The Effect of Overdominance on Characterizing Deleterious Mutations in Large Natural Populations

Jin-Long Li, Jian Li and Hong-Wen Deng

Osteoporosis Research Center and Department of Biomedical Sciences, Creighton University, Omaha, Nebraska 68131

Manuscript received April 20, 1998
Accepted for publication October 30, 1998

ABSTRACT

Alternatives to the mutation-accumulation approach have been developed to characterize deleterious genomic mutations. However, they all depend on the assumption that the standing genetic variation in natural populations is solely due to mutation-selection (M-S) balance and therefore that overdominance does not contribute to heterosis. Despite tremendous efforts, the extent to which this assumption is valid is unknown. With different degrees of violation of the M-S balance assumption in large equilibrium populations, we investigated the statistical properties and the robustness of these alternative methods in the presence of overdominance. We found that for dominant mutations, estimates for U (genomic mutation rate) will be biased upward and those for $\bar{h}$ (mean dominance coefficient) and $\bar{s}$ (mean selection coefficient), biased downward when additional overdominant mutations are present. However, the degree of bias is generally moderate and depends largely on the magnitude of the contribution of overdominant mutations to heterosis or genetic variation. This renders the estimates of U and $\bar{s}$ not always biased under variable mutation effects that, when working alone, cause U and $\bar{s}$ to be underestimated. The contributions to heterosis and genetic variation from overdominant mutations are monotonic but not linearly proportional to each other. Our results not only provide a basis for the correct inference of deleterious mutation parameters from natural populations, but also alleviate the biggest concern in applying the new approaches, thus paving the way for reliably estimating properties of deleterious mutations.

THE genome of any organism is subject to continuous bombardment of mutations, the majority of which are deleterious. Numerous theories based on the assumptions of deleterious genomic mutations have been developed to explain some fundamental phenomena in biology. These phenomena include (but are not limited to) the evolution of sex and recombination (Muller 1964; Kondrashov 1985, 1988; Charlesworth 1990), mate choice (Kirkpatrick and Ryan 1991), diploidy (Kondrashov and Crow 1991), and outbreeding mechanisms (Charlesworth and Charlesworth 1987). Theories also indicate that the parameters of deleterious genomic mutations determine the mutation load in populations at equilibrium (Haldane 1937; Kimura *et al.* 1963; Büger and Hofbauer 1994), the role of deleterious mutations in the extinction of small populations (Lande 1994; Lynch *et al.* 1995, 1996), the rate of input of genetic variance from deleterious mutations per generation (Deng and Lynch 1996, 1997), and the extent to which neutral molecular variation is reduced due to background selection (B. Charlesworth *et al.* 1993; D. Charlesworth *et al.* 1995; Hudson and Kaplan 1995). The validity of all these theories critically depends on the parameters of deleterious mutations.

Corresponding author: Hong-Wen Deng, Osteoporosis Research Center, Creighton University, 601 N. 30th St., Suite 6787, Omaha, NE 68131. E-mail: deng@creighton.edu

For the rest of the Introduction, the following definitions and distinctions between dominance and overdominance are in order. For a locus with alleles A and a, let the three genotypic values of fitness be

AA	Aa	aa
1	$1 - hs$	$1 - s$

Here, h is the dominance coefficient, where $h < 0$ implies overdominance, $h = 0.5$ implies additivity, and $0 \leq h \leq 1.0$ ($h \neq 0.5$) implies dominance. Note that we use "dominant" or "dominance" to refer to cases of complete dominance and partial dominance. Mutations with (over)dominant effects are referred to as (over)-dominant mutations. Deleterious genomic mutations generally refer to dominant mutations. Dominance is compatible with mutation-selection (M-S) balance; overdominance essentially encompasses all kinds of balancing selection at the allelic level (Deng *et al.* 1998a).

Three essential parameters of deleterious genomic mutations are (1) the genomic mutation rate (U), (2) the mean selection coefficient ($\bar{s}$), and (3) the mean dominance coefficient ($\bar{h}$). For the three essential parameters, there are now three approaches for estimation:

1. The mutation-accumulation (M-A) approach (Bateman 1959; Mukai 1964; Mukai *et al.* 1972): This technique estimates U and s. Most estimates have come from this approach applied to *Drosophila melanogaster* (Mukai 1979; Crow and Simmons 1983;

Keightley 1994, 1996) and have been very hard to acquire, requiring large and long-term M-A and special chromosomal constructs or inbred/asexual lines. The data from M-A can also be analyzed by the maximum-likelihood method (Keightley 1994) or the minimum-distance method (Garcia-Dorado 1997).

2. The inbreeding depression approach (Morton *et al.* 1956; Charlesworth *et al.* 1990): Requiring a $\bar{h}$ value that must be assumed or that cannot be estimated without bias (Caballero *et al.* 1997; Deng and Fu 1998; Deng 1998a), this technique *per se* estimates U only. In the highly selfing annual plants Leavenworthia (Charlesworth *et al.* 1994) and Amsinckia (Johnston and Schoen 1995), U estimates from this approach are in line with earlier ones from M-A in Drosophila, suggesting high deleterious genomic mutation rates.
3. The fitness moments approach (Deng and Lynch 1996, 1997; Deng 1998b): This approach estimates U, h, and s. For two outcrossing species of cyclical parthenogenetic Daphnia (a freshwater microcrustacean), preliminary estimates by this approach generally agree with earlier ones from other species (Deng and Lynch 1997) and those from the direct M-A approach in Daphnia (Lynch 1985; Lynch *et al.* 1998).

The last two approaches depend on the change in mean (and genetic variance) of fitness traits upon only one generation of mating in large selfing or outcrossing populations. In comparison, the first approach is much more time-consuming and requires many generations of M-A. None of the current experimental designs and statistical methods can estimate mutation parameters without bias. Under a number of biologically plausible conditions, the statistical properties of the above three approaches were compared (Deng and Fu 1998). We found that, generally speaking, the third approach has the best statistical properties as reflected by the minimum mean square error (MSE). MSE is a composite index of both bias and sampling error for biased estimates.

An essential assumption common to the last two approaches is that all the genetic variation in the study population is maintained under M-S equilibrium. Accordingly, changes in the mean and genetic variance of fitness (or its components) upon inbreeding or outcrossing are solely due to deleterious dominant mutations maintained by M-S balance. Even in large populations, despite tremendous efforts (*e.g.*, Houle 1989, 1994; Houle *et al.* 1996; Charlesworth and Hughes 1998; Deng 1998a; Deng *et al.* 1998a), the validity of this assumption is unknown. In large populations, alternatives to M-S balance, such as functional overdominance or overdominance induced by fluctuating selection, can in principle maintain polymorphisms, although no strong case has emerged for their generality (Deng and Lynch 1996).

The robustness of the approaches applied to natural populations has been investigated under a range of biologically plausible conditions, such as variable and/or epistatic mutation effects, etc. (Charlesworth *et al.* 1990; Deng and Lynch 1996, 1997; Deng and Fu 1998; Deng 1998b). Generally speaking, U and $\bar{h}$ are underestimated and $\bar{s}$ is overestimated. The direction and the magnitude of the bias revealed may provide a numerical basis for the close inference of deleterious genomic mutations. However, estimation under violation of the M-S balance assumption has never been investigated. It is intuitive that violation of the M-S balance assumption will result in biased estimates (Drake *et al.* 1998). However, a critical issue is, What are the statistical properties (the degree of bias and sampling variance, especially the bias) under different degrees of violation of the M-S balance assumption?

The M-S balance assumption can be violated in several scenarios, such as in small populations subject to random genetic drift or in large populations subject to balancing selection due to functional overdominance and/or fluctuating selection at the allelic level. Each scenario deserves careful consideration and thus separate treatment. The two approaches applicable to natural populations were originally devised for large populations at approximate equilibrium. Hence, we investigate estimation in large natural populations with genetic variance maintained by either M-S balance or balancing selection, and with inbreeding depression caused by either dominant or overdominant mutations. The study is conducted by computer simulations using algorithms we devised previously (Deng 1998a) and those we devise here. Other scenarios will be fully investigated in future studies by employing iterative algorithms (*i.e.*, Lynch *et al.* 1995, 1996) to construct populations (in linkage disequilibrium) of various finite sizes.

The experimental designs to characterize deleterious genomic mutations are different depending on the study population's mating type (Morton *et al.* 1956; Charlesworth *et al.* 1990; Deng and Lynch 1996). In outcrossing populations, the outcrossed parents from natural populations are selfed to obtain selfed progeny. In selfing populations, selfed parents from natural populations are outcrossed to obtain outcrossed progeny.

In this article, we first outline the simulations and develop the associated analytical derivations in outcrossing and selfing populations. Then we present the simulation results in these two types of populations for the fitness moments approach and the inbreeding depression approach for both constant and variable mutation effects. Finally, we discuss the implications of our current results for characterizing deleterious genomic mutations from natural populations.

SIMULATIONS

The direction and the magnitude of the bias under balancing selection with overdominance are of particular interest to geneticists. To focus on this, we assume

that genotypic values are measured accurately. In reality, this would require that each genotype be clonally replicated and assayed a large number of times. Ignoring measurement error for genotypic values reduces the sampling error of estimates, but is unlikely to bias either the estimation or the comparison of the techniques, assuming that the same number of genotypes would be handled experimentally. This is supported by our previous investigations (Deng and Lynch 1996; Deng and Fu 1998; Deng *et al.* 1998b). In outcrossing populations, inbreeding (such as sib mating) experiments can be performed for estimation, and selfing is not required (Deng 1998b). To apply the fitness moments approach (Deng and Lynch 1996, 1997), we found (Deng 1998b) that for a given sample size, sampling one selfed progeny is generally more efficient than sampling more selfed progenies from each selfing family. Therefore, for outcrossing populations, selfing experiments in which only one selfed progeny is sampled from each parent are simulated for applying the fitness moments approach.

