

# Deleterious Mutations and the Genetic Variance of Male Fitness Components in *Mimulus guttatus*

John K. Kelly<sup>1</sup>

Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, Kansas 66045

Manuscript received November 11, 2002

Accepted for publication March 10, 2003

## ABSTRACT

Deleterious mutations are relevant to a broad range of questions in genetics and evolutionary biology. I present an application of the “biometric method” for estimating mutational parameters for male fitness characters of the yellow monkeyflower, *Mimulus guttatus*. The biometric method rests on two critical assumptions. The first is that experimental inbreeding changes genotype frequencies without changing allele frequencies; *i.e.*, there is no genetic purging during the experiment. I satisfy this condition by employing a breeding design in which the parents are randomly extracted, fully homozygous inbred lines. The second is that all genetic variation is attributable to deleterious mutations maintained in mutation-selection balance. I explicitly test this hypothesis using likelihood ratios. Of the three deleterious mutation models tested, the first two are rejected for all characters. The failure of these models is due to an excess of additive genetic variation relative to the expectation under mutation-selection balance. The third model is not rejected for either of two log-transformed male fitness traits. However, this model imposes only “weak conditions” and is not sufficiently detailed to provide estimates for mutational parameters. The implication is that, if biometric methods are going to yield useful parameter estimates, they will need to consider mutational models more complicated than those typically employed in experimental studies.

The long-debated question of whether or not genetic variation in fitness primarily reflects contributions of low-frequency deleterious alleles maintained by the balance between selection and mutation, or has a substantial contribution from variants maintained at intermediate frequencies by selection, is still unanswered.

CHARLESWORTH and HUGHES (2000)

**D**ELETERIOUS mutations have been a focus of study in both genetics and evolutionary biology for most of the history of each field (CROW 1993). MULLER (1950) argued that modern medicine would allow deleterious mutations to eventually accumulate to intolerable levels within the human species unless countered by “a rationally directed guidance of reproduction.” Muller’s claims initiated a prolonged debate with other geneticists, most notably Theodosius Dobzhansky (see CROW 1987). Since the Muller-Dobzhansky dispute, biologists have hypothesized that harmful mutations may be critical to a wide range of ecological and evolutionary phenomena. These include the evolution of sex, mating systems, and mate choice (PAMILO *et al.* 1987; KONDRASHOV 1988; CHARLESWORTH *et al.* 1990b; UYENOYAMA *et al.* 1992); the evolution of recombination rates (KONDRASHOV 1984; PALS-SON 2002); the persistence of small populations (LANDE 1994; LYNCH *et al.* 1995; KONDRASHOV 1995); and the maintenance of both molecular and quantitative trait variation (LEWONTIN 1974; TURELLI 1984; CHARLESWORTH *et al.* 1993). The validity of these hypotheses

depends not only on the rate that deleterious mutations occur, but also on the magnitude of their effects, on dominance relations, and on how deleterious alleles at different loci combine to influence the overall fitness of an individual.

The primary mutational parameters are  $U$ ,  $h$ , and  $s$ . The genomic deleterious mutation rate,  $U$ , equals  $2\sum\mu_i$ , where  $\mu_i$  is the deleterious mutation rate at the  $i$ th locus, and the summation is taken over all loci affecting fitness. The selection coefficient,  $s$ , is the proportional fitness reduction caused by the mutation when in homozygous form. The dominance coefficient,  $h$ , characterizes mutational effects in heterozygotes (see Table 1). A mutation with  $s = 0.05$  and  $h = 0.2$  reduces fitness by 1% in heterozygotes and 5% in homozygotes. As mutational effects are likely to vary among loci (and among different mutations at the same locus), estimates for  $s$  and  $h$  may actually represent average values for mutational effects. However, because the values of  $h$  and  $s$  for a given allele are assumed to be constant, this is essentially a model of *unconditionally* deleterious mutations. Given the range of applications outlined above, accurate estimates for  $U$ ,  $h$ , and  $s$  will shed light on a number of unsolved problems in evolutionary biology.

At least three different approaches have been devel-

<sup>1</sup>Address for correspondence: Department of Ecology and Evolutionary Biology, University of Kansas, 1200 Sunnyside Ave., Lawrence, KS 66045. E-mail: jkk@ku.edu

oped to estimate the rate and effects of deleterious mutations: mutation-accumulation experiments, molecular evolution surveys, and biometric analyses. These methods, briefly summarized below, each have their respective strengths and weaknesses. They are applicable to different sorts of data. The procedures also differ in the specific parameters that are estimated and in their underlying assumptions regarding genetics and evolution.

Mutation-accumulation experiments estimate mutational parameters from the observed genetic divergence of initially identical lines (MULLER 1928; BATEMAN 1959; MUKAI *et al.* 1972). These lines are maintained at small population sizes and diverge due to the random accumulation of new mutations. Estimates for mutational parameters can then be extracted from the rate and overall pattern of divergence through a variety of statistical procedures (KEIGHTLEY 1994; GARCIA-DORADO 1997; LYNCH and WALSH 1998; SHAW *et al.* 2002, and references therein). Mutation accumulation is the most direct of the estimation methods and these experiments have provided most of our information regarding mutational parameters (reviewed by SIMMONS and CROW 1977; LYNCH and WALSH 1998, pp. 343–348). Unfortunately, mutation accumulations are highly labor intensive. The parameter estimates are subject to large statistical uncertainty. Perhaps most importantly, mutation-accumulation experiments may fail to detect mutations with small effects on fitness (KEIGHTLEY and EYRE-WALKER 1999; LYNCH *et al.* 1999). As a consequence,  $U$  may be substantially underestimated and  $s$  substantially overestimated.

The molecular survey method is based on the rates of synonymous and nonsynonymous gene sequence evolution within an evolutionary lineage (KONDRASHOV and CROW 1993; EYRE-WALKER and KEIGHTLEY 1999). If synonymous mutations can be treated as selectively neutral, an estimate for  $U$  can be obtained from the deficiency of nonsynonymous substitutions (relative to synonymous substitutions) within the genome as a whole. Molecular surveys measure the time-integrated effects of selection (as opposed to the direct effects of mutations on viability and fertility) and can thus potentially detect very weakly selected mutations. A second advantage is that inferences are based on the actual historical results of natural selection (genetic substitutions). In contrast, the other methods are based on measurements of fitness surrogates (*e.g.*, viability, female fertility, male mating success, etc.) that are usually measured under controlled laboratory conditions. On the negative side, the molecular survey analysis depends on several unrealistic assumptions. In particular, the method assumes that adaptive amino acid substitutions do not occur and that all synonymous changes are strictly neutral; *i.e.*, there is no selection on codon usage (AKASHI 1997). EYRE-WALKER and KEIGHTLEY (1999) argue that deviations from these assumptions will lead to underestimates for  $U$ . As a consequence, molecular surveys may provide a robust lower bound for this quantity.

Biometric methods use patterns of phenotypic variation in fitness components to infer mutational parameters (MORTON *et al.* 1956; CHARLESWORTH *et al.* 1990a; DENG and LYNCH 1996). This approach is grounded in population genetic models of mutation-selection balance (*e.g.*, HALDANE 1927). Biometric quantities, such as the amount of genetic variation in fitness and the magnitude of inbreeding depression, can be expressed as functions of  $U$ ,  $h$ , and  $s$ . Estimates of inbreeding depression or the genetic variance can thus be used to estimate mutational parameters by inverting the population genetic equations. For example, DENG and LYNCH (1996) derive simple estimators for  $U$ ,  $h$ , and  $s$  from the means and genetic variances of inbred and outbred individuals within an experimental population.

Biometric methods are generally less labor intensive than mutation-accumulation experiments. More importantly, these experiments can potentially yield estimators that are relatively low in bias for a wide range of mutational parameters (DENG 1998; DENG *et al.* 2002). However, like the two previous methods, biometric methods are also encumbered with potentially limiting assumptions. Two particularly important assumptions are (1) that all genetic variation in fitness is maintained through mutation-selection balance and (2) that experimental inbreeding changes genotype frequencies without changing allele frequencies. Assumption 1 will break down if balancing selection of some form (*e.g.*, heterozygote advantage, frequency-dependent selection, spatial and/or temporal variation in selection, genotype-by-environment variation, etc.) maintains a substantial fraction of the genetic variance in fitness. If this is so, biometric quantities may deviate substantially from their expected values given  $U$ ,  $h$ , and  $s$ . Assumption 2 breaks down if deleterious alleles are eliminated or “purged” from the experimental population so that the frequencies of deleterious mutations differ between inbred and outbred individuals. This will bias parameter estimates.

This article describes a biometric analysis of male fitness measures in *Mimulus guttatus*, a wildflower commonly known as yellow monkeyflower. My purpose is to address both of the potential pitfalls associated with the biometric method. Mutational parameters are estimated from a breeding experiment that uses randomly extracted, highly inbred lines as parents. This effectively eliminates the problem of purging *within the experiment*. Allele frequencies should not differ (consistently) between outbred and inbred plants. Second, the experiment is designed so that the number of “empirical comparisons” is greater than the number of unknown parameters. This provides the degrees of freedom necessary to test the adequacy of a model prior to parameter estimation. Following the suggestion of DENG and LYNCH (1996, p. 358), I use likelihood ratios to test the assumption that all genetic variation is maintained through mutation-selection balance.

Biometric methods depend on specific features of the experimental design such as the type and levels of in-

breeding, the sorts of experimental crosses, etc. In the following sections, I first describe the experimental design and define the biometric quantities that can be estimated from this design. I then derive the expected values for these quantities under various forms of the deleterious mutation model (DMM). Next, I describe the likelihood-based methods for parameter estimation. A parametric bootstrapping procedure is developed, both to assess the adequacy of the experimental design and to perform hypothesis tests (obtain an appropriate  $P$  value for the various likelihood ratios). Finally, the statistical model is applied to phenotypic data from *M. guttatus*. The characters are flower size and two male fitness measures, pollen production and the pollen size index (PSI). The PSI is strongly correlated with pollen viability in *M. guttatus* (KELLY *et al.* 2002).

