrak{A},\mathfrak{m},\mathfrak{f}\}})$$

$$- (\mathfrak{D}_{\{i\{\mathfrak{A},\mathfrak{m},\mathfrak{f}\}\}} - \mathfrak{P}_{i\{\mathfrak{G}*,\mathfrak{m}\}} + \mathfrak{P}_{i\{\mathfrak{A},\mathfrak{m},\mathfrak{f}\}})(\mathfrak{P}_{j\{\mathfrak{G}*,\mathfrak{m}\}} - \mathfrak{P}_{j\{\mathfrak{A},\mathfrak{m},\mathfrak{f}\}}))$$

$$+ \mathfrak{r}_{\{i\},\{j\}}(\mathfrak{D}_{\{i\{\mathfrak{A},\mathfrak{m},\mathfrak{m}\},j\{\mathfrak{A},\mathfrak{m},\mathfrak{f}\}\}} + \mathfrak{D}_{\{j\{\mathfrak{A},\mathfrak{m},\mathfrak{f}\}\}}(-\mathfrak{P}_{i\{\mathfrak{G}*,\mathfrak{m}\}} + \mathfrak{P}_{i\{\mathfrak{A},\mathfrak{m},\mathfrak{m}\}})$$

$$- (\mathfrak{D}_{\{i\{\mathfrak{A},\mathfrak{m},\mathfrak{m}\}\}} - \mathfrak{P}_{i\{\mathfrak{G}*,\mathfrak{m}\}} + \mathfrak{P}_{i\{\mathfrak{A},\mathfrak{m},\mathfrak{f}\}})\mathfrak{P}_{j\{\mathfrak{G}*,\mathfrak{m}\}} - \mathfrak{P}_{j\{\mathfrak{A},\mathfrak{m},\mathfrak{f}\}}))$$

$$+ \mathfrak{r}_{\{i\},\{j\}}(\mathfrak{D}_{\{i\{\mathfrak{A},\mathfrak{m},\mathfrak{f}\},j\{\mathfrak{A},\mathfrak{m},\mathfrak{m}\}\}} + \mathfrak{D}_{\{j\{\mathfrak{A},\mathfrak{m},\mathfrak{m}\}\}}(-\mathfrak{P}_{i\{\mathfrak{G}*,\mathfrak{m}\}} + \mathfrak{P}_{i\{\mathfrak{A},\mathfrak{m},\mathfrak{f}\}})$$

$$- (\mathfrak{D}_{\{i\{\mathfrak{A},\mathfrak{m},\mathfrak{f}\}\}} - \mathfrak{P}_{i\{\mathfrak{G}*,\mathfrak{m}\}} + \mathfrak{P}_{i\{\mathfrak{A},\mathfrak{m},\mathfrak{f}\}})(\mathfrak{P}_{j\{\mathfrak{G}*,\mathfrak{m}\}} - \mathfrak{P}_{j\{\mathfrak{A},\mathfrak{m},\mathfrak{m}\}}))$$

$$+ \mathfrak{r}_{\{\}\,\{i,j\}}(\mathfrak{D}_{\{i\{\mathfrak{A},\mathfrak{m},\mathfrak{m}\},j\{\mathfrak{A},\mathfrak{m},\mathfrak{m}\}\}} + \mathfrak{D}_{\{j\{\mathfrak{A},\mathfrak{m},\mathfrak{m}\}\}}(-\mathfrak{P}_{i\{\mathfrak{G}*,\mathfrak{m}\}} + \mathfrak{P}_{i\{\mathfrak{A},\mathfrak{m},\mathfrak{m}\}})$$

$$- (\mathfrak{D}_{\{i\{\mathfrak{A},\mathfrak{m},\mathfrak{m}\}\}} - \mathfrak{P}_{i\{\mathfrak{G}*,\mathfrak{m}\}} + \mathfrak{P}_{i\{\mathfrak{A},\mathfrak{m},\mathfrak{m}\}})(\mathfrak{P}_{j\{\mathfrak{G}*,\mathfrak{m}\}} - \mathfrak{P}_{j\{\mathfrak{A},\mathfrak{m},\mathfrak{m}\}})).$$

Such expressions are simplified by defining the reference values (see below).