Large outcrossing populations at equilibrium are constructed with some dominant loci maintained under M-S balance and other overdominant loci maintained by balancing selection. In large selfing populations, overdominance does not contribute to the maintenance of genetic variability (because of constant exposure to the homozygous state under selfing), and mutations of overdominant effects are also maintained by M-S balance (Charlesworth *et al.* 1990; Deng 1998a). For both outcrossing and selfing populations, we study first constant, then variable, mutation effects for dominant mutations. For overdominant mutations, we assume that their effects are constant across loci. This treatment may be at least partially justified by the facts that (1) no theoretical and empirical evidence bearing on the genetic effects across overdominant loci exists and (2) what concerns geneticists most is the estimation under different contributions of overdominant loci to heterosis and standing genetic variation in populations, irrespective of their constant or variable effects. Here, heterosis will refer both to inbreeding depression in outcrossing populations and to outbreeding enhancement in inbred populations. The investigation of the methods under their respective assumptions with constant fitness effects can serve as a starting point for comparison with more realistic situations investigated later in this and future studies.

Mutation effects on fitness across all loci are assumed to be multiplicative throughout, an assumption that is biologically plausible (Morton *et al.* 1956; Crow 1986; Craddock *et al.* 1995; Fu and Ritland 1996) and assumed in the original development of the approaches applied to natural populations (Morton *et al.* 1956; Charlesworth *et al.* 1990; Deng and Lynch 1996). Simulations and algorithms are outlined or developed for outcrossing populations and for selfing populations in the following sections.

Outcrossing populations: *Loci of constant dominant mutation effects mixed with overdominant loci:* At dominant loci at M-S balance, the number of mutations per individual (after selection, all in the heterozygous state) is Poisson distributed with an expectation of $\overline{n} = U/(hs)$ (Deng and Lynch 1996). The population is assumed to be random mating and at linkage equilibrium. Throughout, h and s generally refer to the dominance and selection coefficients of deleterious genomic mutations. In each situation, simulations are performed for different sets of parameters. For each parameter set, K individuals are randomly sampled from both the outcrossed parental and selfed progeny generations (Deng and Lynch 1996; Deng 1998b). Unless otherwise specified, $K = 200$ for outcrossing populations. The total number of genotypes employed in an experiment for the fitness moments approach in outcrossing populations is then 400. For a genotype with n dominant mutations (randomly determined from the Poisson distribution) from the outcrossed parental generation, the fitness is

$$W(n) = W_{\max}\prod_{i=1}^{n}(1 - h_i s_i),$$

where h_i and s_i are the dominance and selection coefficients of the ith locus with mutations. They are assumed to be constant initially and made variable later. $W_{\max}$ is the fitness of a genotype that is free of segregating deleterious genomic mutations in the experimental environment where the measurements are taken. This parameter serves as a scaling factor so that fitness can be on any scale instead of from 0 to 1.

For a genotype sampled from the selfed progeny generation, the fitness is

$$W(n_1,n_2) = W_{\max}\prod_{i=1}^{n_1}(1 - h_i s_i)\prod_{j=n_1+1}^{n_2+n_1}(1 - s_j),$$

where n_1 and n_2 are, respectively, the numbers of loci with mutations in heterozygous and homozygous states. n_1 and n_2 are determined from two levels of random sampling: (1) A number (n) of loci is randomly determined from the Poisson distribution with mean $\overline{n} = U/(hs)$. (2) Each of these n loci has a probability of 1/4 in the selfed progeny of being homozygous for the normal A allele, a probability of 1/2 of being a heterozygote Aa, and a probability of 1/4 of being a homozygote aa. After the genotypic status of each locus is determined as above, n_1 and n_2 are, respectively, the sum of loci heterozygous and homozygous for mutations.

Now consider the overall individual fitness with N polymorphic overdominant loci in the genome in addition to those dominant loci at M-S balance. At an overdominant locus with effect $h_o < 0$ and s_o in large populations, the equilibrium frequency of the more fit allele B is $p = (h_o - 1)/(2h_o - 1)$ (Crow 1986) and that of the less fit allele b is $q = h_o/(2h_o - 1)$. With N such additional overdominant polymorphic loci in the popu-

lation, the overall fitness of a random parental individual now becomes

$$W(n,n_3,n_4) = W_{\max}[\prod_{i=1}^{n}(1 - h_i s_i)](1 - h_o s_o)^{n_3}(1 - s_o)^{n_4},$$

where n_3 and n_4 are, respectively, the numbers of overdominant loci with genotypes *Bb* and *bb* in the genome of this individual, and n is defined earlier for the dominant loci. n_3 and n_4 are determined by random sampling, N times, from the trinomial distribution, with genotypes *Bb* and *bb* having the frequencies of $2pq$ and q^2, respectively. Recall that the population is assumed to be randomly mating. Different N values are assumed in simulations.

Upon selfing, the overall fitness of a selfed progeny whose parent has n_3 overdominant loci with the *Bb* genotype and n_4 overdominant loci with the *bb* genotype, is

$$W(n_1,n_2,n_4,n_5,n_6) = [W_{\max}\prod_{i=1}^{n_1}(1 - h_i s_i)\prod_{j=n_1+1}^{n_2+n_1}(1 - s_j)] \times (1 - h_o s_o)^{n_5}(1 - s_o)^{n_4+n_6},$$

where the term in the brackets has been explained earlier for the dominant loci. n_5 and n_6 are the number of loci in the selfed progeny heterozygous (*Bb*) and homozygous (*bb*) at the n_3 heterozygous overdominant loci of the parent. They are determined by random segregation during selfing of the parent, in the same manner as n_1 and n_2.

Once the desired samples of K individuals from the parent and selfed progeny generations are simulated, we estimate the parameters of deleterious genomic mutations on the basis of the assumption of pure dominant mutations maintained under M-S balance (Deng and Lynch 1997). Let $\overline{w}_o$ and σ_o^2 denote the mean and genetic variance of fitness in the outcrossed parental generation, respectively; and let $\overline{w}_s$ and σ_s^2 denote the corresponding values among the selfed progeny generation, respectively. These can be computed easily from simulated data (Deng 1998b). We define x, y, and z as follows:

$$x = \ln\left(\frac{\sigma_o^2}{\overline{w}_o^2} + 1\right), \quad y = \ln\left(\frac{\overline{w}_s}{\overline{w}_o}\right), \quad z = \ln\left(\frac{\sigma_s^2}{\overline{w}_s^2} + 1\right). \tag{1}$$

Then

$$h = \frac{1}{2\sqrt{z/x - \frac{1}{2}}} \tag{2a}$$

$$U = \frac{4hy}{2h - 1} \tag{2b}$$

$$s = \frac{x}{Uh}. \tag{2c}$$

If a value of h is assumed by external knowledge or estimated by other experimental designs and estimation methods, U can then be estimated from the change in the mean fitness upon selfing by Equation 2b (Morton *et al.* 1956). The method of Deng (1998a) is employed to estimate h. Unlike Mukai's method (Mukai *et al.* 1972), Deng's method does not require construction of homozygous lines. When applied to outcrossing populations, it can achieve about the same quality of estimation as Mukai's method. The data needed are the genotypic value of the parent fitness (w), and the mean genotypic fitness value (z) of the multiple selfed progeny within each selfed family. Let $t = 4z - 2w$. Then

$$h = \frac{\mathrm{Cov}(w,t)}{\mathrm{Var}(t)}. \tag{3}$$

For investigating the inbreeding depression approach in outcrossing populations, a set of M selfing families, each having S selfed progeny, is simulated as above to estimate h (Equation 3). This set of simulated selfing families and another set of L selfing families (each with one selfing parent and one selfed offspring) are employed to estimate inbreeding depression and then U (Equation 2b). Unless otherwise specified, $M = 10$, $S = 40$, and $L = 20$. This choice of parameters is shown to be efficient on the basis of our previous investigation (Deng 1998a; Deng and Fu 1998). The total number of genotypes employed in an experiment is then 450.

Two aspects of overdominant mutations concern geneticists most and are directly relevant to characterizing deleterious genomic mutations from natural populations. One is the contribution of dominant mutations to heterosis (the mean fitness of the outcrossed generation to the inbred generation) relative to that of overdominant mutations. The other is the magnitude of genetic variation due to dominant mutations maintained under M-S balance relative to that due to overdominant mutations maintained by balancing selection.

The contribution to the total heterosis upon selfing from the dominant mutations can be measured by the index

$$\alpha = \frac{E(\ln\,(w_{dp}/w_{do}))}{E(\ln\,(w_{tp}/w_{to}))} = \frac{E(\ln\,w_{dp}) - E(\ln\,w_{do})}{E(\ln\,w_{tp}) - E(\ln\,w_{to})}, \tag{4}$$

where w_{dp} and w_{do} are, respectively, the fitness in the parent and selfed offspring generations (denoted by p and o, respectively, in the second term of the subscript) if there were only dominant (d in the first term of the subscript) mutations in the genome. w_{tp} and w_{to} are, respectively, the fitness in the parent and selfed offspring generations if the total mutation effects (including both dominant and overdominant mutations, denoted by t in the first term of the subscript) are considered. E denotes the mathematical expectation. The derivation of the four expectation terms in Equation 4 is technical and tedious; thus, we present them in appendix a (Equations A1–A4).