#### BIOMETRIC QUANTITIES AND THEIR EXPECTED VALUES

Consider a breeding design in which the parents are randomly extracted, fully homozygous inbred lines. Such lines can be formed in several different ways. For select species, completely homozygous lines can be formed in a single generation via chromosomal manipulations, gametophytic self-fertilization, or double-haploid formation. Repeated rounds of inbreeding provide a more generally applicable, albeit slower, method of line formation. Successive generations of self-fertilization can yield lines that are almost fully inbred in only five to six generations. The parental lines used in this study were developed by John Willis through six successive generations of self-fertilization (WILLIS 1999a,b).

Regardless of the method, genetic purging is likely to accompany the formation of the inbred lines. Lines that become homozygous for mutations that are lethal or cause sterility in homozygous form will go extinct and eliminate these alleles from the experimental population. In contrast, deleterious mutations with smaller effects should fix randomly within lines (WILLIS 1999a,b). As a consequence, a genetic analysis using these lines as parents is informative only about the nonlethal and nonsterile fraction of mutations. Parameter estimates are thus specific to this class of mutations, termed “detrimentals” by SIMMONS and CROW (1977).

A major advantage of using fully inbred lines as parents in a biometric breeding design is that they are already fully purged. Subsequent inbreeding or outcrossing should result in predictable changes in homozygosity without changes in allele frequency (apart from the random changes due to finite sample sizes). Consider the breeding design described by Figure 1. Each “extended family” is constituted from the progeny of three randomly selected inbred lines (labeled L1, L2, and L3). One line (L1) is assigned as the “sire” and is mated to the other two lines, the “dams.” The progeny of these crosses comprise two distinct but related outbred families (the families in Figure 1, bottom). Each line is

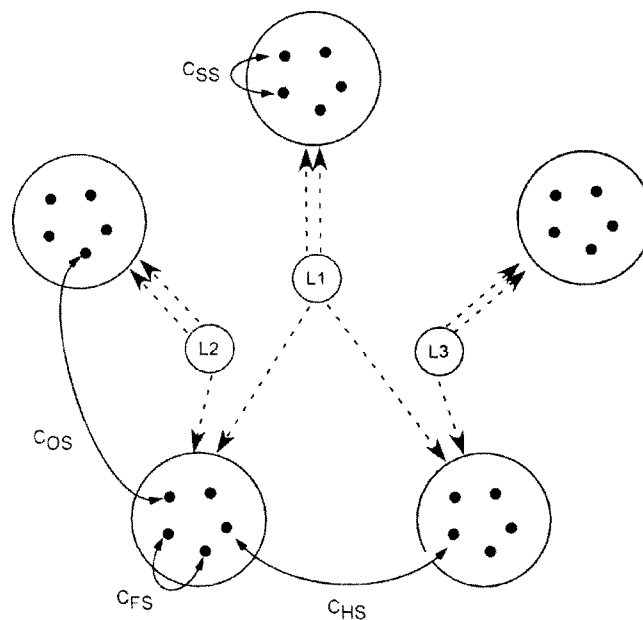


FIGURE 1.—The relationships among individuals within an extended family derived from a particular set of lines (L1, L2, and L3). Individual progeny, denoted by solid circles, are contained within families. Dashed lines denote lines of descent (a double line for selfed progeny). Of the three parental lines, L1 is the sire for the outbred families, while L2 and L3 are the dams.

also self-fertilized to produce three inbred families. The inbred families are unrelated to each other, but each is related to one or both of the outbred families. This same mating scheme is applied to other sets of lines generating a large number of extended families.

The measurement of progeny from this design provides data sufficient to estimate the mean phenotype of both inbred and outbred plants as well as a number of (co)variance components. Let  $\beta$  denote the difference in mean phenotype between inbred and outbred plants. Four different comparisons among relatives can be quantified as covariances (Figure 1).  $C_{SS}$  is the covariance among plants within selfed families (selfed siblings).  $C_{FS}$  is the covariance among plants within outbred families (full siblings).  $C_{OS}$  is the covariance between outbred plants and their inbred siblings (related through a single parent). Finally,  $C_{HS}$  is the covariance of plants from different outbred families (half-siblings related through the sire line). These  $C$  terms are “observational variance components” in the terminology of FALCONER (1989). The observational components can be expressed as a function of the “causal components”  $V_A$ ,  $C_{AD}$ ,  $V_D$ , and  $V_{DI}$  (see COCKERHAM and WEIR 1984; SHAW *et al.* 1998; KELLY and ARATHI 2003). I do not use these relationships here. Instead, the observational components are expressed directly as functions of the mutational parameters (see Equations 1–7 below).

Estimates for  $\beta$  and the observational components are used to assess the sufficiency of a particular deleterious mutation model to explain genetic variability. If demon-

TABLE 1

The basic parameterization of a single locus in which the deleterious allele (a) has population frequency  $q$

Genotype	Frequency in a randomly mating population	Fitness	Phenotypic effect on a fitness component
AA	$p^2$	1	0
Aa	$2pq$	$1 - hs$	$-\alpha hs$
aa	$q^2$	$1 - s$	$-\alpha s$

See text for additional details.

strated as sufficient, the same data can then be used to estimate the mutational parameters of that model. For both purposes, it is necessary to express  $\beta$ ,  $C_{SS}$ ,  $C_{FS}$ ,  $C_{OS}$ , and  $C_{HS}$  as functions of the mutational parameters. This requires a model for genotypic effects on both phenotype and fitness. Table 1 summarizes the standard fitness model for an individual locus: the deleterious allele (a) exists at population frequency  $q$ . Following CHARLESWORTH and HUGHES (2000), I distinguish the evolutionary fitness of a genotype (determined by  $h$  and  $s$ ) from its contribution to a measurable fitness component. The quantity  $\alpha$  “translates” between phenotype and fitness.

Predicted values for  $\beta$ ,  $C_{SS}$ ,  $C_{FS}$ ,  $C_{OS}$ , and  $C_{HS}$  can be obtained by applying quantitative genetic models developed for inbreeding populations (COCKERHAM and WEIR 1984; see also HARRIS 1964; JACQUARD 1974). As these quantities depend on the contributions of many loci, it is necessary to subscript the various quantities in Table 1. As shown in APPENDIX A,

$$\beta = - \sum_i \alpha_i q_i s_i (1 - 2h_i), \tag{1a}$$

$$C_{SS} = \sum_i \alpha_i^2 q_i s_i^2, \tag{1b}$$

$$C_{OS} = \sum_i \alpha_i^2 q_i h_i s_i^2, \tag{1c}$$

$$C_{FS} = 2 \sum_i \alpha_i^2 q_i (h_i s_i)^2, \tag{1d}$$

$$C_{HS} = \sum_i \alpha_i^2 q_i (h_i s_i) = \frac{1}{2} C_{FS}, \tag{1e}$$

where  $q_i$ ,  $\alpha_i$ ,  $h_i$ , and  $s_i$  are the allele frequency and effects associated with the  $i$ th locus (see also CHARLESWORTH and HUGHES 2000). The summations are taken over all loci harboring deleterious alleles. These equations define the genetic covariances among relatives. Environmental factors may also affect the phenotypic resemblance of relatives. In this treatment, these effects are characterized by other variables (see HYPOTHESIS TESTING AND PARAMETER ESTIMATION).

Equations 1 depend on two important assumptions. The first is that the loci contribute additively to the phenotype. Oftentimes, it is assumed that the effects of deleterious mutations at different loci combine multiplica-

tively. In this situation, trait values should be log-transformed prior to analysis. As the appropriate scale is not obvious *a priori*, here I apply the methods to both log-transformed and untransformed data (see RESULTS and DISCUSSION). Second, each component is linear in  $q_i$  because I have neglected terms of order  $q_i^2$  (see APPENDIX A). This approximation should be quite accurate because deleterious alleles will not usually obtain high frequencies in a large population.

The frequency of deleterious alleles at a given locus ( $q$ ) depends on the mutational parameters. I now assume that the natural population is large, randomly mating, and in mutation-selection balance equilibrium. I further assume that deleterious mutations have some phenotypic effect in heterozygotes ( $h > 0$ ), an assumption supported by genetic data (SIMMONS and CROW 1977). As shown by HALDANE (1927), the expected frequency  $q_i$  of a partially recessive mutation is equal to  $\mu_i / (h_i s_i)$ . Substitution of this expression into Equations 1 yields predictions for the biometric quantities in terms of mutational parameters.

The final step is to characterize the mutational spectrum, the joint density function of  $\alpha_i$ ,  $h_i$ , and  $s_i$  across loci. I consider three alternative models. In model I, mutational effects are fixed and do not vary across loci. In model II, mutational effects vary according to an exponential distribution. Model III considers variable mutational effects without imposing distributional assumptions.

**Model I:** All new mutations have fixed effects  $\alpha$ ,  $h$ , and  $s$ . With constant mutational effects,

$$\beta = U\alpha \left( 1 - \frac{1}{2h} \right), \tag{2a}$$

$$C_{SS} = \frac{U\alpha^2 s}{2h}, \tag{2b}$$

$$C_{OS} = \frac{U\alpha^2 s}{2}, \tag{2c}$$

$$C_{FS} = U\alpha^2 hs, \tag{2d}$$

$$C_{HS} = \frac{C_{FS}}{2}, \tag{2e}$$

where  $U = 2\sum\mu_i$ , as defined previously.

The five estimable biometric quantities are functions of only three unknown quantities:  $U\alpha$ ,  $h$ , and  $s\alpha$ . Because both  $U$  and  $s$  are always included in a product with  $\alpha$ , neither parameter can be estimated individually (unless a particular value for  $\alpha$  is assumed). Only the composite parameters,  $U\alpha$  and  $s\alpha$ , can be estimated. The fact that the five measurable statistics are a function of only three unknowns allows us to test the sufficiency of model I. The model imposes the constraints that  $C_{FS} = 2C_{OS}^2 / C_{SS}$  and  $C_{HS} = \frac{1}{2} C_{FS}$ .