Stabilizing selection: Stabilizing selection is represented by defining fitness as a function of genotype. It is convenient to define the trait and fitness separately for each sex, even though they are in fact the same. The expressions can be simplified later:

$$z_{\mathfrak{m}} = \Sigma_{i\in\{j,k\}}b_i(X_{i_{\{\mathfrak{Z},\mathfrak{m},\mathfrak{m}\}}} + X_{i_{\{\mathfrak{Z},\mathfrak{m},\mathfrak{f}\}}} - 1);$$

$$z_{\mathfrak{f}} = \Sigma_{i\in\{j,k\}}b_i(X_{i_{\{\mathfrak{Z},\mathfrak{f},\mathfrak{m}\}}} + X_{i_{\{\mathfrak{Z},\mathfrak{f},\mathfrak{f}\}}} - 1);$$

$$W_{\mathfrak{m}} = 1 - \frac{s}{2}Z_{\mathfrak{m}}^2; \quad W_{\mathfrak{f}} = 1 - \frac{s}{2}Z_{\mathfrak{f}}^2.$$

Fitness depends on two loci, j and k. This defines the set of positions that influence fitness of males and females:

$$\mathbb{W}_{\mathfrak{m}} = \{_{j\{\mathfrak{Z},\mathfrak{m},\mathfrak{m}\},\ j\{\mathfrak{Z},\mathfrak{m},\mathfrak{f}\},\ k\{\mathfrak{Z},\mathfrak{m},\mathfrak{m}\},\ k\{\mathfrak{Z},\mathfrak{m},\mathfrak{f}\}}\};$$

$$\mathbb{W}_{\mathfrak{f}} = \{_{j\{\mathfrak{Z},\mathfrak{f},\mathfrak{m}\},\ j\{\mathfrak{Z},\mathfrak{f},\mathfrak{f}\},\ k\{\mathfrak{Z},\mathfrak{f},\mathfrak{m}\},\ k\{\mathfrak{Z},\mathfrak{f},\mathfrak{f}\}}\}.$$

The fitness is now an explicit polynomial function of the states of genes in zygotes:

$$W_{\mathfrak{m}} = 1 - \frac{1}{2}s((-1 + X_{j_{\{\mathfrak{Z},\mathfrak{m},\mathfrak{f}\}}} + X_{j_{\{\mathfrak{Z},\mathfrak{m},\mathfrak{m}\}}})\,b_j$$

$$+ (-1 + X_{k_{\{\mathfrak{Z},\mathfrak{m},\mathfrak{f}\}}} + X_{k_{\{\mathfrak{Z},\mathfrak{m},\mathfrak{m}\}}})\,b_k)^2.$$

Since we assume two alleles per locus, the fitness must be simplified to remove superfluous higher powers of the X's:

W2$_{\mathfrak{m}}$ = Expand[$W_{\mathfrak{m}}$] $/\!\!/.$Biallelic[ ];

 W2$_{\mathfrak{f}}$ = Expand [$W_{\mathfrak{f}}$] $/\!\!/.$Biallelic [ ]; W2$_{\mathfrak{m}}$ $/\!\!/$Simplify

$$\frac{1}{2}(2 + s(-1 + X_{j_{\{\mathfrak{Z},\mathfrak{m},\mathfrak{f}\}}}(1 - 2X_{j_{\{\mathfrak{Z},\mathfrak{m},\mathfrak{m}\}}}) + X_{j_{\{\mathfrak{Z},\mathfrak{m},\mathfrak{m}\}}})\,b_j^2$$

$$- 2s(-1 + X_{j_{\{\mathfrak{Z},\mathfrak{m},\mathfrak{f}\}}} + X_{j_{\{\mathfrak{Z},\mathfrak{m},\mathfrak{m}\}}})(-1 + X_{k_{\{\mathfrak{Z},\mathfrak{m},\mathfrak{f}\}}} + X_{k_{\{\mathfrak{Z},\mathfrak{m},\mathfrak{m}\}}})\,b_j b_k$$

$$+ s(-1 + X_{k_{\{\mathfrak{Z},\mathfrak{m},\mathfrak{f}\}}}(1 - 2X_{k_{\{\mathfrak{Z},\mathfrak{m},\mathfrak{m}\}}}) + X_{k_{\{\mathfrak{Z},\mathfrak{m},\mathfrak{m}\}}})\,b_k^2).$$

The rule Biallelic[] implements the reduction formula that led to Equation 5.