The index α plays an important role. Compared with a similar index (α, constructed on the original fitness

scale) of Deng (1998a), α here represents the proportion of heterosis on the log fitness scale that is attributable to dominant mutations. Therefore, α ranges from 0 to 1. If $\alpha = 1$, the sole cause of heterosis is dominance; if $\alpha = 0$, it is overdominance. The smaller the α, the larger the contribution to heterosis from overdominant mutations.

To measure the magnitude of genetic variation from dominant mutations maintained under M-S balance relative to that from overdominant mutations maintained by balancing selection, we define the index

$$\beta = \frac{\sigma^2(\ln w_{dp})}{\sigma^2(\ln w_{tp})}. \quad (5)$$

β is the proportion of the standing genetic variation on the log fitness scale in the parental generation that is attributable to dominant mutations maintained under M-S balance; accordingly, $1 - \beta$ is that attributable to overdominant mutations maintained by balancing selection. β ranges from 0 to 1. The smaller the β, the larger the contribution to the standing genetic variation from overdominance. The numerator and denominator of Equation 5 can be expressed in terms of parameters for dominant and overdominant mutations (Equations A6 and A8 in appendix a).

Dominant loci with variable mutation effects mixed with overdominant loci: Deleterious mutation effects h_i and s_i, across loci are unlikely to be constant. For example, s_i may vary anywhere from 0 (neutral mutation) to 1 (lethal mutation). The rate of mutations with different effects may also vary so that mutations of smaller effects may occur at higher rates. To evaluate the direction and the magnitude of bias introduced jointly by variable mutation effects and overdominant mutations, as in Deng and Lynch (1996), we adopt an exponentially distributed mutation rate for mutations of variable effect s_i:

$$\mu(s_i) = \frac{1}{\bar{s}}\exp(-s_i/\bar{s}). \quad (6a)$$

Also we let

$$h(s_i) = \frac{1}{2}\exp(-13s_i). \quad (6b)$$

By Equation 6b, h_i and s_i are correlated. These are in rough accordance with the few available data (Gregory 1965; Crow and Simmons 1983; Mackay *et al.* 1992; Keightley 1994) and with biochemical arguments (Kacser and Burns 1981). In Equation 6b, $\bar{h} = 0.36$ when $\bar{s} = 0.03$, $h \to 0.5$ as $s \to 0$, and $h \to 0.0$ as $s \to 1.0$, all in rough accordance with the data (Crow and Simmons 1983). However, true mutational spectra may be such that the dominance of individual mutations is broadly scattered around such a function (Caballero and Keightley 1994). With variable effects for deleterious genomic mutations, the indices α (Equation 4) and β (Equation 6) can be constructed using the results in appendix a.

In simulations, we divide the entire range of s $(0 - 1)$ into 100 discrete classes of width 0.01. Within each class, mutations have constant effects (h_i and s_i). Each individual from the outcrossed parental generation in the simulation is assigned a number (n_i) of heterozygous mutations from the ith of these classes by drawing from a Poisson distribution with expectation $Up_i/(h_is_i)$, where p_i is the density of the mutational distribution in the ith class. For an individual from the selfed progeny generation, n_i's are first determined as above. Then for each of the n_i loci, the genotype is, as before, determined by randomly sampling from the trinomial probabilities so that probabilities for different genotypes are 1/4 for *AA*, 1/2 for *Aa*, and 1/4 for *aa*, respectively (due to random segregation during selfing of parents). This discrete treatment closely approximates the continuous distribution of mutation effects (H.-W. Deng, unpublished data).

Selfing populations: To estimate deleterious genomic mutations, selfed individuals from natural selfing populations are crossed randomly to obtain outcrossed progeny. In selfing populations, new mutations in the genome most likely follow a Poisson distribution, whether they involve dominant or overdominant mutations. In highly selfing populations, mutant alleles will be maintained by M-S balance, regardless of their (over)dominance (Deng 1998a). Hence, as in the dominant case, we assume that the number of loci with overdominant mutants (n_7), all in the homozygous state, per genome in selfing populations is Poisson distributed with mean $\bar{n}_o$ and constant effects h_o and s_o. If the genomic mutation rate to the overdominant (but less fit) allele *a* is U_o, it can be easily shown that at M-S equilibrium, $\bar{n}_o = U_o/(2s_o)$ (Charlesworth *et al.* 1990; Deng 1998a).

In each situation, a variable number K of individuals is randomly sampled from the selfed parental and outcrossed progeny generations, respectively. For a genotype with n dominant and n_7 overdominant mutations [randomly determined from the Poisson distribution with mean $U/(2s)$ and $U_o/(2s_o)$, respectively] from the selfed parental generation, the fitness is

$$W(n) = W_{max}(1 - s_o)^{n_7}\prod_{i=1}^{n}(1 - s_i).$$

For an outcrossed progeny resulting from crossing two selfed parents (with n_f, n_7 and n_m, n_8 homozygous loci for dominant and overdominant mutations, respectively, where the subscript f indicates female parent and m the male parent), its fitness is

$$W(n_f,n_m,n_7,n_8) = W_{max}(1 - h_os_o)^{n_7+n_8} \times \prod_{i=1}^{n_i}(1 - h_is_i)\prod_{j=1}^{n_m}(1 - h_js_j).$$

h_i and s_i are the dominance and selection coefficients

of the *i*th locus with dominant mutations. They are assumed to be constant initially and made variable later.

In selfing populations, the indices α and β defined in Equations 4 and 5 can be constructed from the derivations in appendix b for the constant and variable dominant mutation effects, respectively. In simulated populations, the genome contains both dominant and overdominant loci, all at M-S equilibrium. In the parental generation, the number of homozygous dominant loci in each individual is determined by random sampling from a Poisson distribution of mean $U/(2s)$, and the number for the overdominant loci is determined from a Poisson distribution with mean $\bar{n}_o$ $[=U_o/(2s_o)]$ (Charlesworth *et al.* 1990; Deng 1998a). In the outcrossed offspring generation, the number of dominant loci in each individual is sampled from a Poisson distribution of mean U/s, and the number of overdominant loci in each individual is determined from a Poisson distribution with mean U_o/s_o (Charlesworth *et al.* 1990; Deng 1998a), all in the heterozygous state. The variable dominant mutations are modeled by Equation 6 and are simulated by discrete classes of mutations, in a manner similar to that in outcrossing populations as described earlier.

Once the desired samples of K individuals from the selfed parent and the outcrossed progeny generations are simulated, the estimation developed on the basis of the assumption of pure dominant mutants maintained under M-S balance (Deng and Lynch 1996) is applied. Unless otherwise specified, $K = 200$ for selfing populations. The total sample size is then 400 for the application of the fitness moments approach. Let $\bar{w}_o$, σ_o^2 and $\bar{w}_s$, σ_s^2 be the mean and genetic variance of the fitness in the outcrossed progeny and selfed parental generations, respectively. Let x, y, z be defined as in Equation 1; then

$$h = \sqrt{\frac{x}{2z}}, \tag{7a}$$

$$U = \frac{2y}{2h - 1}, \tag{7b}$$

$$s = \frac{2z}{U}. \tag{7c}$$

To apply the inbreeding depression approach to estimate U (Charlesworth *et al.* 1990), the value for h must be assumed or estimated by other experimental designs and methods. Mukai's method (Mukai *et al.* 1972) is employed to estimate h. It estimates h approximately by the slope of the regression of the outcrossed-progeny fitness (x) on the fitness sum (y) of the two corresponding parental homozygotes:

$$h = \frac{\mathrm{Cov}(x, y)}{\mathrm{Var}(y)}.$$

Once h is estimated, U can be estimated by Equation 7b. A sample of 200 outcrossing families, each consisting of two selfed parents and one outcrossed offspring, is simulated to implement the inbreeding depression approach. The total number of genotypes employed in the experiment is 600.

In simulations, we arbitrarily let $W_{max} = 1$, as the values of W_{max} do not influence the estimation for the mutation parameters (Deng and Lynch 1996). For each set of parameters, we perform 500 simulations. Unless otherwise specified, in all the simulations presented (except with pure overdominant mutations in the genome, *i.e.*, when $\alpha = \beta = 0$), $U = 1.0$, $\bar{h} = 0.36$, and $\bar{s} = 0.03$, which are close to the most often cited values estimated by Mukai *et al.* (1972; Lynch *et al.* 1995). The experimental designs have been laid out earlier for different estimations in different populations. Results for other simulation parameters (*e.g.*, $U = 0.1$–4.0 and $\bar{s} = 0.01$–0.05) and experimental designs have also been performed. The results are similar and thus not presented. Because almost all the results are biased, the MSE is presented together with one standard deviation (SD) computed over the repeated simulations.