**Model II:** New mutations vary in their effects according to a simple probability model. Following DENG



and LYNCH (1996), model II posits an exponential density for  $s$ ,

$$f(s) = \frac{1}{\bar{s}} \text{Exp}\left(\frac{-s}{\bar{s}}\right), \quad (3)$$

where  $\bar{s}$  is the average homozygous effect. Since mutations with larger effects tend to be more recessive (MACKAY *et al.* 1992; CABALLERO and KEIGHTLEY 1994), the dominance coefficient is a function of  $s$ ,

$$h(s) = \frac{1}{2} \text{Exp}(Ks), \quad (4)$$

where  $(-1/\bar{s}) < K < 0$ . DENG and LYNCH (1996) suggest a value of  $K = -13$  from a consideration of genetic data. Here, it is a parameter to be estimated from the data.

The parameters of model II are  $U$ ,  $K$ ,  $\bar{s}$ , and  $\alpha$ . I assume that  $\alpha$  is the same across loci. The average dominance of new mutations,  $\bar{h}$ , is

$$\bar{h} = \frac{1}{2(1 - K\bar{s})}. \quad (5)$$

Integrating over the distributions of mutational effects (see APPENDIX B), we obtain predicted values for the biometric quantities:

$$\beta = U\alpha \left[ 1 - \frac{1}{1 + K\bar{s}} \right] \quad (6a)$$

$$C_{SS} = \frac{U\alpha^2\bar{s}}{(1 + K\bar{s})^2} \quad (6b)$$

$$C_{OS} = \frac{U\alpha^2\bar{s}}{2} \quad (6c)$$

$$C_{FS} = \frac{U\alpha^2\bar{s}}{2(1 - K\bar{s})^2} \quad (6d)$$

$$C_{HS} = \frac{C_{FS}}{2}. \quad (6e)$$

Again, our five observable quantities are functions of three unknowns. These three estimable quantities are composite functions of the parameters  $U\alpha$ ,  $K\bar{s}$ , and  $\bar{\alpha}$ . Model II constrains the value of  $C_{FS}$ ,

$$C_{FS} = \frac{h^* C_{SS}}{4 - 4\sqrt{2h^*} + 2h^*}, \quad (7)$$

where  $h^* = C_{OS}/C_{SS}$ .

**Model III (weak conditions):** Model II represents only one way in which mutational effects may vary across loci (or between different mutations at the same locus). A range of alternative models could be considered. For example, CABALLERO and KEIGHTLEY (1994) consider a model that allows variable dominance of mutations, even if they have the same value for  $s$ . Other generalizations are also possible (ZHANG *et al.* 2002 and references therein). Given the range of possibilities, it is fortunate that we can make a few statements about the relative

magnitudes of the biometric quantities without an explicit model of mutational effects. As long as deleterious alleles are rare and  $h_i < 0.5$  at all loci, then Equations 1 imply that  $C_{OS} < \frac{1}{2}C_{SS}$ ,  $C_{FS} < C_{OS}$ , and  $C_{HS} = \frac{1}{2}C_{FS}$ . Model III is not sufficiently detailed to allow parameter estimation. However, a model with these constraints can be tested against the unconstrained model. The conditions associated with model III are “weak” because they may often hold even if a substantial portion of the variance in fitness is maintained by balancing selection.

## STUDY SPECIES AND METHODS

*M. guttatus* ( $2n = 28$ ; Scrophulariaceae) is a small wildflower, increasingly used as a model organism for genetic studies of floral variation and the evolution of plant mating systems (MACNAIR and CUMBES 1989; RITLAND 1989; CARR and FENSTER 1994; FENSTER and RITLAND 1994; ROBERTSON *et al.* 1994). It occurs throughout western North America and local populations may be either annual or perennial. The plants of this study are derived from Iron Mountain, which is a large, annual (or winter annual) population located in the Cascade Mountains of central Oregon (see WILLIS 1996, 1999a,b). *M. guttatus* is a self-compatible, but the Iron Mountain population is predominantly outcrossing (WILLIS 1993b).

John H. Willis initiated  $\sim 1200$  independent lines of *M. guttatus* in August of 1995. Each line was founded from the seed set of a separate field-collected plant from Iron Mountain. Each line was subsequently maintained in the greenhouse by single-seed descent (self-fertilization) for six generations. Six successive generations of self-fertilization yield an inbreeding coefficient of  $>0.98$  and I treat the parents as fully inbred ( $f = 1$ ) for the purpose of calculating covariances among relatives. As expected, these lines are almost completely homozygous at highly polymorphic microsatellite loci with different lines fixed for different alleles (WILLIS 1999a).

Three hundred of these sixth-generation lines were used to found 100 extended families following the experimental design of Figure 1. A single plant from the sire line was grown to maturity, self-fertilized, and used as a pollen source for a single plant from each of the dam lines. Each of the dams was also self-fertilized. A maximum of eight pots were seeded per family (40 per extended family) on March 9, 2001 in the University of Kansas greenhouses. At 2 weeks after seeding, all pots were thinned to a single plant (randomly selected) and were subsequently fertilized once per week. A series of morphological measurements were taken on the first flower produced by each plant (see KELLY and ARATHI 2003) and the anthers were collected into a microcentrifuge tube. Flowering was monitored daily until 60 days postseeding, by which time the great majority of plants had flowered. In total, 2345 plants were measured (1113 outbred and 1232 inbred). The distribution of plants across families was not balanced and many extended

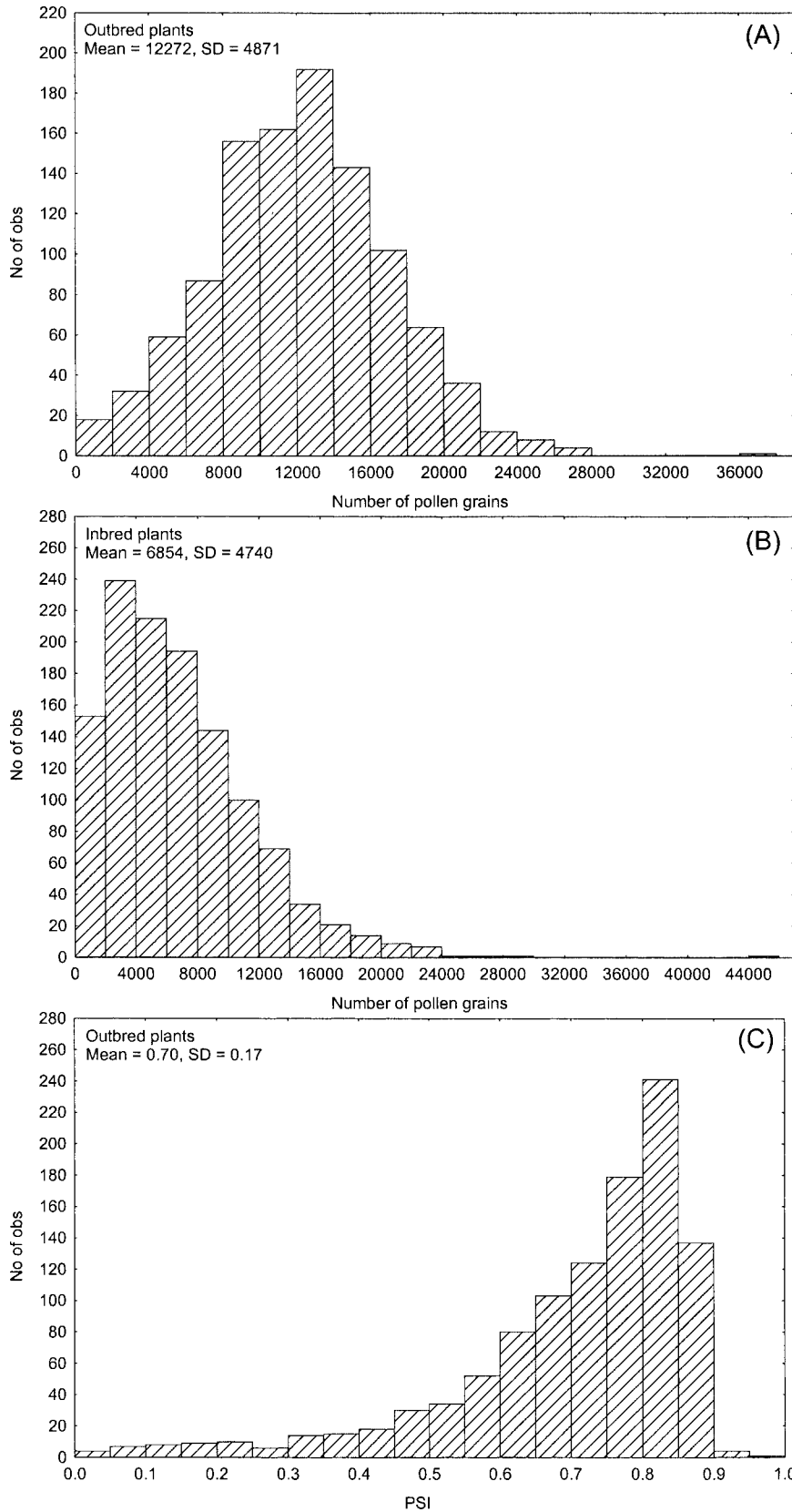


FIGURE 2.—The distributions of male fitness components (untransformed) among outbred and inbred plants, respectively.

families were completely missing one or more of their component families. This is relevant to the analysis. Maximum likelihood accommodates unbalanced de-

signs better than least-squares or “ANOVA-style” methods do (SEARLE *et al.* 1992).