The effect of selection on the mean contribution of a position is

$$\mathfrak{m}_{i\{\mathfrak{A},\mathfrak{m},\mathfrak{f}\}}\ /.\text{Selection}[\{\text{Sex}\}]/.$$

$$\{\alpha_{\text{sex\_,U\_}}:\rightarrow (\text{MakeSelectionCoefficient}\,[\mathbb{U},\,W2_{\text{sex}}]/\text{wb})\}/.$$

$$\wp_{i_{\{\mathfrak{Z},\_\_\}}} :> \rho_i/.\text{ReferencePointZero}\,[\mathfrak{Z}]\,/\!\!/\text{Simplify}$$

$$\rho_i - \frac{s}{2\text{wb}}(((-1 + 2\rho_j)\,\mathfrak{D}_{\{i\{\mathfrak{G},\mathfrak{m},\mathfrak{f}\},j\{\mathfrak{Z},\mathfrak{m},\mathfrak{f}\}\}}$$

$$+ (-1 + 2\rho_j)\,\mathfrak{D}_{\{i\{\mathfrak{G},\mathfrak{m},\mathfrak{f}\},j\{\mathfrak{G},\mathfrak{m},\mathfrak{m}\}\}} + 2\mathfrak{D}_{\{i\{\mathfrak{G},\mathfrak{m},\mathfrak{f}\},j\{\mathfrak{G},\mathfrak{m},\mathfrak{f}\},j\{\mathfrak{Z},\mathfrak{m},\mathfrak{m}\}\}})\,b_j^2$$

$$+ 2((-1 + 2\rho_k)\,\mathfrak{D}_{\{i_{\{\mathfrak{Z},\mathfrak{m},\mathfrak{f}\}},j\{\mathfrak{G},\mathfrak{m},\mathfrak{f}\}\}} + (-1 + 2\rho_k)\,\mathfrak{D}_{\{i_{\{\mathfrak{G},\mathfrak{m},\mathfrak{f}\}},j\{\mathfrak{Z},\mathfrak{m},\mathfrak{m}\}\}}$$

$$- \mathfrak{D}_{\{i\{\mathfrak{G},\mathfrak{m},\mathfrak{f}\},k\{\mathfrak{G},\mathfrak{m},\mathfrak{f}\}\}} + 2\rho_j\mathfrak{D}_{\{i\{\mathfrak{G},\mathfrak{m},\mathfrak{f}\},k\{\mathfrak{G},\mathfrak{m},\mathfrak{f}\}\}} - \mathfrak{D}_{\{i\{\mathfrak{G},\mathfrak{m},\mathfrak{f}\},k\{\mathfrak{G},\mathfrak{m},\mathfrak{m}\}\}}$$

$$+ 2\rho_j\mathfrak{D}_{\{i\{\mathfrak{G},\mathfrak{m},\mathfrak{f}\},k\{\mathfrak{G},\mathfrak{m},\mathfrak{m}\}\}} + \mathfrak{D}_{\{i\{\mathfrak{G},\mathfrak{m},\mathfrak{f}\},j\{\mathfrak{G},\mathfrak{m},\mathfrak{f}\},k\{\mathfrak{G},\mathfrak{m},\mathfrak{f}\}\}} + \mathfrak{D}_{\{i\{\mathfrak{G},\mathfrak{m},\mathfrak{f}\},j\{\mathfrak{G},\mathfrak{m},\mathfrak{f}\},k\{\mathfrak{G},\mathfrak{m},\mathfrak{m}\}\}}$$

$$+ \; \mathfrak{D}_{\{i_{\{\mathcal{G},\mathfrak{m},\mathfrak{f}\}},j_{\{\mathcal{G},\mathfrak{m},\mathfrak{m}\}},k_{\{\mathcal{G},\mathfrak{m},\mathfrak{f}\}}\}} \; + \; \mathfrak{D}_{\{i_{\{\mathcal{G},\mathfrak{m},\mathfrak{f}\}},j_{\{\mathcal{G},\mathfrak{m},\mathfrak{m}\}},k_{\{\mathcal{G},\mathfrak{m},\mathfrak{m}\}}\}} b_j b_k$$

$$+ \; ((-1 + 2\rho_k) \mathfrak{D}_{\{i_{\{\mathcal{G},\mathfrak{m},\mathfrak{f}\}},k_{\{\mathcal{G},\mathfrak{m},\mathfrak{f}\}}\}} \; + \; (-1 + 2\rho_k) \mathfrak{D}_{\{i_{\{\mathcal{G},\mathfrak{m},\mathfrak{f}\}},k_{\{\mathcal{G},\mathfrak{m},\mathfrak{m}\}}\}}$$

$$+ \; 2 \mathfrak{D}_{\{i_{\{\mathcal{G},\mathfrak{m},\mathfrak{f}\}},k_{\{\mathcal{G},\mathfrak{m},\mathfrak{f}\}},k_{\{\mathcal{G},\mathfrak{m},\mathfrak{m}\}}\}}) b_k^2).$$