RESULTS

Outcrossing populations

Constant dominant mutation effects: *1. The fitness moments approach (Table 1):* With only deleterious dominant loci in the genome ($N = 0$ and $\alpha = \beta = 1$), the estimates for U, h, and s are unbiased. Recall that N is the number of polymorphic overdominant loci in the population and α and β are, respectively, the proportion of heterosis and genetic variation on the log fitness scale that is attributable to dominance mutations. With overdominant loci coexisting in the genome with deleterious dominant loci ($N > 0$ and $0 < \alpha, \beta < 1$), $\hat{U}$ ($\hat{}$ indicates an estimated value) is an overestimate, while h and s are underestimated. The degree of bias increases with increasing contributions from overdominance to heterosis (decreasing α) and to the standing genetic variation in the population (decreasing β). Generally, the bias is not dramatic so that estimates of the upper bound of U and lower bounds of h and s can be obtained, and these estimates are close to the true parameter values. All the sampling errors are quite small. Even with only overdominant mutations in the genome ($\alpha = \beta = 0$), estimates of U, $\bar{h}$, and $\bar{s}$ can still be obtained, although the parameter values do not exist for the dominant mutations. In this case, it is not incorrect to treat $\hat{U}$ as an upper limit for the true U of zero. This is understandable, because, upon selfing (or outcrossing in selfing populations), overdominant mutations will also cause mean and genetic variance of fitness to change, similar to those changes caused by dominant mutations. This will be similar in every case, and thus will not be repeated. The estimation bias is relatively more sensitive

TABLE 1

Characterizing constant deleterious genomic mutations in the presence of overdominant mutations with the fitness moments approach in outcrossing populations

h_o	s_o	N	α	β	$\hat{U}$	$\hat{h}$	$\hat{s}$
−0.1	0.03	1308	0.00	0.00	3.10 (0.45)	0.23 (0.02)	0.01 (0.00)
		581	0.20	0.71	3.11 (0.61)	0.30 (0.02)	0.02 (0.00)
					[2.19]	[0.06]	[0.01]
		218	0.39	0.87	1.97 (0.40)	0.33 (0.02)	0.02 (0.00)
					[1.04]	[0.04]	[0.01]
		97	0.59	0.94	1.52 (0.34)	0.35 (0.02)	0.02 (0.01)
					[0.63]	[0.03]	[0.01]
		36	0.80	0.98	1.23 (0.34)	0.35 (0.02)	0.03 (0.01)
					[0.40]	[0.02]	[0.01]
−0.2	0.03	701	0.00	0.00	7.06 (1.79)	0.33 (0.03)	0.01 (0.00)
		312	0.20	0.56	4.74 (1.44)	0.35 (0.03)	0.01 (0.00)
					[4.01]	[0.03]	[0.02]
		117	0.39	0.78	2.47 (0.71)	0.35 (0.02)	0.02 (0.00)
					[1.64]	[0.03]	[0.01]
		52	0.59	0.89	1.70 (0.44)	0.36 (0.02)	0.02 (0.01)
					[0.82]	[0.02]	[0.01]
		20	0.79	0.95	1.31 (0.34)	0.36 (0.02)	0.03 (0.01)
					[0.46]	[0.02]	[0.01]
−0.2	0.01	2118	0.00	0.00	7.24 (2.09)	0.33 (0.03)	0.00 (0.00)
		941	0.20	0.80	5.04 (1.43)	0.35 (0.02)	0.01 (0.00)
					[4.29]	[0.02]	[0.02]
		353	0.39	0.91	2.59 (0.68)	0.36 (0.02)	0.01 (0.00)
					[1.72]	[0.02]	[0.01]
		157	0.59	0.96	1.77 (0.43)	0.36 (0.02)	0.01 (0.00)
					[0.88]	[0.02]	[0.01]
		59	0.80	0.98	1.33 (0.37)	0.36 (0.02)	0.02 (0.01)
					[0.49]	[0.02]	[0.01]
0	0	0	1.00	1.00	1.05 (0.24)	0.36 (0.02)	0.03 (0.01)

Numbers within parentheses are standard deviations; those within brackets are the square roots of MSE (for biased estimates only). Unless otherwise specified, in all the simulations presented (except with pure overdominant mutations in the genome, *i.e.*, when $\alpha = \beta = 0.0$), $U = 1.0$, $\bar{h} = 0.36$, and $\bar{s} = 0.03$. N is determined by the magnitude of α and β, by the parameters for mutations, and by Equations of 4 and 5 and those in the appendices.

to a change of h_o than to a change of s_o. With a larger absolute value of h_o, the degree of bias increases.

2. The inbreeding depression approach (Table 2): With $N = 0$ and $\alpha = \beta = 1$, the estimates for U and h are nearly unbiased. With $N > 0$ and $0 < \alpha, \beta < 1$, U is generally overestimated, while h is underestimated. The degree of bias generally increases with decreasing α and β. Compared with the fitness moments approach, the bias is larger for $\hat{h}$ and smaller for $\hat{U}$. The smaller bias of $\hat{U}$ is largely due to the greatly underestimated $\hat{h}$. This can be understood from Equation 2b or Figure 1 in Deng and Fu (1998). Although the presence of overdominant mutations will tend to bias $\hat{U}$ upward, the bias will be greatly dampened by a greatly underestimated $\hat{h}$. However, the estimation of U suffers from large sampling errors, even though the number of genotypes employed (450) is larger than that for the fitness moments approach (400). When both sampling error and bias are considered, the estimation of U by the inbreeding depression approach is generally worse than that by the fitness moments approach, as reflected by the larger MSE. The statistical properties (mean and sampling variance) of $\hat{U}$ are relatively unstable with changes of a and β. This instability is largely due to the relatively small sample size employed. When overdominance contributes importantly to the heterosis and standing genetic variation in natural populations (with small α and β), $\hat{U}$ is unacceptable even as an estimate for the upper limit because of the large sampling error. $\hat{h}$ estimated by Deng's (1998b) method can serve well as a lower bound of the true h as evidenced by its small sampling error.

Variable dominant mutation effects: *The fitness moments approach (Table 3):* With $N = 0$ and $\alpha = \beta = 1$, U and $\bar{h}$ are underestimated and $\bar{s}$ is overestimated. With $N > 0$ and $0 < \alpha, \beta < 1$, $\hat{\bar{h}}$ is always biased downward, and the magnitude of bias and sampling variance do not change much with changing α and β. The degree of bias is relatively small so that $\hat{\bar{h}} \approx 0.67\ \bar{h}$. The small bias and sampling variance of $\hat{\bar{h}}$ render it an ideal esti-

TABLE 2

Estimates of h with Deng's method and U with the inbreeding depression approach to characterize constant dominant mutations in the presence of overdominant mutations in outcrossing populations

h_o	s_o	N	α	β	$\hat{h}$	$\hat{U}$
−0.1	0.03	1308	0.00	0.00	0.00 (0.09)	0.16 (0.13)
		581	0.20	0.71	0.17 (0.09) [0.21]	1.37 (1.53) [1.57]
		218	0.39	0.87	0.25 (0.08) [0.14]	1.45 (4.65) [4.67]
		97	0.59	0.94	0.29 (0.06) [0.09]	1.07 (0.76) [0.77]
		36	0.80	0.98	0.32 (0.04) [0.06]	0.98 (0.48) [0.48]
−0.2	0.03	701	0.00	0.00	0.01 (0.12)	0.44 (2.41)
		312	0.20	0.56	0.18 (0.11) [0.21]	1.36 (6.32) [6.33]
		117	0.39	0.78	0.26 (0.09) [0.13]	0.64 (15.03) [15.04]
		52	0.59	0.89	0.30 (0.07) [0.09]	1.40 (5.28) [5.30]
		20	0.79	0.95	0.32 (0.05) [0.06]	1.01 (0.54) [0.54]
−0.2	0.01	2118	0.00	0.00	0.00 (0.12)	0.31 (1.36)
		941	0.20	0.80	0.27 (0.08) [0.13]	1.90 (56.13) [56.14]
		353	0.39	0.91	0.30 (0.06) [0.09]	1.64 (2.78) [2.85]
		157	0.59	0.96	0.32 (0.05) [0.06]	1.34 (0.70) [0.78]
		59	0.80	0.98	0.33 (0.04) [0.04]	1.07 (0.48) [0.49]
0	0	0	1.00	1.00	0.34 (0.03)	0.88 (0.27)

See Table 1 legend for detailed explanation.

mate of the lower limit for the true $\bar{h}$, and it is close to the true parameter value. The bias of $\hat{U}$ and $\hat{\bar{s}}$ changes so that $\hat{U}$ and $\hat{\bar{s}}$ are not always biased. When α and β are relative large, so that overdominance does not contribute substantially to the heterosis and to the genetic variation in the population, U and $\hat{\bar{s}}$ are both underestimated. When α and β gradually decrease, so that overdominance contributes more to the heterosis and the standing genetic variation in the population, $\hat{U}$ and $\hat{\bar{s}}$ become unbiased and then overestimated. For the same magnitude of α or β, with different parameters h_o and s_o, the degree of bias for $\hat{U}$, $\hat{\bar{h}}$, and $\hat{\bar{s}}$ is different. This is also true throughout this study and is not repeated.

It should be noted that with different h_o and s_o parameters for overdominant mutations, the same α may correspond to a different β. This can be inferred from the corresponding Equations 4 and 5 and those in appendices a and b. It is also evident in every case as can be seen from the numerical values of Tables 1–8 for outcrossing and selfing populations and for constant and variable mutation effects. To illustrate the monotonic but nonlinear relationship between α and β, Figure 1 plots the values of α and β for constant and variable mutation effects in both outcrossing and selfing populations.