Pollen number and the PSI were estimated for each

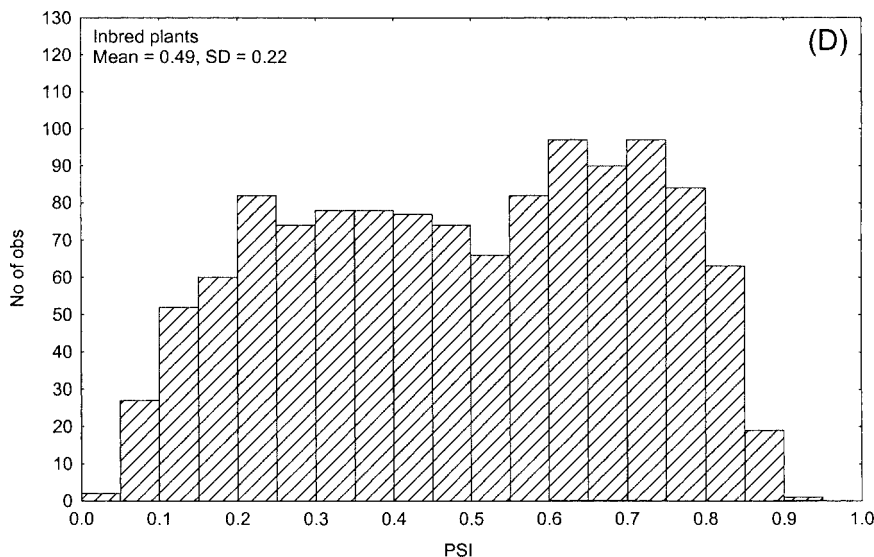


FIGURE 2.—Continued.

sample using a Coulter counter model Z1 dual (Coulter, Miami). The microcentrifuge tubes were left open for 1 week after pollen collection to allow the anthers to dry and release their pollen. Each sample was then diluted into 20 ml of electrolyte solution and run through the Coulter counter. The machine counts the number of particles between two size thresholds (10 and 25  $\mu\text{M}$ ) and the number of particles that are larger than the upper threshold ( $>25 \mu\text{M}$ ). The PSI is the proportion of grains in the upper size category. The size thresholds were set on the basis of results from a previous study showing a strong positive correlation between PSI and pollen viability as measured through standard staining techniques (KELLY *et al.* 2002).

The distributions of pollen number and PSI within both inbred and outbred plants are given in Figure 2. A genetic analysis of the floral morphology is presented elsewhere (KELLY and ARATHI 2003). One floral trait, corolla width, is included here as a comparison to the male fitness measures (outbred plants, mean = 19.13 mm, SD = 2.65; inbred plants, mean = 17.32, SD = 2.95). For analysis, the original measurements were divided by the outbred mean value for that trait. This linear transformation is commonly used in biometric studies as it allows comparison of parameter estimates for traits that differ in the magnitudes of their measurement values (HOULE 1992; CHARLESWORTH and HUGHES 2000). We also considered log-transformed values for pollen number and PSI, as this may be more appropriate if loci combine multiplicatively.

#### HYPOTHESIS TESTING AND PARAMETER ESTIMATION

Maximum likelihood is used for both testing the various models and estimating parameters. By methods described below, I find the parameter values for each of

the mutational models (I, II, and III) that produce the highest likelihood for the data. These values are the maximum-likelihood estimates (MLEs) for the parameters of that model. The sufficiency of that model is tested by comparing the maximum-likelihood value for the model with the maximum likelihood of the “unconstrained model.” The latter is based on direct estimates for the biometric quantities without the particular restrictions imposed by the various mutational models (see section on constraints below).

The resemblance of relatives may depend on environmental factors as well as genetic factors. These factors must be accounted for in order to infer genetic parameters from phenotypic covariances. Maternal effects are a potentially important source of resemblance among relatives. The variance of maternal effects,  $V_M$ , can be estimated from this design because the comparison between inbred and outbred sibs (estimating  $C_{OS}$ ) involves both male and female parents (Figure 1). Also, the magnitude of environmental deviations may be greater for inbred individuals than for outbred individuals (LERNER 1954; KELLY and ARATHI 2003). I thus estimate different environmental variances for outbred and inbred plants,  $V_{E(o)}$  and  $V_{E(i)}$ .

With regard to hypothesis testing, it is important to distinguish universal constraints from model-specific constraints. The universal constraints apply to all of the models including the unconstrained model. These follow from basic features of inheritance, independent of allele frequencies or the nature of genetic effects. The universal constraints are that  $V_{E(o)}$  and  $V_{E(i)}$  must be positive;  $V_M$ ,  $C_{SS}$ ,  $C_{FS}$ ,  $C_{OS}$ , and  $C_{HS}$  must be greater than or equal to zero;  $C_{HS} \leq \frac{1}{2}C_{FS}$ ; and the magnitude of  $C_{OS}$  is limited by  $C_{SS}$  and  $C_{FS}$  (KELLY and ARATHI 2003).

Subject to these universal constraints, MLEs were obtained for the unconstrained model by finding the set of values for  $M_O$ ,  $\beta$ ,  $C_{SS}$ ,  $C_{FS}$ ,  $C_{OS}$ ,  $C_{HS}$ ,  $V_M$ ,  $V_{E(o)}$ , and  $V_{E(i)}$

that maximize  $l$ , the log-likelihood function. Here,  $M_0$  is the mean phenotype of outbred plants. The predicted mean phenotype of inbred plants is  $M_0 + \beta$ . I use the standard multivariate normal model for the log-likelihood function,

$$l = C - \frac{1}{2} \ln |V| - \frac{1}{2} (z - X\eta)^T V^{-1} (z - X\eta), \quad (8)$$

where  $V$  is the variance-covariance matrix of individual measurements,  $z$  is the vector of trait values,  $X$  is an "incidence matrix" for fixed effects,  $\eta$  is the vector of fixed effects, and  $C$  is a constant determined by the total sample size (SHAW 1987; SEARLE *et al.* 1992). The (co)variance parameters determine the numerical values of elements in  $V$  and  $\eta = [M_0, \beta]^T$ . Equation 8 assumes a multivariate normal distribution of trait values, but is quite robust to deviations from this distribution (SEARLE *et al.* 1992; J. K. KELLY, unpublished simulation results).

The maximum of  $l$  was determined iteratively by using the first derivatives of  $l$  with regard to each parameter and the asymptotic dispersion matrix. Following SEARLE *et al.* (1992), the vector of derivatives of  $l$  with regard to each of the fixed effects is

$$\frac{\partial l}{\partial \eta} = X^T V^{-1} (z - X\eta). \quad (9)$$

The derivative of  $l$  with regard to a particular variance component  $\theta$  is

$$\frac{\partial l}{\partial \theta} = -\frac{1}{2} \text{tr} \left[ V^{-1} \frac{\partial}{\partial \theta} (V) \right] + \frac{1}{2} (z - X\eta)^T V^{-1} \frac{\partial}{\partial \theta} (V) V^{-1} (z - X\eta), \quad (10)$$

where  $(\partial/\partial\theta)(V)$  is the derivative of the variance-covariance matrix with regard to  $\theta$  and  $\text{tr}[*]$  is the trace of a matrix.

Optimization was performed using two methods simultaneously, steepest ascent and scoring (SEARLE *et al.* 1992, Chap. 6; ELIASON 1993). By steepest ascent, the direction of movement within the parameter space is simply the vector of first derivatives. The direction for scoring is the product of the dispersion matrix and the derivatives vector. A log-likelihood value was calculated at 10 points along each directional vector using Equation 8. The highest of these 20 likelihood values was chosen to determine the set of parameters for the next iteration. In most cases, the optimization routine used the scoring vector in the first few iterations and the steepest ascent vector close to the optimum. These calculations, as well as the simulations described below, were performed by a series of computer programs similar to those described in KELLY and ARATHI (2003). These programs, written in the C programming language, are available from the author upon request.

The maximum log-likelihoods for each mutation model were also obtained using Equation 8, although

with different values in the variance-covariance matrix,  $V$ . The  $V$  matrix is a function of four genetic parameters in the unconstrained model ( $C_{SS}$ ,  $C_{FS}$ ,  $C_{OS}$ , and  $C_{HS}$ ), in addition to the environmental components. For model I,  $V$  contains only two genetic parameters,  $C_{SS}$  and  $h$ , in addition to the same set of environmental components. The various comparisons between relatives within  $V$  are functions of these two parameters:  $C_{OS} = hC_{SS}$ ,  $C_{FS} = 2h^2C_{SS}$ , and  $C_{HS} = h^2C_{SS}$ . For model II,  $V$  also depends on only two genetic parameters:  $C_{OS} = h^*C_{SS}$ ,  $C_{FS}$  is given by Equation 7, and  $C_{HS} = C_{FS}/2$ . For model III, the elements of  $V$  are functions of three genetic parameters ( $C_{SS}$ ,  $C_{FS}$ , and  $C_{OS}$ ) with the genetic covariance of half-sibs equated to  $C_{FS}/2$ . However, two of these parameters,  $C_{FS}$  and  $C_{OS}$ , are constrained in value:  $C_{OS} < C_{SS}/2$  and  $C_{FS} < C_{OS}$ . These model-specific constraints affect not only the numerical values within  $V$ , but also the derivatives of  $V$  with regard to each parameter [the  $(\partial/\partial\theta)(V)$  matrices of Equation 10].

Models I, II, and III were tested by comparing the maximum likelihood obtained under that hypothesis ( $l_0$ ) to the maximum likelihood obtained from the unconstrained model ( $l_1$ ). The model was rejected if the likelihood-ratio statistic  $2(l_1 - l_0)$  was greater than the appropriate critical value. It is generally assumed that, under the null hypothesis,  $2(l_1 - l_0)$  follows a chi-square distribution with the number of degrees of freedom equal to the number of parameters distinguishing the two models. This would suggest a critical value 5.99 (for  $P < 0.05$ ) for models I and II because the unconstrained model has two more parameters than either deleterious mutation model. There is no obvious choice for model III because only one parameter is eliminated ( $C_{HS} = \frac{1}{2}C_{FS}$ ) while other parameters are constrained in value ( $C_{OS} < \frac{1}{2}C_{SS}$ ,  $C_{FS} < C_{OS}$ ). These inequalities do not translate in a simple way into degrees of freedom. Even the critical values for models I and II may be suspect because the universal constraints apply to the unconstrained model (see SHAW *et al.* 1998). For these reasons, I employ a parametric bootstrapping routine to evaluate the significance of likelihood-ratio tests.