MakeSelectionCoefficient[$\mathbb{U}$, W2$_{\text{sex}}$] generates the coefficient $a_{\mathbb{U}}$ from the fitness function. ReferencePoint Zero[$\mathfrak{Z}$] sets the reference values in zygotes equal to the current allele frequencies, which are assumed equal across the sexes. The mean fitness in the denominator is represented by wb and is best substituted later.

The complete life cycle: This set of rules defines the change over the whole life cycle:

rules = (Recombination [ ] ∨ SymmetricRecombination [ ]

       ∨ ReferencePointSame [$\mathcal{G}$, {$\mathfrak{A}$, $\mathfrak{Z}$}]

       ∨ ReferencePointZero [{$\mathcal{G}$*, $\mathcal{G}$}]

       ∨ Selection[{Sex}] ∨ UnionOfGametes [ ] ∨ RandomUnion [$\mathcal{G}$]

       ∨ Symmetrize [Disequilibrium] ∨ SymmetricSexes [ ]);

The mean contributions in the two sexes are assumed equal, as are all the linkage disequilibria; thus, SymmetricSexes[] is applied. Symmetrize[Disequilibrium] simplifies $\mathfrak{D}_{\{k,j\}}$ to $\mathfrak{D}_{\{j,k\}}$. ReferencePointSame[$\mathcal{G}$,{$\mathfrak{A}$,$\mathfrak{Z}$}] sets the reference points in diploid stages to be inherited from the gamete stage, where they are equal to the current allele frequencies. UnionOfGametes[] derives genes at the zygote stage from those in the gamete stage, and RandomUnion[$\mathcal{G}$] makes the further assumption that gametes unite at random.

The new mean depends on selection coefficients up to fourth order:

nm = $\mathfrak{m}_{j_{\{\mathcal{G}*,\mathfrak{m}\}}}$ //.rules//Simplify

$$\times \frac{1}{2}(2\mathfrak{m}_{j_{\{\mathcal{G},\mathfrak{f}\}}} \; + \; \mathfrak{D}_{\{j_{\{\mathcal{G},\mathfrak{f}\}},k_{\{\mathcal{G},\mathfrak{f}\}}\}} a_{\mathfrak{m},\{k_{\{\mathcal{G},\mathfrak{m},\mathfrak{f}\}}\}}$$

$$+ \; \mathfrak{D}_{\{j_{\{\mathcal{G},\mathfrak{f}\}},k_{\{\mathcal{G},\mathfrak{f}\}}\}} a_{\mathfrak{m},\{k_{\{\mathcal{G},\mathfrak{m},\mathfrak{m}\}}\}} \; + \; \mathfrak{D}_{\{j_{\{\mathcal{G},\mathfrak{f}\}},j_{\{\mathcal{G},\mathfrak{f}\}},k_{\{\mathcal{G},\mathfrak{f}\}}\}} \; a_{\mathfrak{m},\{j_{\{\mathcal{G},\mathfrak{m},\mathfrak{f}\}},k_{\{\mathcal{G},\mathfrak{m},\mathfrak{f}\}}\}}$$

$$+ \; \mathfrak{D}_{\{j_{\{\mathcal{G},\mathfrak{f}\}},j_{\{\mathcal{G},\mathfrak{f}\}},k_{\{\mathcal{G},\mathfrak{f}\}}\}} a_{\mathfrak{m},\{j_{\{\mathcal{G},\mathfrak{m},\mathfrak{m}\}},k_{\{\mathcal{G},\mathfrak{m},\mathfrak{m}\}}\}}$$

$$+ \; \mathfrak{D}_{\{j_{\{\mathcal{G},\mathfrak{f}\}},j_{\{\mathcal{G},\mathfrak{f}\}}\}}(a_{\mathfrak{m},\{j_{\{\mathcal{G},\mathfrak{m},\mathfrak{f}\}}\}} \; + \; a_{\mathfrak{m},\{j_{\{\mathcal{G},\mathfrak{m},\mathfrak{m}\}}\}}$$