The inbreeding depression approach (Table 4): With $N = 0$ and $\alpha = \beta = 1$, the estimates for U and $\bar{h}$ are both biased downward. With $N > 0$ and $0 < \alpha, \beta < 1$, U is generally underestimated when α and β are relatively large and is only overestimated when α and β are quite small. However, the sampling variance of $\hat{U}$ is usually large. On the other hand, $\hat{\bar{h}}$ is always biased downward and the sampling variance is miniscule. With decreasing a and β, the degree of bias of $\hat{\bar{h}}$ increases. $\hat{\bar{h}}$ can serve reasonably well as a lower bound of the true $\bar{h}$.

Selfing populations

Constant dominant mutation effects: *The fitness moments approach (Table 5):* With $N = 0$ and $\alpha = \beta = 1$, the estimates for U, h, and s are unbiased. With $N > 0$ and $0 < \alpha, \beta < 1$, U is overestimated, while h and s are underestimated. The degree of bias increases with decreasing α and β. However, the bias is not so dramatic that the upper bound of U and lower bounds of h and s can be estimated, and that they are not wildly far away

TABLE 3

Characterizing variable deleterious genomic mutations in the presence of overdominant mutations with the fitness moments approach in outcrossing populations

h_o	s_o	N	α	β	$\hat{U}$	$\hat{\bar{h}}$	$\hat{\bar{s}}$
−0.1	0.03	2070	0.00	0.00	4.89 (0.73)	0.23 (0.02)	0.01 (0.02)
		920	0.20	0.52	2.90 (0.43) [1.94]	0.24 (0.02) [0.07]	0.02 (0.00) [0.01]
		345	0.40	0.74	1.45 (0.19) [0.49]	0.24 (0.02) [0.06]	0.03 (0.00) [0.00]
		153	0.60	0.87	0.97 (0.12) [0.13]	0.25 (0.02) [0.06]	0.05 (0.01) [0.01]
		58	0.80	0.95	0.71 (0.09) [0.30]	0.25 (0.02) [0.06]	0.02 (0.01) [0.02]
−0.2	0.03	1110	0.00	0.00	11.29 (3.03)	0.33 (0.03)	0.01 (0.00)
		493	0.20	0.37	4.47 (0.90) [3.58]	0.29 (0.02) [0.03]	0.01 (0.00) [0.01]
		185	0.40	0.61	1.85 (0.29) [0.89]	0.27 (0.02) [0.04]	0.02 (0.00) [0.01]
		82	0.60	0.78	1.10 (0.15) [0.18]	0.26 (0.02) [0.05]	0.03 (0.01) [0.02]
		31	0.80	0.90	0.75 (0.10) [0.27]	0.26 (0.02) [0.03]	0.04 (0.01) [0.02]
−0.2	0.01	3352	0.00	0.00	11.61 (2.98)	0.33 (0.03)	0.00 (0.00)
		1500	0.20	0.64	3.78 (0.57) [2.83]	0.27 (0.02) [0.04]	0.01 (0.00) [0.02]
		559	0.40	0.82	1.66 (0.23) [0.70]	0.26 (0.02) [0.05]	0.02 (0.00) [0.01]
		248	0.60	0.91	1.02 (0.13) [0.13]	0.25 (0.02) [0.05]	0.03 (0.01) [0.01]
		93	0.80	0.97	0.73 (0.10) [0.28]	0.25 (0.02) [0.05]	0.04 (0.01) [0.02]
0	0	0	1.00	1.00	0.41 (0.18) [0.62]	0.20 (0.00) [0.10]	0.11 (0.07) [0.11]

See Table 1 legend.

from the true parameter values. The estimation bias is not very sensitive to changes in h_o and s_o, especially for $\hat{h}$ and $\hat{s}$.

The inbreeding depression approach (Table 6): With $N = 0$ and $\alpha = \beta = 1$, the estimates for U and h are nearly unbiased. With $N > 0$ and $0 < \alpha, \beta < 1$, U is generally overestimated, while h is underestimated. The degree of bias generally increases with decreasing α and β. Compared with the fitness moments approach, the bias is larger for $\hat{h}$ and smaller for $\hat{U}$. The smaller bias of $\hat{U}$ is largely due to the greatly underestimated $\hat{h}$. This can be understood from Equation 2b or Figure 1 in Deng and Fu (1998). Although the presence of overdominant mutations will tend to bias $\hat{U}$ upward, the bias will be greatly dampened by a largely underestimated $\hat{h}$. Compared with outcrossing populations under constant mutation effects with a comparable sample size of genotypes, the sampling error for $\hat{U}$ is relatively small, and hence $\hat{U}$ can serve well as an estimate for the upper limit. $\hat{h}$ estimated by Mukai's method (Mukai *et al.* 1972) can also serve well as a lower bound of the true h as evidenced by its small sampling error.

Variable dominant mutation effects: *The fitness moments approach (Table 7):* With $N = 0$ and $\alpha = \beta = 1$, the estimates for U and $\bar{h}$ are biased downward and the estimates for $\bar{s}$ are biased upward. With $N > 0$ and $0 < \alpha, \beta < 1$, $\hat{\bar{h}}$ is always biased downward, and the magnitude of the bias increases slightly with decreasing α and β, while its sampling variance remains relatively stable. $\hat{\bar{h}}$ ranges from 0.77 $\bar{h}$ to 0.5 $\bar{h}$. The relatively small bias and sampling variance of $\hat{\bar{h}}$ render it an ideal estimate of the lower limit for $\bar{h}$. The direction and the magnitude of the bias of $\hat{U}$ and $\hat{\bar{s}}$ change so that $\hat{U}$ and $\hat{\bar{s}}$ are not always biased. When α and β are relatively large, so that overdominance does not contribute substantially to the heterosis and the standing genetic variation in the population, U and $\hat{\bar{s}}$ are both underestimates. When α and β gradually decrease, $\hat{U}$ and $\hat{\bar{s}}$ become unbiased and then overestimated. However, for $\hat{U}$ and $\hat{\bar{s}}$ to become biased upward, α and β need to be quite small ($\alpha < \sim 0.56$, $\beta < 0.84$) so that overdominance contributes substantially to heterosis and the standing genetic variation in the populations.

The inbreeding depression approach (Table 8): With $N = 0$ and $\alpha = \beta = 1$, the estimates for U and $\bar{h}$ are biased. With $N > 0$ and $0 < \alpha, \beta < 1$, U is generally underesti-

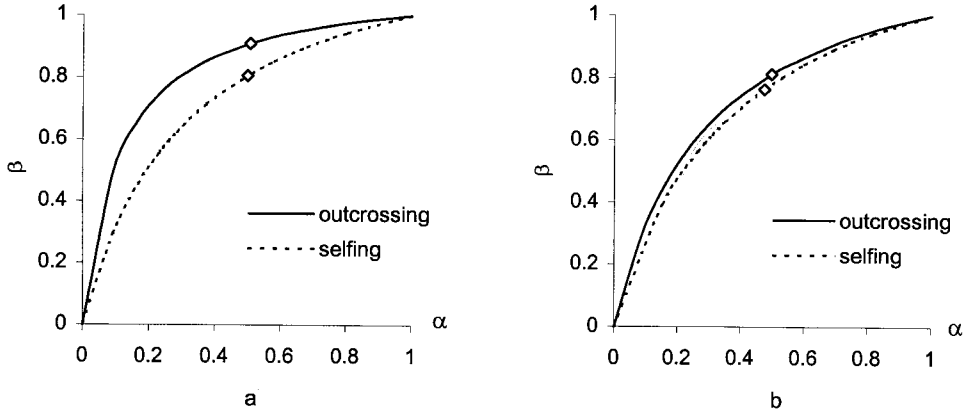

Figure 1.— Differential contribution of overdominant mutations to heterosis (as measured by α) and to the standing genetic variation (as measured by β) in outcrossing populations (solid line) and selfing populations (dotted lines). Plots a and b are for constant and variable mutation effects, respectively. On each plot, the curve to the left of the diamond point is obtained by fixing N (number of polymorphic overdominant loci in outcrossing populations) or $\bar{n}_o$ (mean number of overdominant loci in selfing populations) and letting U vary. In plot a, $N = 141$ for outcrossing populations and $\bar{n}_o = 4$ for selfing populations; in plot b, $N = 233$ for outcrossing populations and $\bar{n}_o = 5$ for selfing populations. The curve to the right of the diamond point is obtained by fixing U (=1) and letting N or $\bar{n}_o$ vary. All other parameters are the same: $\bar{h} = 0.36$, $\bar{s} = 0.03$; $h_o = -0.1$, $s_o = 0.03$.

mated when α and β are relatively large and is only overestimated when α and β are quite small. It should be noted that, as with the case for the outcrossing populations, when overdominant mutations are present but do not contribute substantially to heterosis and genetic variation, the bias of $\hat{U}$ is *smaller* than under dominant mutations. This is because the directions of estimation bias caused by overdominant mutations and variable effects of dominant mutations are opposite and they cancel each other, resulting in smaller (or no) bias. The

TABLE 4

Estimates of $\bar{h}$ with Deng's method and U with the inbreeding depression approach to characterize variable dominant mutations in the presence of overdominant mutations in outcrossing populations

h_o	s_o	N	α	β	$\hat{\bar{h}}$	$\hat{U}$
−0.1	0.03	2070	0.00	0.00	0.00 (0.09)	0.21 (1.23)
		920	0.20	0.52	0.10 (0.00) [0.20]	0.98 (1.08) [1.08]
		345	0.40	0.75	0.15 (0.00) [0.16]	0.74 (0.54) [0.60]
		153	0.60	0.87	0.17 (0.00) [0.13]	0.59 (0.50) [0.64]
		58	0.80	0.95	0.19 (0.00) [0.11]	0.50 (0.36) [0.62]
−0.2	0.03	1110	0.00	0.00	0.04 (0.12)	11.55 (47.59)
		493	0.20	0.37	0.11 (0.00) [0.20]	1.22 (1.92) [1.93]
		185	0.40	0.61	0.16 (0.00) [0.15]	1.04 (4.55) [4.55]
		82	0.60	0.78	0.18 (0.00) [0.13]	0.60 (0.38) [0.56]
		31	0.80	0.90	0.19 (0.00) [0.11]	0.49 (0.27) [0.58]
−0.2	0.01	3352	0.00	0.00	0.01 (0.12)	0.45 (1.68)
		1490	0.20	0.64	0.16 (0.00) [0.14]	1.77 (1.51) [1.69]
		559	0.40	0.83	0.19 (0.00) [0.12]	1.02 (0.63) [0.63]
		248	0.60	0.91	0.19 (0.00) [0.11]	0.67 (0.32) [0.46]
		93	0.80	0.97	0.20 (0.00) [0.10]	0.51 (0.23) [0.54]
0	0	0	1.00	1.00	0.25 (0.07) [0.06]	0.56 (0.07) [0.45]

See Table 2 legend.