The parametric bootstrap involves repeated simulation of the data using the parameter estimates from the relevant mutational model (I, II, or III). Genotypic values for the parental lines are first sampled (from a normal distribution with variance  $C_{SS}$ ) and the progeny groups are then formed conditional on these parental values (and the values of the various parameter estimates). The numbers of individuals in each family are equivalent to the actual sample sizes from the experiment. Once the simulated data set is created,  $l_1$  and  $l_0$  are determined by applying the maximum-likelihood programs for the unconstrained and deleterious mutation models, respectively. I simulated 1000 data sets per test and the various models were fitted to each data set. The distribution of values for  $2(l_1 - l_0)$ , across simulated data sets, estimates the sampling distribution of the like-



TABLE 2  
Simulation results for selected parameter values using the parametric bootstrapping programs

True parameter values: model I	Critical value for LR	Median and percentiles for estimates		
		$U$	$h$	$s$
$U = 1.0, h = 0.2, s = 0.05, V_{E(o)} = V_{E(i)} = C_{SS}$	5.01	1.00 (0.73, 1.31)	0.20 (0.16, 0.23)	0.049 (0.040, 0.059)
$U = 1.0, h = 0.4, s = 0.05, V_{E(o)} = V_{E(i)} = C_{SS}$	5.22	0.99 (0.67, 1.81)	0.40 (0.36, 0.44)	0.049 (0.027, 0.073)
$U = 1.0, h = 0.1, s = 0.05, V_{E(o)} = V_{E(i)} = C_{SS}$	3.77	1.04 (0.62, 1.50)	0.10 (0.07, 0.14)	0.050 (0.042, 0.059)
$U = 1.0, h = 0.2, s = 0.05, V_{E(o)} = V_{E(i)} = 3C_{SS}$	3.88	0.99 (0.54, 1.72)	0.20 (0.13, 0.27)	0.049 (0.035, 0.067)
$U = 1.0, h = 0.2, s = 0.005, V_{E(o)} = V_{E(i)} = C_{SS}$	4.62	0.98 (0.73, 1.27)	0.20 (0.16, 0.23)	0.005 (0.004, 0.006)

True parameter values: model II	Critical value for LR	$U$	$K$	$\bar{s}$
$U = 1, K = -10, \bar{s} = 0.05, V_{E(o)} = V_{E(i)} = C_{SS}$	6.15	1.00 (0.73, 1.31)	-10.3 (-13.1, -8.2)	0.048 (0.039, 0.060)
$U = 1, K = -5, \bar{s} = 0.05, V_{E(o)} = V_{E(i)} = C_{SS}$	6.91	1.01 (0.80, 1.43)	-5.06 (-6.05, -4.10)	0.049 (0.038, 0.066)
$U = 1, K = -15, \bar{s} = 0.05, V_{E(o)} = V_{E(i)} = C_{SS}^{**}$	3.54	1.03 (0.46, 1.61)	-15.1 (-29.3, -10.5)	0.049 (0.030, 0.063)
$U = 1, K = -10, \bar{s} = 0.05, V_{E(o)} = V_{E(i)} = 3C_{SS}^{**}$	3.58	1.00 (0.52, 1.63)	-10.4 (-15.3, -7.7)	0.047 (0.036, 0.057)
$U = 1, K = -100, \bar{s} = 0.005, V_{E(o)} = V_{E(i)} = C_{SS}$	6.24	1.01 (0.75, 1.30)	-101 (-124, -85)	0.005 (0.004, 0.006)

For each parameter set, we used the distribution of likelihood-ratio (LR) values to determine an appropriate critical value for hypothesis testing (for that particular array of parameter values). The median estimate and the 10th and 90th percentiles of the distribution of 1000 estimates are given for the appropriate model. In all cases,  $\alpha = 1$  and  $V_M = 0$ . \*\*Distributions of parameter estimates exclude simulations that yield estimates out of the proper range.

likelihood-ratio statistic. The  $P$  value associated with a particular test is  $(M + 1)/(R + 1)$ , where  $M$  is the number of simulated values that exceed the actual likelihood-ratio value (calculated from the original data) and  $R$  is the number of simulations (DAVISON and HINKLEY 1997, p. 148).

In addition to their use in hypothesis testing, the parametric bootstrapping programs can be used to estimate the joint sampling distribution of our ML estimators for any mutational model in this experimental design. These simulations thus provide a means to evaluate properties such as bias and sampling variance. The distribution of simulated data sets demonstrates the range of possible experimental outcomes, including values for the likelihood-ratio statistic, when the null hypothesis is correct (when genetic variation conforms to the mutational model under consideration). I conducted simulation studies of both models I and II for a range of parameter sets to investigate (1) the null distribution of the likelihood-ratio statistic and (2) the statistical properties of the estimators. Each simulated data set was equivalent in terms of sample sizes to the experimental study.

## RESULTS

**Simulation study:** Table 2 summarizes a representative set of results from the simulation study. Cases 1–5 are simulations from models I, while cases 6–10 are from model II. Each simulated data set was first fit to the appropriate mutational model and then to the unconstrained model. The second column of Table 2 gives an empirically determined “critical value” for that set

of parameters. Ninety-five percent (950 of 1000) of the simulations yielded likelihood-ratio statistics that were less than this value. For model I, the empirical critical values differed among parameter sets but were always less than the chi-square value of 5.99. For model II, there was also variation among parameter sets, but empirical critical values were somewhat higher than those for model I.

The maximum-likelihood parameter estimates for both models are approximately median unbiased (Table 2). They exhibit relatively low sampling variances. The procedures do occasionally fail to yield estimates when certain variance components approach boundary values. However, this occurred in only a small fraction of simulations (not at all for most parameter combinations) and these cases are excluded from Table 2.

**Mimulus study:** Parameter estimates with standard errors (SEs) are given in Table 3 for each trait under the unconstrained model. The SEs are derived from the asymptotic dispersion matrix. These are only approximate and significantly nonzero estimates are routinely within 2 SE of zero (SHAW *et al.* 1998; KELLY and ARATHI 2003). The inbred genetic variance ( $C_{SS}$ ) is greater than the outbred variance ( $C_{FS}$ ) for each trait, although the magnitude of the difference varies greatly. For each trait,  $C_{FS} > C_{OS}$ . This implies that the point estimates for these components violate even the weak conditions (model III). Finally, maternal effects are evident for flower size but have no apparent effect on the male fitness traits. The  $V_M$  parameter is thus eliminated from the model for subsequent analyses of male fitness traits.

Maximum-likelihood values for each trait under model I are far lower than their associated values under

TABLE 3  
Maximum-likelihood estimates for each trait for the unconstrained model

Trait	$\beta$	$C_{SS}$	$C_{OS}$	$C_{FS}$	$C_{HS}$	$V_{E(i)}$	$V_{E(i)}$	$V_M$
Corolla width	-0.093 (0.008)	0.0088 (0.0016)	0.0030 (0.0009)	0.0048 (0.0015)	0.0024 (0.0009)	0.0119 (0.0005)	0.0126 (0.0006)	0.0023 (0.0012)
Pollen number	-0.447 (0.021)	0.0424 (0.0090)	0.0099 (0.0046)	0.0245 (0.0084)	0.0076 (0.0053)	0.133 (0.006)	0.106 (0.005)	0.00 (0.0065)
Ln (pollen number)	-0.732 (0.036)	0.1790 (0.0299)	0.0234 (0.0110)	0.0297 (0.0175)	0.0049 (0.0078)	0.236 (0.011)	0.411 (0.019)	0.00 (0.0157)
PSI	-0.308 (0.016)	0.036 (0.006)	0.0050 (0.0026)	0.0088 (0.0040)	0.0044 (0.0020)	0.0490 (0.0023)	0.0630 (0.0028)	0.00 (0.0035)
Ln(PSI)	-0.453 (0.028)	0.1065 (0.0176)	0.0146 (0.0068)	0.0173 (0.0105)	0.0087 (0.0046)	0.1416 (0.0066)	0.2302 (0.0104)	0.00 (0.0094)

Corolla width, pollen number, and PSI were divided by their respective outbred mean values prior to analysis. Standard errors (in parentheses) are derived from the asymptotic dispersion matrix.

the unconstrained model (Table 4). Each likelihood-ratio test is highly significant by both the parametric bootstrap and  $\chi^2$ . A  $P$  value  $<0.001$  for the parametric bootstrap implies that the observed likelihood-ratio (LR) statistic was greater than any of the 1000 simulated values. The MLEs for each trait are also given. Confidence intervals could be established for these estimates by resampling (bootstrapping) over extended families. I do not bother to justify the parameter estimates here because the model on which they are based is rejected.

There is generally better fit of the data to model II, although the model is still rejected for all traits (Table 5). As expected, the likelihood ratios are much lower for model III (Table 6). In fact, genetic variation in both Ln(pollen number) and Ln(PSI) are statistically consistent with the weak conditions. For all of the models, the likelihood ratios are lower for Ln-transformed male fitness values than for untransformed trait values.

DISCUSSION

This study clearly illustrates both the strengths and the faults of biometric estimates for mutational parameters. On the positive side, I was able to calibrate several different mutational models with an experiment of only moderate size (2345 plants). These models are surprisingly general, allowing variable mutational effects, maternal effects, and distinct environmental deviations for inbred and outbred plants. The parametric bootstrap simulations demonstrate that, if genetic variation is maintained according to one of the mutational models, the MLEs are surprisingly accurate (Table 2). Taken together, these observations suggest that biometric methods have the potential to yield accurate estimates for mutational parameters with much lower effort than is typically necessary for mutation-accumulation experiments.

The primary weakness associated with our parameter estimates is that they are based on models insufficient to explain the data. Likelihood-ratio tests reject models I and II for each of the traits (Tables 4 and 5). This is not surprising for flower size, as a previous study demonstrated that this trait is inconsistent with a deleterious mutation model of genetic variation (KELLY and WILLIS 2001). However, it is unexpected that models I and II were rejected for the male fitness traits. Evolutionary analyses of life-history (fitness) traits frequently assume that genetic variation is maintained by mutation-selection balance (CROW 1993; HOULE *et al.* 1996; DENG and LYNCH 1997). Our results call this assumption into question.