$$+ \; \mathfrak{D}_{\{j_{\{\mathcal{G},\mathfrak{f}\}},k_{\{\mathcal{G},\mathfrak{f}\}}\}}(a_{\mathfrak{m},\{j_{\{\mathcal{G},\mathfrak{m},\mathfrak{f}\}},j_{\{\mathcal{G},\mathfrak{m},\mathfrak{f}\}}\}}$$

$$+ \; a_{\mathfrak{m},\{j_{\{\mathcal{G},\mathfrak{m},\mathfrak{f}\}},j_{\{\mathcal{G},\mathfrak{m},\mathfrak{m}\}},k_{\{\mathcal{G},\mathfrak{m},\mathfrak{m}\}}\}}))$$

$$+ \; \mathfrak{D}^2_{\{j_{\{\mathcal{G},\mathfrak{f}\}},k_{\{\mathcal{G},\mathfrak{f}\}}\}} a_{\mathfrak{m},\{j_{\{\mathcal{G},\mathfrak{m},\mathfrak{f}\}},k_{\{\mathcal{G},\mathfrak{m},\mathfrak{f}\}}\}}$$

$$+ \; \mathfrak{D}^2_{\{j_{\{\mathcal{G},\mathfrak{f}\}},k_{\{\mathcal{G},\mathfrak{f}\}}\}} a_{\mathfrak{m},\{j_{\{\mathcal{G},\mathfrak{m},\mathfrak{m}\}},k_{\{\mathcal{G},\mathfrak{m},\mathfrak{f}\}},k_{\{3,\mathfrak{m},\mathfrak{m}\}}\}}$$

$$+ \; 2 \mathfrak{D}_{\{j_{\{\mathcal{G},\mathfrak{f}\}},k_{\{\mathcal{G},\mathfrak{f}\}}\}} \mathfrak{D}_{\{j_{\{\mathcal{G},\mathfrak{f}\}},j_{\{\mathcal{G},\mathfrak{f}\}},k_{\{\mathcal{G},\mathfrak{f}\}}\}} a_{\mathfrak{m},\{j_{\{\mathcal{G},\mathfrak{m},\mathfrak{f}\}},j_{\{\mathcal{G},\mathfrak{m},\mathfrak{m}\}},k_{\{\mathcal{G},\mathfrak{m},\mathfrak{f}\}},k_{\{\mathcal{G},\mathfrak{m},\mathfrak{m}\}}\}})$$

We now substitute in the selection coefficients for this specific scheme:

nms = nm//. {$a_{\text{sex}\_,\mathbb{U}\_}$:→ (MakeSelectionCoefficient [$\mathbb{U}$, W2$_{\text{sex}}$]/wb)}

       ∨ Biallelic [ ] ∨ rules/. {$\mathfrak{m}_{i\_} \to \rho_i$}//Simplify

$$\rho_{j_{\{\mathcal{G},\mathfrak{f}\}}} + \frac{s}{\text{wb}}\left(b_j(1 - \rho_{j_{\{\mathcal{G},\mathfrak{f}\}}})\rho_{j_{\{\mathcal{G},\mathfrak{f}\}}}\left(\frac{1}{2}(1 - 2\rho_{j_{\{\mathcal{G},\mathfrak{f}\}}})b_j + (1 - 2\rho_{k_{\{\mathcal{G},\mathfrak{f}\}}})b_k\right)\right.$$

$$\left. + \; \mathfrak{D}_{\{j_{\{\mathcal{G},\mathfrak{f}\}},k_{\{\mathcal{G},\mathfrak{f}\}}\}} b_k\left(\frac{1}{2}(1 - 2\rho_{k_{\{\mathcal{G},\mathfrak{f}\}}})b_k\right)\right)$$

This expression is the same as Equation 42 with two loci.

Other processes: Non-Mendelian inheritance is handled by interpreting recombination rates appropriately. For example, a cytoplasmically inherited gene is certain to be inherited from the mother:

$\{\mathfrak{r}_{\{i,j\},\{k\}} \; \mathfrak{r}_{\{j\},\{i,k\}}\}$/.NonMendelian[CytoplasmicLoci → {i}]

$\{\mathfrak{r}_{\{j\}, \{k\}}, 0\}$

$\mathfrak{r}_{\{i,j\},\{k\}}$ is the chance that genes i, j are inherited from the mother and k from the father; if i is cytoplasmically inherited, then this is equal to the recombination rate for the autosomal loci, $\mathfrak{r}_{\{j\},\{k\}}$. The converse, $\mathfrak{r}_{\{j\},\{i,k\}}$, is impossible if i is cytoplasmically inherited.