TABLE 5

Characterizing constant deleterious genomic mutations in the presence of overdominant mutations with the fitness moments approach in selfing populations

h_o	s_o	N	α	β	$\hat{U}$	$\hat{h}$	$\hat{s}$
−0.1	0.03	36	0.00	0.00	3.24 (0.07)	0.10 (0.01)	0.02 (0.00)
		16	0.20	0.51	3.11 (0.27)	0.27 (0.02)	0.02 (0.00)
					[2.13]	[0.09]	[0.01]
		6	0.40	0.74	1.93 (0.26)	0.31 (0.02)	0.02 (0.00)
					[0.97]	[0.05]	[0.01]
		3	0.57	0.85	1.55 (0.27)	0.34 (0.02)	0.02 (0.01)
					[0.61]	[0.03]	[0.01]
		1	0.80	0.94	1.22 (0.23)	0.35 (0.02)	0.03 (0.01)
					[0.23]	[0.03]	[0.01]
−0.2	0.03	31	0.00	0.00	4.35 (0.21)	0.20 (0.01)	0.01 (0.00)
		14	0.20	0.54	3.64 (0.39)	0.30 (0.02)	0.02 (0.00)
					[2.67]	[0.07]	[0.01]
		5	0.41	0.77	2.10 (0.30)	0.33 (0.02)	0.02 (0.00)
					[1.14]	[0.04]	[0.01]
		2	0.63	0.89	1.51 (0.27)	0.35 (0.02)	0.02 (0.01)
					[0.58]	[0.03]	[0.01]
		1	0.77	0.94	1.28 (0.25)	0.35 (0.02)	0.03 (0.01)
					[0.38]	[0.03]	[0.01]
−0.2	0.01	93	0.00	0.00	4.37 (0.22)	0.20 (0.01)	0.00 (0.00)
		42	0.20	0.78	4.44 (0.71)	0.33 (0.02)	0.01 (0.00)
					[3.51]	[0.04]	[0.02]
		16	0.39	0.91	2.52 (0.52)	0.35 (0.03)	0.01 (0.00)
					[1.61]	[0.03]	[0.02]
		7	0.60	0.96	1.73 (0.89)	0.36 (0.02)	0.02 (0.00)
					[1.15]	[0.03]	[0.01]
		3	0.78	0.98	1.35 (0.32)	0.36 (0.03)	0.02 (0.01)
					[0.48]	[0.03]	[0.01]
0	0	0	1.00	1.00	1.04 (0.22)	0.36 (0.03)	0.03 (0.01)

See Table 1 legend.

extent of the bias depends on the parameters under estimation and α and β parameter values. The sampling variance of $\hat{U}$ is small. $\hat{\bar{h}}$ is always biased downward and the sampling variance is miniscule. With decreasing α and β, the degree of bias of $\hat{\bar{h}}$ increases. $\hat{\bar{h}}$ can serve well as a lower bound of the true $\bar{h}$.

DISCUSSION

Using extensive simulations, we investigated the effect of overdominant mutations on characterizing deleterious dominant mutations by the two existing estimation approaches (Morton *et al.* 1956; Charlesworth *et al.* 1990; Deng and Lynch 1996, 1997; Deng 1998b). We developed two important indices and associated analytical derivations to characterize the relative contributions of overdominant mutations to heterosis and genetic variation. The simulation algorithms and the analytical derivations developed are useful for investigating other issues in genetics concerning the mixture of dominant and overdominant mutations in the genome. Estimates for U are biased upward and those for $\bar{h}$ and $\bar{s}$ biased downward by overdominant mutations. However, the degree of bias is generally moderate and depends on the magnitude of the contribution of overdominant mutations to heterosis or genetic variation. This renders the estimates of U and $\bar{s}$ not invariably biased under variable mutation effects, which when working independently will almost always cause U and $\bar{s}$ to be underestimated. We also note that the contributions to heterosis and genetic variation from overdominant mutations are monotonic but not linearly proportional to each other. Our results may not only provide a basis for correct inferences about deleterious mutations from natural populations, but may also alleviate the biggest concern and obstacle in applying the inbreeding depression and fitness moments approaches, thus paving the way for efficiently characterizing deleterious genomic mutations from large natural populations.

Although it is intuitive that the two approaches will yield biased estimates (Drake *et al.* 1998), it is not clear what the magnitude and the direction of the bias will be for different estimates without the extensive simulations conducted here. Overdominant mutations, when acting together with variable mutation effects and depending on their contributions to heterosis and the standing

TABLE 6

Estimates of h with Mukai's method and U with the inbreeding depression approach to characterize constant dominant mutations in the presence of overdominant mutations in selfing populations

h_o	s_o	N	α	β	$\hat{h}$	$\hat{U}$
−0.1	0.03	36	0.00	0.00	−0.10 (0.00)	2.17 (0.02)
		16	0.20	0.51	0.13 (0.02) [0.23]	1.95 (0.09) [0.96]
		6	0.40	0.74	0.24 (0.01) [0.13]	1.35 (0.09) [0.36]
		3	0.57	0.85	0.29 (0.01) [0.08]	1.16 (0.08) [0.18]
		1	0.80	0.94	0.33 (0.01) [0.03]	1.04 (0.08) [0.08]
−0.2	−0.3	31	0.00	0.00	−0.20 (0.00)	1.87 (0.02)
		14	0.20	0.54	0.10 (0.02) [0.26]	1.84 (0.10) [0.84]
		5	0.41	0.77	0.23 (0.02) [0.13]	1.29 (0.09) [0.30]
		2	0.63	0.89	0.30 (0.01) [0.06]	1.11 (0.09) [0.14]
		1	0.77	0.94	0.33 (0.01) [0.04]	1.05 (0.08) [0.10]
−0.2	0.01	93	0.00	0.00	−0.20 (0.00)	1.86 (0.01)
		42	0.20	0.78	0.24 (0.01) [0.12]	2.78 (0.16) [1.79]
		16	0.39	0.91	0.30 (0.01) [0.06]	1.85 (0.13) [0.86]
		7	0.60	0.96	0.33 (0.01) [0.03]	1.41 (0.09) [0.42]
		3	0.78	0.98	0.34 (0.01) [0.02]	1.17 (0.07) [0.18]
0	0	0	1.00	1.00	0.36 (0.00)	0.98 (0.07)

See Table 1 legend.

genetic variation, may actually render estimates of U and $\bar{s}$ unbiased. It has been stipulated (Deng and Fu 1998; Drake *et al.* 1998) that the inbreeding depression and fitness moments approaches may be least affected by overdominant mutations in selfing populations, because overdominant mutations cannot be maintained by balancing selection there. However, as shown in Tables 1–8, with comparable contributions from overdominant mutations to heterosis and standing genetic variation, the estimation will be affected to a similar degree in outcrossing and selfing populations. We also note that the influence on the estimation from overdominant mutations will depend not only on their contributions to heterosis and the standing genetic variation, but also on the parameters of overdominant mutations such as h_o and s_o, although such dependence does not seem to be large.