The rejection of mutation-selection balance for male fitness traits is clearly tentative. Model III was not rejected for either Ln(pollen number) or Ln(PSI). This may reflect the fact that these “weak conditions” are consistent with a range of genetic architectures, at least within the scope of estimation error. However, it may

TABLE 4

Likelihood-ratio tests and parameter estimates for model I (fixed mutational effects) applied to each trait

Trait	LR	$P_{boot}$	$P[\chi^2]$	$U\alpha$	$h$	$s\alpha$
Corolla width	16.14	<0.001	<0.001	0.33	0.39	0.018
Pollen number	24.58	<0.001	<0.001	5.70	0.46	0.006
Ln(pollen number)	14.48	<0.001	<0.001	0.51	0.20	0.139
PSI	34.88	<0.001	<0.001	0.40	0.28	0.049
Ln(PSI)	19.28	<0.001	<0.001	0.38	0.23	0.124

Two  $P$  values are provided for each likelihood-ratio statistic, the first from the parametric bootstrap simulations and the second from the  $\chi^2$  distribution with 2 d.f.

also mean that variation is governed by a deleterious mutation model that is simply more complicated than models I or II (e.g., CABALLERO and KEIGHTLEY 1994; ZHANG *et al.* 2002). While our experiment is not sufficiently complex to both calibrate and test a more complicated mutational model, it is straightforward to expand the design to include more comparisons.

Models I and II fail because there is too much additive genetic variation in both flower size and male fitness traits. The genetic covariance of full siblings ( $C_{FS}$ ) is equal to the additive variance when the parents are fully inbred. The genetic covariance of half siblings ( $C_{HS}$ ) is equal to one-half the additive variance (APPENDIX A). Under the deleterious mutation models, the amount of additive variation relative to other components depends on the average dominance of deleterious mutations. For example, with  $h = 0.2$  in model I, the variance among fully homozygous genotypes ( $C_{SS}$ ) should be 12.5 times greater than the variance among outbred genotypes ( $C_{FS}$ ) and  $C_{OS}$  should be 2.5 times greater than  $C_{FS}$ .

The  $C_{SS}$  estimates are substantially greater than  $C_{FS}$  estimates, particularly for log-transformed male fitness traits. However,  $C_{OS}$  is invariably less than  $C_{FS}$  when the unconstrained model is fit to the data (Table 3). The low estimates of  $C_{OS}$  relative to  $C_{SS}$  indicate that, if variation is attributable to rare alleles, these alleles must be fairly recessive on average. However, the more recessive they are, the less additive variation there should be. Estimating  $h$  just from  $C_{OS}$  and  $C_{SS}$  in Table 3, we would expect the additive variances of the male fitness traits (as measured by  $C_{FS}$  and  $C_{HS}$ ) to be only  $\sim 20\%$  of their actual estimated values. An excess of additive genetic variation

has also been documented for several life-history traits of *Drosophila melanogaster* and used to reject mutation-selection balance as a sufficient explanation of variation (MUKAI and NAGANO 1983; TAKANO *et al.* 1987; CHARLESWORTH and HUGHES 2000).

The higher-than-expected estimates for  $C_{FS}$  and  $C_{HS}$  also explain why the estimated dominance coefficients for male fitness traits under models I and II (Tables 4 and 5) are greater than those obtained previously. WILLIS (1999a) used a subset of these same inbred lines to estimate the average dominance of deleterious mutations by the “regression method” (MUKAI *et al.* 1972; MUKAI and YAMAGUCHI 1974; CABALLERO *et al.* 1997). He randomly paired lines and mated them to produce  $F_1$  families. The regression of  $F_1$  family mean trait values onto the corresponding sum of mean values for their parental lines provides an estimate for the average dominance of deleterious alleles. By this method, WILLIS (1999a) estimated the average dominance to be  $\sim 0.1$  for both pollen number and pollen viability (both measurements log-transformed). In contrast, our estimates of  $h$  (from model I) are about twice as large (Table 4) and the estimates of  $\bar{h}$  (from model II) are 3.5 times greater (Table 5).

The comparison of  $F_1$  families with parental lines is actually contained within this design (Figure 1) and it is straightforward to show that the regression method is actually estimating  $C_{OS}/C_{SS}$ . If we calculate  $C_{OS}/C_{SS}$  from the estimates in Table 3, the values are quite close to those obtained by WILLIS (1999a): 0.13 for Ln(pollen) and 0.14 for Ln(PSI). In contrast, the maximum-likelihood estimates for  $h$  and  $\bar{h}$  (Tables 4 and 5) depend

TABLE 5

Likelihood-ratio tests and parameter estimates for model II applied to each trait

Trait	LR	$P_{boot}$	$P[\chi^2]$	$U\alpha$	$K\bar{s}$	$\bar{s}$	$\bar{h}$
Corolla width	15.28	<0.001	<0.001	0.35	-0.210	0.014	0.41
Pollen number	23.78	<0.001	<0.001	3.40	-0.115	0.010	0.45
Ln(pollen number)	9.06	0.012	0.011	0.94	-0.439	0.063	0.35
PSI	24.76	<0.001	<0.001	0.61	-0.334	0.029	0.38
Ln(PSI)	11.04	0.007	0.004	0.70	-0.389	0.059	0.36

not only on  $C_{OS}$  and  $C_{SS}$  but also on  $C_{FS}$  and  $C_{HS}$ . Maximum likelihood finds the best “compromise” given the relative magnitudes of all four variance components. This requires dominance coefficients to be adjusted upward to account for the higher-than-expected values for  $C_{FS}$  and  $C_{HS}$ . If intermediate frequency alleles make a substantial contribution to variation in fitness, the regression method ( $C_{OS}/C_{SS}$ ) may actually provide a more accurate estimate for the dominance of rare alleles than the procedure used here (J. K. KELLY, unpublished results).

**Assumptions:** Models I and II employ an equilibrium model to predict the frequencies of deleterious alleles as a function of mutational parameters. These frequencies are determined by the action of selection in nature (if the model is accurate), but the effects are estimated from variation measured in the greenhouse. This is noteworthy given that both the relationship between trait values and evolutionary fitness and the relationship between genotype and phenotype are likely to be different in the greenhouse than in the wild. Character heritabilities and genetic correlations may change when organisms are assayed in a novel environment (SERVICE and ROSE 1985). In addition, the magnitude of inbreeding depression may be substantially greater under natural conditions than in the relatively benevolent greenhouse environment (e.g., DUDASH 1990; CARR and EUBANKS 2002). This may lead to biased estimates of mutational parameters. This difficulty is abrogated, to some extent, by the inclusion of  $\alpha$  in the model (Table 1). This parameter effectively translates between the measured phenotype and evolutionary fitness.

The important question for this study is whether the change in growth environment is responsible for the poor fit of models I and II to the data. Could the transfer cause rejection of the mutational models even if the natural population is in mutation-selection equilibrium? It is difficult to provide a definitive answer to this question, but it seems unlikely given that the hypothesis tests do not impose stringent constraints on the values of mutational parameters. Instead, parameters such as  $h$  and  $s$  of model I are estimated from the data. Thus, we should not falsely reject the null model, even if mutational effects are different in the lab.

The allele frequencies at QTL are the key to the constraints. A false rejection of the mutational models might occur if alleles that are deleterious in nature become advantageous in the lab environment. Such alleles might increase in frequency and the resulting lab-adapted population would exhibit an excess of additive genetic variation relative to the field population. However, the opportunity for this kind of selection was greatly limited in this study. The inbred lines that constitute the parents of the breeding design were generated rapidly by self-fertilization, with random selection of progeny within families (WILLIS 1999a,b). Each line was initiated from a single family (see STUDY SPECIES AND METHODS). A rare allele, present within a single family, might fix within the line founded by that family. However, it could not

spread within the laboratory population as a whole because each family contributed to one and only one line. Adaptation to lab conditions might be a more serious concern for randomly mating populations that propagate themselves for a number of generations prior to estimation of genetic parameters (see MATOS *et al.* 2000; HOFFMANN *et al.* 2001; LINNEN *et al.* 2001; SGRO and PARTRIDGE 2001).

An alternative approach would be to let the natural population adapt to the lab environment prior to biometric analyses. This has the advantage that population allele frequencies will settle to values dictated by the selective conditions of the lab environment (the relevant values of  $U$ ,  $h$ , and  $s$  if variation in fitness is maintained in mutation-selection balance). Of course, any parameter estimates would be specific to the lab environment. Moreover, a satisfactory fit of the data to a mutational model would not conclusively indicate that natural variation is maintained by mutation-selection balance. Variation maintained by variable selection, in either space or time, or genotype-by-environment interaction, would likely be lost in the process of lab adaptation.

A second potential concern with this study is that the deleterious mutation models assume random mating in the ancestral population. While the Iron Mountain population of *M. guttatus* is primarily outcrossing (WILLIS 1993b), some self-fertilization does occur. Inbreeding should reduce the frequency of deleterious mutations relative to the random-mating expectation,  $\mu/(hs)$ . This would reduce the absolute value of the variance components but not their relative values (to a first approximation when deleterious alleles are rare). Because the relative values of variance components determine the sufficiency of the models in the likelihood-ratio tests, inbreeding is not the cause of failures for models I and II. In this vein, it is noteworthy that *M. guttatus* does exhibit large amounts of inbreeding depression in fitness-related traits despite its mating system (WILLIS 1993a,b; DUDASH and CARR 1998). Moreover, most of this inbreeding depression was not purged during the formation of the lines used for this study (WILLIS 1999b).

Scale of measurement is another important concern for analyses based on patterns of variation. Scale transformations, such as the logarithm or square root of measurements, are routinely used in quantitative genetics to “stabilize” the variance (WRIGHT 1952; LYNCH and WALSH 1998, Chap. 11). In the present study, log-transformation substantially improved the fit of male fitness traits to each of the deleterious mutation models (Tables 4–6). This may mean that loci affecting these traits combine in a way that is more nearly multiplicative than additive. However, it could also mean that log-transformation simply obscures deviations from the models.

The most appropriate scale for the application of the present methods is the one in which loci contribute in the most nearly additive way to trait variation. On this



TABLE 6

Likelihood-ratio values and their associated *P* values for model III applied to each trait

Trait	LR	<i>P</i> <sub>boot</sub>
Corolla width	6.04	0.02
Pollen number	12.06	<0.001
Ln(pollen number)	1.96	0.15
PSI	4.38	0.03
Ln(PSI)	0.58	0.38

scale, the mean phenotype should decline linearly with the inbreeding coefficient (WRIGHT 1951; KEMPTHORNE 1957). Several studies have attempted to determine the functional relationship between mean trait values and the level of inbreeding in *M. guttatus* (WILLIS 1993a; CARR and DUDASH 1997). However, in each of these studies, the effects of changing homozygosity are confounded with the effects of genetic purging. As a consequence, it is difficult to determine the best scale of measurement for estimating mutational parameters. More data in this area are clearly necessary.