Migration is implemented in a similar way to recombination, as a set of transmission rules. For example, this is the cross-genome association between loci j and k among juveniles, $\tilde{\mathfrak{J}}$, after migration:

$\mathfrak{D}_{\{j_{\{\tilde{\mathfrak{J}},1,\mathfrak{m},\mathfrak{m}\}},k_{\{\tilde{\mathfrak{J}},1,\mathfrak{m},\mathfrak{f}\}}\}}$/.Migration[{1, 2} &]

$\text{M}_{1,1} \, (\mathfrak{D}_{\{j_{\{\mathcal{G},1,\mathfrak{m},\mathfrak{m}\}},k_{\{\mathcal{G},1,\mathfrak{m},\mathfrak{f}\}}\}} \; + \; \mathfrak{D}_{\{k_{\{\mathcal{G},1,\mathfrak{m},\mathfrak{f}\}}\}}(-\mathfrak{P}_{j_{\{\tilde{\mathfrak{J}},1,\mathfrak{m},\mathfrak{m}\}}} \; + \; \mathfrak{P}_{j_{\{\mathcal{G},1,\mathfrak{m},\mathfrak{m}\}}})$

$\; + \; \mathfrak{D}_{\{j_{\{\mathcal{G},1,\mathfrak{m},\mathfrak{m}\}}\}}(-\rho_{k_{\{\tilde{\mathfrak{J}},1,\mathfrak{m},\mathfrak{f}\}}} \; + \; \mathfrak{P}_{k_{\{\mathcal{G},1,\mathfrak{m},\mathfrak{f}\}}}) \; + \; (-\mathfrak{P}_{j_{\{\tilde{\mathfrak{J}},1,\mathfrak{m},\mathfrak{m}\}}} \; + \; \mathfrak{P}_{j_{\{\mathcal{G},1,\mathfrak{m},\mathfrak{m}\}}})(-\mathfrak{P}_{k_{\{\tilde{\mathfrak{J}},1,\mathfrak{m},\mathfrak{f}\}}} \; + \; \mathfrak{P}_{k_{\{\mathcal{G},1,\mathfrak{m},\mathfrak{f}\}}}))$

$\; + \; \text{M}_{1,2} \, (\mathfrak{D}_{\{j_{\{\mathcal{G},2,\mathfrak{m},\mathfrak{m}\}},k_{\{\mathcal{G},2,\mathfrak{m},\mathfrak{f}\}}\}} \; + \; \mathfrak{D}_{\{k_{\{\tilde{\mathfrak{J}},2,\mathfrak{m},\mathfrak{f}\}}\}} \; (-\mathfrak{P}_{j_{\{\tilde{\mathfrak{J}},1,\mathfrak{m},\mathfrak{m}\}}} \; + \; \mathfrak{P}_{j_{\{\mathcal{G},2,\mathfrak{m},\mathfrak{m}\}}})$

$\; + \; \mathfrak{D}_{\{j_{\{\mathcal{G},2,\mathfrak{m},\mathfrak{m}\}}\}} \; (-\mathfrak{P}_{k_{\{\tilde{\mathfrak{J}},1,\mathfrak{m},\mathfrak{f}\}}} \; + \; \mathfrak{P}_{k_{\{\mathcal{G},2,\mathfrak{m},\mathfrak{f}\}}}) \; + \; (-\mathfrak{P}_{j_{\{\tilde{\mathfrak{J}},1,\mathfrak{m},\mathfrak{m}\}}} \; + \; \mathfrak{P}_{j_{\{\mathcal{G},2,\mathfrak{m},\mathfrak{m}\}}})(-\mathfrak{P}_{k_{\{\tilde{\mathfrak{J}},1,\mathfrak{m},\mathfrak{f}\}}} \; + \; \mathfrak{P}_{k_{\{\mathcal{G},2,\mathfrak{m},\mathfrak{f}\}}}))$

These expressions are linear sums, weighted rates, $\text{M}_{i,j}$. The set of possible source demes is defined by {1,2} &—that is, all demes can receive immigrants from demes 1 or 2. The deme in which a gene is found is indicated by an extra element in the context, which takes value 1 or 2 in this example.