Our simulation results not only reveal the robustness and statistical properties of the current approaches to characterize deleterious dominant mutations in natural populations, but also shed light on the relative efficiencies of the different approaches in different populations. Although the relative efficiencies of all the three available approaches (as outlined in the Introduction) were investigated earlier (Deng and Fu 1998), the investigations were not conducted under conditions of mixed dominant and overdominant mutations in the genome. In the present study, the sample sizes implemented in simulations for the two approaches investigated were deliberately set to be either comparable, or those for the inbreeding depression approach were actually larger. Recall that the number of genotypes employed for the fitness moments approach is 400 in outcrossing and selfing populations, while those for the inbreeding depression approach were 450 and 600, respectively, in outcrossing and selfing populations. However, it can be seen from Tables 1–8 that the estimation by the fitness moments approach is often better than the inbreeding depression approach. This is especially true for outcrossing populations and for the estimation of $\bar{h}$. The inbreeding depression approach is sometimes better for the estimation of U; however, the better estimation is achieved because of a greatly biased estimation of $\bar{h}$. Therefore, it is not the original inbreeding depression approach *per se* that achieves the better estimation for U. It is actually the greatly underestimated $\bar{h}$ by the

TABLE 7

Characterizing constant deleterious genomic mutations in the presence of overdominant mutations with the fitness moments approach in selfing populations

h_o	s_o	N	α	β	$\hat{U}$	$\hat{\bar{h}}$	$\hat{\bar{s}}$
−0.1	0.03	35	0.00	0.00	3.16 (0.07)	0.10 (0.01)	0.02 (0.00)
		15	0.20	0.52	2.41 (0.12)	0.22 (0.01)	0.02 (0.00)
					[1.42]	[0.14]	[0.01]
		6	0.39	0.73	1.42 (0.08)	0.25 (0.01)	0.03 (0.00)
					[0.43]	[0.11]	[0.00]
		3	0.56	0.84	1.06 (0.07)	0.26 (0.01)	0.03 (0.00)
					[0.09]	[0.10]	[0.01]
		1	0.79	0.94	0.80 (0.06)	0.28 (0.01)	0.04 (0.01)
					[0.21]	[0.08]	[0.01]
−0.2	0.03	30	0.00	0.00	4.20 (0.21)	0.20 (0.02)	0.01 (0.00)
		13	0.20	0.55	2.76 (0.20)	0.25 (0.02)	0.02 (0.00)
					[1.77]	[0.11]	[0.01]
		5	0.40	0.76	1.51 (0.11)	0.27 (0.02)	0.03 (0.00)
					[0.52]	[0.10]	[0.01]
		2	0.62	0.89	1.01 (0.07)	0.28 (0.02)	0.03 (0.00)
					[0.07]	[0.09]	[0.01]
		1	0.77	0.94	0.84 (0.06)	0.28 (0.01)	0.04 (0.01)
					[0.17]	[0.08]	[0.01]
−0.2	0.01	90	0.00	0.00	4.22 (0.21)	0.20 (0.02)	0.00 (0.00)
		40	0.20	0.79	3.04 (0.20)	0.27 (0.02)	0.01 (0.00)
					[2.05]	[0.10]	[0.02]
		15	0.40	0.91	1.57 (0.10)	0.27 (0.02)	0.02 (0.00)
					[0.59]	[0.09]	[0.01]
		7	0.59	0.95	1.10 (0.08)	0.28 (0.02)	0.03 (0.00)
					[0.13]	[0.08]	[0.00]
		3	0.77	0.98	0.85 (0.06)	0.28 (0.02)	0.04 (0.01)
					[0.17]	[0.08]	[0.01]
0	0	0	1.00	1.00	0.65 (0.05)	0.28 (0.01)	0.05 (0.01)
					[0.35]	[0.08]	[0.02]

See Table 1 legend.

estimation methods chosen that leads to the less biased U in the inbreeding depression approach. Therefore, the estimation of U by the inbreeding depression approach would greatly depend on the methods chosen for the estimation of $\bar{h}$. With less biased estimates or assumed values for $\bar{h}$, simulation results not shown here indicate that the U estimation by the inbreeding depression approach is much worse statistically than that of the fitness moments approach.

The issue of dominance and overdominance has been under debate for decades in genetics (Davenport 1908; East 1908; Shull 1908; Crow 1952; Sprague 1983; Wallace 1989; Houle 1989, 1994; Crow 1993; Deng *et al.* 1998a). The debate has far-reaching significance for agriculture, human health, evolution, and conservation biology, among other areas. While most of the data are consistent with the dominance hypothesis, overdominance cannot be ruled out in many situations (Simmons and Crow 1977; Charlesworth and Charlesworth 1987; Barrett and Charlesworth 1991; Stuber *et al.* 1992; Crow 1993; Mitton 1993). Given the current status of the debate, instead of favoring one hypothesis over the other, it may be more sensible to examine the issues concerned under mixed dominant and overdominant mutations in the genome, with mutations of each type having different contributions (*e.g.*, to heterosis and/or genetic variation, etc.). The theoretical machinery for measuring the relative importance of dominance and overdominance has not been available. The development of two important indices, α and β, provides a basis for investigating a number of other genetic issues related to the contribution of dominant and overdominant mutations to inbreeding and the standing genetic variation in natural populations.

It has long been recognized that, when dominant and overdominant mutations coexist, the heterosis and standing genetic variation will be affected by both. However, the disproportional contributions of overdominant mutations to heterosis and to standing genetic variation have not been documented before. This phenomenon may form a basis for discerning the relative importance of dominant and overdominant mutations in the genome. Studies have been initiated along this line of research. It is worthy of note that, for overdominant mutations to contribute relatively importantly to the standing genetic variation, a substantial proportion

TABLE 8

Estimates of $\bar{h}$ with Mukai's method and U with the inbreeding depression approach to characterize variable dominant mutations in the presence of overdominant mutations in selfing populations

h_o	s_o	N	α	β	$\hat{\bar{h}}$	$\hat{U}$
−0.1	0.03	35	0.00	0.00	−0.10 (0.00)	2.11 (0.02)
		15	0.20	0.52	0.09 (0.01) [0.27]	1.67 (0.05) [0.67]
		6	0.39	0.73	0.15 (0.01) [0.21]	1.02 (0.04) [0.04]
		3	0.56	0.84	0.19 (0.01) [0.17]	0.79 (0.04) [0.21]
		1	0.79	0.94	0.22 (0.01) [0.14]	0.65 (0.03) [0.36]
−0.2	0.03	30	0.00	0.00	−0.20 (0.00)	1.81 (0.02)
		13	0.20	0.55	0.10 (0.01) [0.26]	1.72 (0.06) [0.73]
		5	0.40	0.76	0.16 (0.01) [0.20]	1.03 (0.04) [0.05]
		2	0.62	0.89	0.20 (0.01) [0.16]	0.77 (0.04) [0.24]
		1	0.77	0.94	0.22 (0.01) [0.14]	0.67 (0.04) [0.33]
−0.2	0.01	90	0.00	0.00	−0.20 (0.00)	1.80 (0.01)
		40	0.20	0.79	0.17 (0.01) [0.19]	2.11 (0.08) [0.12]
		15	0.40	0.91	0.21 (0.01) [0.15]	1.22 (0.05) [0.23]
		7	0.59	0.95	0.23 (0.01) [0.13]	0.88 (0.04) [0.12]
		3	0.77	0.98	0.24 (0.01) [0.12]	0.71 (0.04) [0.29]
0	0	0	1.00	1.00	0.25 (0.01) [0.11]	0.57 (0.03) [0.43]

Note that we estimate the constant h in Table 6 and thus $\hat{h}$ is given there. In this table we estimate the average $\bar{h}$ of variable h_i across loci; thus $\hat{\bar{h}}$ is given here.

See legend to Table 1.

of heterosis must be caused by overdominant mutations. This is especially true when overdominant mutations contribute to less than half of the heterosis ($\alpha > 0.5$; Figure 1).

For any theory to be of great significance, its underlying assumptions must be examined closely and the important parameters must be estimated. There is no doubt that any genome is subject to continuous bombardment of deleterious genomic mutations. However, no amount of theoretical argument can resolve the issues concerning the importance of deleterious genomic mutations without the important parameters being estimated. Indisputably, characterizing deleterious genomic mutations is extremely important. However, even if the importance is realized by more and more scientists and revealed in more and more biological aspects, the estimates are astonishingly few and thus are imperatively needed (Peck and Eyre-Walker 1997). Among the three approaches currently available, the statistical properties and the robustness of the fitness moments approach are investigated most thoroughly and best known. Investigation of the other two available approaches (the M-A approach and the inbreeding-depression approach) is also extremely important and is beginning to appear in studies (Deng and Fu 1998; Deng *et al.* 1998b). Different approaches have different peculiar assumptions whose validity may be difficult to consolidate in a specific experimental setting (Keightley 1994; Peck and Eyre-Walker 1997; Deng and Fu 1998; Lynch *et al.* 1998). Examples of these assumptions are M-S balance in the fitness-moments approach and in the inbreeding-depression approach, no line losses because of selection during M-A, no gene conversion for the M-A chromosome in Drosophila, etc. Applying multiple approaches to the same organism and/or characterizing deleterious mutations in diverse organisms may provide a cross-check of the results (and of the underlying assumptions to derive these results) and eventually may crystallize the deleterious mutation parameters.

H.-W. Deng thanks Professor M. Lynch for years of advice, continuous encouragement, and support. We are very grateful to Professor Marjorie A. Asmussen and three anonymous reviewers for their ex-

tremely careful comments that helped to improve the article. We thank Drs. Robert R. Recker and Mark Johnson and Ms. Carolyn Meeks for careful editing of the manuscript. The work was partially supported by a grant from National Institutes of Health (R01 AR45349) and a Health Future Foundation grant from Creighton University, Nebraska, and by graduate student tuition waiver to J.-L.L. and J.L. from the Department of Biomedical Sciences of Creighton University.

LITERATURE CITED

Barrett, S. C. H., and D. Charlesworth, 1991 Effects of a change in the level of inbreeding on the genetic load. Nature **352:** 522–524.

Bateman, A. J., 1959 The viability of near-normal irradiated chromosomes. Int. J. Radiat. Biol. **1:** 170–180.

Büger, R., and J. Hofbauer, 1994 Mutation load and quantitative genetic traits. J. Math. Biol. **32:** 193–218.

Caballero, A., and P. D. Keightley, 1994 A pleiotropic nonadditive 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.