**Concluding comments:** More than anything, this study illustrates the need to combine model evaluation (hypothesis testing) with parameter estimation in biometric studies. Three distinct deleterious mutation models were tested. The first two models were rejected by the data for all characters, primarily because of an excess of additive genetic variation relative to other variance components. Thus, while most of the estimates in Tables 4 and 5 are reasonable, they cannot be treated as unbiased. However, model III was not rejected for either of the log-transformed male fitness components (Table 6). This suggests that a more elaborate deleterious mutation model may prove sufficient.

How generally will mutation-selection balance prove sufficient as an explanation for variation in fitness components? This is clearly the issue that will confront future applications of the biometric method for estimating mutational parameters. The ubiquity of genetic variation is one of our most general observations of natural populations. The question of how evolutionary processes maintain this variation is one of the oldest in evolutionary biology. It remains one of the most important.

This article benefited from comments by Norm Slade, Scott Williamson, Miles Stearns, Liza Holeski, Tara Marriage, Bruce Walsh, and two anonymous reviewers. Tyler Ternes, Scott Jablonski, and H. S. Arathi helped to collect many of the measurements. This research was supported by funds from National Science Foundation grant DEB-9903758 and the Murphy Scholarship fund.

## LITERATURE CITED

- AKASHI, H., 1997 Codon bias evolution in *Drosophila*. Population genetics of mutation-selection drift. *Gene* **205**: 269–278.  
 BATEMAN, A. J., 1959 The viability of near normal irradiated chromosomes. *Int. J. Radiat. Biol.* **1**: 170–180.  
 CABALLERO, A., and P. D. KEIGHTLEY, 1994 A pleiotropic nonaddi-

- tive model of variation in quantitative traits. *Genetics* **138**: 883–900.  
 CABALLERO, A., P. D. KEIGHTLEY and M. TURELLI, 1997 Average dominance for polygenes: drawbacks of regression estimates. *Genetics* **147**: 1487–1490.  
 CARR, D. E., and M. R. DUDASH, 1997 The effects of five generations of enforced selfing on potential male and female function in *Mimulus guttatus*. *Evolution* **51**: 1797–1807.  
 CARR, D. E., and M. D. EUBANKS, 2002 Inbreeding alters resistance to insect herbivory and host plant quality in *Mimulus guttatus* (Scrophulariaceae). *Evolution* **56**: 22–30.  
 CARR, D. E., and C. B. FENSTER, 1994 Levels of genetic variation and covariation for *Mimulus* (Scrophulariaceae) floral traits. *Heredity* **72**: 606–618.  
 CHARLESWORTH, B., and K. A. HUGHES, 2000 The maintenance of genetic variation in life history traits, pp. 369–392 in *Evolutionary Genetics From Molecules to Morphology*, edited by R. S. SINGH and C. B. KRIMBAS. Cambridge University Press, Cambridge, UK.  
 CHARLESWORTH, B., D. CHARLESWORTH and M. T. MORGAN, 1990a Genetic loads and estimates of mutation-rates in highly inbred plant-populations. *Nature* **347**: 380–382.  
 CHARLESWORTH, D., M. T. MORGAN and B. CHARLESWORTH, 1990b Inbreeding depression, genetic load, and the evolution of outcrossing rates in a multilocus system with no linkage. *Evolution* **44**: 1469–1489.  
 CHARLESWORTH, B., M. T. MORGAN and D. CHARLESWORTH, 1993 The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**: 1289–1303.  
 COCKERHAM, C. C., and B. S. WEIR, 1984 Covariances of relatives stemming from a population undergoing mixed self and random mating. *Biometrics* **40**: 157–164.  
 CROW, J. F., 1987 Muller, Dobzhansky, and overdominance. *J. Hist. Biol.* **20**: 351–380.  
 CROW, J. F., 1993 Mutation, mean fitness, and genetic load. *Oxf. Surv. Evol. Biol.* **9**: 3–42.  
 DAVIDSON, A. C., and D. V. HINKLEY, 1997 *Bootstrap Methods and Their Application*. Cambridge University Press, Cambridge, UK.  
 DENG, H.-W., 1998 Characterizing deleterious mutations in outcrossing populations. *Genetics* **150**: 945–956.  
 DENG, H.-W., and M. LYNCH, 1996 Estimation of genomic mutation parameters in natural populations. *Genetics* **144**: 349–360.  
 DENG, H.-W., and M. LYNCH, 1997 Inbreeding depression and inferred deleterious mutation parameters in *Daphnia*. *Genetics* **147**: 147–155.  
 DENG, H. W., G. GAO and J.-L. LI, 2002 Estimation of deleterious genomic mutation parameters in natural populations by accounting for variable mutation effects across loci. *Genetics* **162**: 1487–1500.  
 DUDASH, M. R., 1990 Relative fitness of selfed and outcrossed progeny in a self-compatible, protandrous species, *Silene angularis* (Gentianaceae): a comparison in three environments. *Evolution* **44**: 1129–1139.  
 DUDASH, M. R., and D. E. CARR, 1998 Genetics underlying inbreeding depression in *Mimulus* with contrasting mating systems. *Nature* **393**: 682–684.  
 ELIASON, S. R., 1993 *Maximum Likelihood Estimation: Logic and Practice*. Sage Publications, London.  
 EYRE-WALKER, A., and P. D. KEIGHTLEY, 1999 High genomic deleterious mutation rates in hominids. *Nature* **397**: 344–347.  
 FALCONER, D. S., 1989 *An Introduction to Quantitative Genetics*. John Wiley & Sons, New York.  
 FENSTER, C. B., and K. RITLAND, 1994 Quantitative genetics of mating system divergence in the yellow monkeyflower species complex. *Heredity* **73**: 422–435.  
 GARCIA-DORADO, A., 1997 The rate and effects distribution of viable mutation in *Drosophila*: minimum distance estimation. *Evolution* **51**: 1130–1139.  
 HALDANE, J. B. S., 1927 A mathematical theory of natural and artificial selection. Part V. Selection and mutation. *Proc. Camb. Philos. Soc.* **23**: 838–844.  
 HARRIS, D. L., 1964 Genotypic covariances between inbred relatives. *Genetics* **50**: 1319–1348.  
 HOFFMANN, A. A., R. HALLAS, C. SINCLAIR and L. PARTRIDGE, 2001 Rapid loss of stress resistance in *Drosophila melanogaster* under adaptation to laboratory culture. *Evolution* **55**: 436–438.  
 HOULE, D., 1992 Comparing evolvability and variability of quantitative traits. *Genetics* **130**: 195–204.

- HOULE, D., B. MORIKAWA and M. LYNCH, 1996 Comparing mutational variabilities. *Genetics* **143**: 1467–1483.
- JACQUARD, A., 1974 *The Genetic Structure of Populations*. Springer-Verlag, New York.
- KEIGHTLEY, P. D., 1994 The distribution of mutation effects on viability in *Drosophila melanogaster*. *Genetics* **138**: 1315–1322.
- KEIGHTLEY, P. D., and A. EYRE-WALKER, 1999 Terumi Mukai and the riddle of deleterious mutation rates. *Genetics* **153**: 515–523.
- KELLY, J. K., 1999 An experimental method for evaluating the contribution of deleterious mutations to quantitative trait variation. *Genet. Res.* **73**: 263–273.
- KELLY, J. K., and H. S. ARATHI, 2003 Inbreeding and the genetic variance of floral traits in *Mimulus guttatus*. *Heredity* **90**: 77–83.
- KELLY, J. K., and J. H. WILLIS, 2001 Deleterious mutations and genetic variation for flower size in *Mimulus guttatus*. *Evolution* **55**: 937–942.
- KELLY, J. K., A. RASCH and S. KALISZ, 2002 A method to estimate pollen viability from pollen size variation. *Am. J. Bot.* **89**: 1021–1023.
- KEMPTHORNE, O., 1957 *An Introduction to Genetic Statistics*. Wiley, New York.
- KONDRASHOV, A. S., 1984 Deleterious mutations as an evolutionary factor. I. The advantage of recombination. *Genet. Res.* **44**: 199–217.
- KONDRASHOV, A. S., 1988 Deleterious mutations and the evolution of sexual reproduction. *Nature* **336**: 435–440.
- KONDRASHOV, A. S., 1995 Contamination of the genome by very slightly deleterious mutations: Why have we not died 100 times over? *J. Theor. Biol.* **175**: 583–594.
- KONDRASHOV, A. S., and J. F. CROW, 1993 A molecular approach to estimating the human deleterious mutation rate. *Hum. Mutat.* **2**: 229–234.
- LANDE, R., 1994 Risk of population extinction from fixation of new deleterious mutations. *Evolution* **48**: 1460–1469.
- LERNER, I. M., 1954 *Genetic Homeostasis*. John Wiley & Sons, New York.
- LEWONTIN, R. C., 1974 *The Genetic Basis of Evolutionary Change*. Columbia University Press, New York.
- LINNEN, C., M. TATAR and D. PROMISLOW, 2001 Cultural artifacts: a comparison of senescence in natural, laboratory-adapted and artificially selected lines of *Drosophila melanogaster*. *Evol. Ecol. Res.* **3**: 877–888.
- LYNCH, M., and B. WALSH, 1998 *Genetics and Analysis of Quantitative Characters*. Sinauer Associates, Sunderland, MA.
- LYNCH, M., J. CONERY and R. BURGER, 1995 Mutational meltdowns in sexual populations. *Evolution* **49**: 1067–1080.
- LYNCH, M., J. BLANCHARD, D. HOULE, T. KIBOTA, S. SCHULTZ *et al.*, 1999 Spontaneous deleterious mutation. *Evolution* **53**: 645–663.
- MACKAY, T. F. C., R. LYMAN and M. S. JACKSON, 1992 Effects of P elements on quantitative traits in *Drosophila melanogaster*. *Genetics* **130**: 315–332.
- MACNAIR, M. R., and Q. J. CUMBES, 1989 The genetic architecture of interspecific variation in *Mimulus*. *Genetics* **122**: 211–222.
- MATOS, M., M. R. ROSE, M. T. R. PITE, C. REGO and T. AVELAR, 2000 Adaptation to the laboratory environment in *Drosophila subobscura*. *J. Evol. Biol.* **13**: 9–19.
- MORTON, N. E., J. F. CROW and H. J. MULLER, 1956 An estimate of the mutational damage in man from data on consanguineous marriages. *Proc. Natl. Acad. Sci. USA* **42**: 855–863.
- MUKAI, T., and S. NAGANO, 1983 The genetic structure of natural populations of *Drosophila melanogaster*. XVI. Excess of additive genetic variance of viability. *Genetics* **105**: 115–134.
- MUKAI, T., and O. YAMAGUCHI, 1974 The genetic structure of natural populations of *Drosophila melanogaster*. XI. Genetic variability in a local population. *Genetics* **76**: 339–366.
- MUKAI, T., S. CHIGUSA, L. E. METTLER and J. F. CROW, 1972 Mutation rate and dominance of genes affecting viability in *Drosophila melanogaster*. *Genetics* **72**: 335–355.
- MULLER, H. J., 1928 The measurement of gene mutation rate in *Drosophila*, its high variability, and its dependence upon temperature. *Genetics* **13**: 279–357.
- MULLER, H. J., 1950 Our load of mutations. *Am. J. Hum. Genet.* **2**: 111–176.
- PALSSON, S., 2002 Selection on a modifier of recombination rate due to linked deleterious mutations. *J. Hered.* **93**: 22–26.
- PAMILO, P., M. NEI and W. H. LI, 1987 Accumulation of mutations in sexual and asexual populations. *Genet. Res.* **49**: 135–146.
- RITLAND, K., 1989 Genetic differentiation, diversity, and inbreeding in the mountain monkeyflower (*Mimulus-Caespitiosus*) of the Washington Cascades. *Can. J. Bot. Rev. Can. Bot.* **67**: 2017–2024.
- ROBERTSON, A. W., A. DIAZ and M. R. MACNAIR, 1994 The quantitative genetics of floral characters in *Mimulus guttatus*. *Heredity* **72**: 300–311.
- SEARLE, S. R., G. CASELLA and C. E. MCCULLOCH, 1992 *Variance Components*. John Wiley & Sons, New York.
- SERVICE, P. M., and M. R. ROSE, 1985 Genetic covariation among life history components: the effect of novel environments. *Evolution* **39**: 943–945.
- SGRO, C. M., and L. PARTRIDGE, 2001 Laboratory adaptation of life history in *Drosophila*. *Am. Nat.* **158**: 657–658.
- SHAW, F. H., C. J. GEYER and R. G. SHAW, 2002 A comprehensive model of mutations affecting fitness and inferences for *Arabidopsis thaliana*. *Evolution* **56**: 453–463.
- SHAW, R. G., 1987 Maximum-likelihood approaches applied to quantitative genetics of natural populations. *Evolution* **41**: 812–826.
- SHAW, R. G., D. L. BYERS and F. H. SHAW, 1998 Genetic components of variation in *Nemophila menziesii* undergoing inbreeding: morphology and flowering time. *Genetics* **150**: 1649–1661.
- SIMMONS, M. J., and J. F. CROW, 1977 Mutations affecting fitness in *Drosophila* populations. *Annu. Rev. Genet.* **11**: 49–78.
- TAKANO, T., S. KUSAKABE and T. MUKAI, 1987 The genetic structure of natural populations of *Drosophila melanogaster*. XX. Comparisons of genotype environment interaction in viability between a northern and a southern population. *Genetics* **117**: 245–254.
- TURELLI, M., 1984 Heritable genetic variation via mutation-selection balance: Lerch's zeta meets the abdominal bristle. *Theor. Popul. Biol.* **25**: 138–193.
- UYENOYAMA, M. K., K. E. HOLSINGER and D. M. WALLER, 1992 Ecological and genetic factors directing the evolution of self-fertilization. *Oxf. Surv. Evol. Biol.* **8**: 327–381.
- WILLIS, J. H., 1993a Effects of different levels of inbreeding on fitness components in *Mimulus guttatus*. *Evolution* **47**: 864–876.
- WILLIS, J. H., 1993b Partial self fertilization and inbreeding depression in two populations of *Mimulus guttatus*. *Heredity* **71**: 145–154.
- WILLIS, J. H., 1996 Measures of phenotypic selection are biased by partial inbreeding. *Evolution* **50**: 1501–1511.
- WILLIS, J. H., 1999a Inbreeding load, average dominance, and the mutation rate for mildly deleterious alleles in *Mimulus guttatus*. *Genetics* **153**: 1885–1898.
- WILLIS, J. H., 1999b The role of genes of large effect on inbreeding depression in *Mimulus guttatus*. *Evolution* **53**: 1678–1691.
- WRIGHT, S., 1951 The genetical structure of populations. *Ann. Eugen.* **15**: 323–354.
- WRIGHT, S., 1952 The genetics of quantitative variability, pp. 5–41 in *Quantitative Inheritance*, edited by E. C. R. REEVE and C. H. WADDINGTON. Her Majesty's Stationary Office, London.
- ZHANG, X.-S., J. WANG and W. G. HILL, 2002 Pleiotropic model of maintenance of quantitative genetic variation at mutation-selection balance. *Genetics* **161**: 419–433.

Communicating editor: J. B. WALSH

## APPENDIX A

COCKERHAM and WEIR (1984) describe the methods necessary to express the observational components ( $C_{SS}$ ,  $C_{FS}$ ,  $C_{OS}$ , and  $C_{HS}$ ) in terms of causal components. For our case, where the parents are completely inbred, the appropriate equations are

$$C_{SS} = 2V_A + 4C_{AD} + V_{DI} \quad (A1)$$

$$C_{OS} = V_A + C_{AD} \quad (A2)$$

$$C_{FS} = V_A + V_D \quad (A3)$$

$$C_{HS} = \frac{1}{2} V_A, \tag{A4}$$

where  $V_A$  is the standard additive genetic variance,  $V_D$  is the standard dominance variance,  $V_{DI}$  is the inbred dominance variance (the variance of dominance effects with complete homozygosity), and  $C_{AD}$  is the covariance of additive and homozygous dominance effects (KELLY and ARATHI 2003).

Each of the causal components ( $V_A$ ,  $V_D$ ,  $V_{DI}$ , and  $C_{AD}$ ) can be written as a function of allele frequencies and genetic effects across loci (COCKERHAM and WEIR 1984; KELLY 1999). Using the parameterization of Table 1, I find that

$$V_A = \sum_i 2q_i(1 - q_i)\alpha_i^2 s_i^2 (\frac{1}{2} + (\frac{1}{2} - h_i)(2q_i - 1))^2 \tag{A5}$$

$$V_D = \sum_i 4q_i^2(1 - q_i)^2 \alpha_i^2 s_i^2 (\frac{1}{2} - h_i)^2 \tag{A6}$$

$$C_{AD} = \sum_i 2q_i(1 - q_i)(1 - 2q_i)\alpha_i^2 s_i^2 (\frac{1}{2} - h_i)(\frac{1}{2} + (\frac{1}{2} - h_i)(2q_i - 1)) \tag{A7}$$

$$V_{DI} = \sum_i 4q_i(1 - q_i)(1 - 2q_i)^2 \alpha_i^2 s_i^2 (\frac{1}{2} - h_i)^2 \tag{A8}$$

and

$$\beta = \sum_i -2q_i(1 - q_i)\alpha_i s_i (\frac{1}{2} - h_i). \tag{A9}$$

Since deleterious mutations should be rare, we expect that  $q_i \gg q_i^2 \gg q_i^3$ , etc. Equations 1a–1e in the text are obtained by neglecting terms containing higher powers of  $q_i$  in Equations A5–A9. Equations A5–A8 are then substituted into Equations A1–A4. Because the leading terms

in  $V_D$  are of order  $q_i^2$ , this term vanishes under rare-alleles models of inheritance. This explains why  $C_{HS} = C_{FS}/2$  in each of the deleterious mutation models.

APPENDIX B

We need to integrate over the probability density function of mutational effects,  $f(s)$ . Consider  $\beta$  for a particular locus,

$$\beta = - \int_0^\infty f(s) q(s) \alpha s (1 - 2h(s)) \partial s, \tag{A10}$$

where  $f(s)$  and  $h(s)$  are given in the text and

$$q(s) = \frac{\mu}{sh(s)}. \tag{A11}$$

Substituting  $f(s)$ ,  $q(s)$ , and  $h(s)$  into Equation A10, we find that

$$\beta = - \frac{2\mu\alpha}{s} \int_0^\infty \left[ \text{Exp}\left(-s\left(\frac{1}{s} + K\right)\right) - \text{Exp}\left(\frac{-s}{s}\right) \right] \partial s = 2\mu\alpha \left(1 - \frac{1}{1 + \bar{s}K}\right). \tag{A12}$$

Summing over loci, we obtain text Equation 6a. The single-locus integrals for the variance components are

$$C_{SS} = \frac{2\mu\alpha^2}{\bar{s}} \int_0^\infty \left[ s \text{Exp}\left(-s\left(\frac{1}{s} + K\right)\right) \right] \partial s = \frac{2\mu\alpha^2 \bar{s}}{(1 + \bar{s}K)^2} \tag{A13}$$

$$C_{OS} = \frac{\mu\alpha^2}{\bar{s}} \int_0^\infty \left[ s \text{Exp}\left(\frac{-s}{s}\right) \right] \partial s = \mu\alpha^2 \bar{s} \tag{A14}$$

$$C_{FS} = \frac{\mu\alpha^2}{\bar{s}} \int_0^\infty \left[ s \text{Exp}\left(-s\left(\frac{1}{s} - K\right)\right) \right] \partial s = \frac{\mu\alpha^2 \bar{s}}{(1 - \bar{s}K)^2} \tag{A15}$$

and  $C_{HS} = C_{FS}/2$ .