Charlesworth, B., 1990 Mutation-selection balance and the evolutionary advantage of sex and recombination. Genet. Res. **55:** 199–221.

Charlesworth, D., and B. Charlesworth, 1987 Inbreeding depression and its evolutionary consequences. Annu. Rev. Ecol. Syst. **18** (Suppl.): 237–268.

Charlesworth, B., and K. A. Hughes, 1998 The maintenance of genetic variation in life-history traits, in *Evolutionary Genetics from Molecules to Morphology*, edited by R. S. Singh and C. B. Krimbas. Cambridge University Press, Cambridge, UK (in press).

Charlesworth, B., D. Charlesworth and M. T. Morgan, 1990 Genetic loads and estimates of mutation rates in highly inbred plant populations. Nature **347:** 380–382.

Charlesworth, B., M. T. Morgan and D. Charlesworth, 1993 The effects of deleterious mutations on neutral molecular variation. Genetics **134:** 1289–1303.

Charlesworth, D., E. E. Lyons and L. B. Litchfield, 1994 Inbreeding depression in two highly inbreeding populations of Leavenworthia. Proc. R. Soc. Lond. Ser. B Biol. Sci. **258:** 209–214.

Charlesworth, D., B. Charlesworth and M. T. Morgan, 1995 The pattern of neutral molecular variation under the background selection model. Genetics **141:** 1619–1632.

Craddock, N., V. Khodel, P. V. Eerdewegh and T. Reich, 1995 Mathematical limits of multilocus models: the genetic transmission of bipolar disorder. Am. J. Hum. Genet. 57: 690–702.

Crow, J. F., 1952 Dominance and overdominance, pp. 282–297 in *Heterosis*, edited by J. W. Gowen. Iowa State College Press, Ames, IA.

Crow, J. F., 1986 *Basic Concepts in Population, Quantitative and Evolutionary Genetics.* W. H. Freeman and Company, New York.

Crow, J. F., 1993 How much do we know about spontaneous human mutation rates? Environ. Mol. Mutagen. **21:** 122–129.

Crow, J. F., and M. J. Simmons, 1983 The mutation load in Drosophila, pp. 1–35 in *The Genetics and Biology of Drosophila*, Vol. 3c, edited by M. Ashburner, H. L. Carson and J. N. Thompson, Academic Press, London/NY.

Davenport, C. B., 1908 Degeneration, albinism and inbreeding. Science **28:** 454–455.

Deng, H.-W., 1998a Estimating (over)dominance coefficient and discriminating dominance *vs.* overdominance as the genetic cause of heterosis. Genetics **148:** 2003–2014.

Deng, H.-W., 1998b Characterizing deleterious genomic mutations in outcrossing populations. Genetics **151:** 945–956.

Deng, H.-W., and Y.-X. Fu, 1998 On the three methods for estimating deleterious genomic mutation parameters. Genet. Res. **71:** 223–236.

Deng, H.-W., and M. Lynch, 1996 Estimation of the 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., Y.-X. Fu and M. Lynch, 1998a Inferring the major genomic mode of dominance and overdominance. Genetica **102/103:** 559–567.

Deng, H.-W., J. Li and J. L. Li, 1998b On the experimental designs and data analyses of mutation-accumulation experiments. Genet. Res. (in press).

Drake, J. W., B. Charlesworth, D. Charlesworth and J. F. Crow, 1998 Rates of spontaneous mutations. Genetics **148:** 1667–1686.

East, E. M., 1908 Inbreeding in corn. Rep. Conn. Agric. Exp. Stn. (1907) 419–428.

Fu, Y.-B., and K. Ritland, 1996 Marker-based inference about epistasis for gene influencing inbreeding depression. Genetics **144:** 339–348.

Garcia-Dorado, A., 1997 The rate and effects distribution of viability mutation in Drosophila: minimum distance estimation. Evolution **51:** 1130–1139.

Gregory, W. C., 1965 Mutation frequency, magnitude of change and the probability of improvement in adaptation. Radiat. Bot. **5** (Suppl.): 429–441.

Haldane, J. B. S., 1937 The effects of variation on fitness. Am. Nat. **71:** 337–349.

Houle, D., 1989 Allozyme-associated heterosis in *Drosophila melanogaster*. Genetics **123:** 789–801.

Houle, D., 1994 Adaptive distance and the genetic basis of heterosis. Evolution **48:** 1410–1417.

Houle, D., B. Morikawa and M. Lynch, 1996 Comparing mutational variabilities. Genetics **143:** 1467–1483.

Hudson, R., and N. L. Kaplan, 1995 Deleterious background selection with recombination. Genetics **141:** 1605–1617.

Johnston, M., and D. J. Schoen, 1995 Mutation rates and dominance levels of genes affecting total fitness in two angiosperm species. Science **267:** 226–229.

Kacser, H., and J. A. Burns, 1981 The molecular basis of dominance. Genetics **97:** 639–666.

Keightley, P. D., 1994 The distribution of mutation effects on viability in *Drosophila melanogaster*. Genetics **138:** 1315–1322.

Keightley, P. D., 1996 Nature of deleterious mutation load in Drosophila. Genetics **144:** 1993–1999.

Kimura, M., T. Maruyama and J. F. Crow, 1963 The mutation load in small populations. Genetics **48:** 1303–1312.

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

Kondrashov, A. S., 1985 Deleterious mutations as an evolutionary factor. II. Facultative apomixis and selfing. Genetics **111:** 635–653.

Kondrashov, A. S., 1988 Deleterious mutations and the evolution of sexual reproduction. Nature **336:** 435–440.

Kondrashov, A. S., and J. F. Crow, 1991 Haploidy or diploidy: which is better? Nature **351:** 314–315.

Lande, R., 1994 Risk of population extinction from new deleterious mutations. Evolution **48:** 1460–1469.

Lynch, M., 1985 Spontaneous mutations for life-history characters in an obligate parthenogen. Evolution **39:** 804–818.

Lynch, M., 1988 The rate of polygenic mutation. Genet. Res. **51:** 173–148.

Lynch, M., J. Conery and R. Burger, 1995 Mutation accumulation and the extinction of small populations. Am. Nat. **146:** 489–518.

Lynch, M., J. Conery and R. Burger, 1996 Mutation meltdowns in sexual populations. Evolution **49:** 1067–1080.

Lynch, M., L. Latta, J. Hicks and M. Giorgianni, 1998 Mutation, selection, and maintenance of life-history variation in a natural population. Evolution **52:** 727–733.

Mackay, T. F. C., R. F. Lyman and M. S. Jackson, 1992 Effects of *P* element insertions on quantitative traits in *Drosophila melanogaster*. Genetics **130:** 315–322.

Mitton, J. B., 1993 Theory and data pertinent to the relationship between heterozygosity and fitness, pp. 17–41 in *The History of Inbreeding and Outbreeding*, edited by N. D. Thornhill. The University of Chicago Press, Chicago.

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., 1964 Spontaneous mutation rate of polygenes controlling viability. Genetics **50:** 1–19.

Mukai, T., 1979 Polygenic mutations, pp. 177–196 in *Quantitative*

Genetic Variation, edited by J. N. Thompson, Jr., and J. M. Thoday. Academic Press, New York.
Mukai, T., S. I. 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., 1964 The relation of recombination to mutational advance. Mutat. Res. **1:** 1–9.
Peck J. R., and A. Eyre-Walker, 1997 The muddle about mutations. Nature **387:** 135–136.
Shull, G. H., 1908 The composition of a field of maize. Rep. Am. Breed. Assoc. **4:** 296–301.
Simmons, M. J., and J. F. Crow, 1977 Mutations affecting fitness in *Drosophila* populations. Annu. Rev. Genet. **11:** 49–78.
Sprague, G. F., 1983 Heterosis in maize: theory and practice, pp. 47–70 in *Heterosis: Reappraisal of Theory and Practice*, edited by R. Frankel. Springer-Verlag, Berlin.
Stuber, C. W., S. E. Lincoln, D. W. Wolff, T. Helentijaris and E. S. Lander, 1992 Identification of genetic factors in a hybrid from two elite maize inbred lines using molecular markers. Genetics **132:** 823–839.
Wallace, B., 1989 One selectionist's perspective. Q. Rev. Biol. **64:** 127–145.

Communicating editor: M. A. Asmussen

APPENDIX A: DETERMINATION OF α AND β IN OUTCROSSING POPULATIONS

Constant dominant mutation effects: Unless otherwise defined in the appendices, all the notations used are the same as in the text. Multiplicative fitness function and Hardy-Weinberg and linkage equilibrium are assumed. Let $p(n)$ denote a Poisson distribution with mean $\bar{n}$ and density function $p(n) = \bar{n}^n e^{-\bar{n}}/n!$. Let the subscripts d and o for $\bar{n}$, h, and s denote the parameters for dominant and overdominant mutations, respectively. In relation to the parameters for dominant and overdominant mutations, the expectation terms in Equation 4 in the text can be obtained. The expectations of ln(fitness) due to dominant mutations alone in the outcrossed parental and selfed progeny generations are, respectively,

$$E(\ln w_{dp}) = \sum_{n=0}^{\infty} p(n)\ln(1 - h_d s_d)^n = \bar{n}_d \ln(1 - h_d s_d), \quad \text{(A1)}$$

and

$$E(\ln w_{do}) = \sum_{n=0}^{\infty}\left\{p(n)\sum_{i=0}^{n}\sum_{j=0}^{n-i}\left\{\binom{n}{i}\binom{n-i}{j}\left(\frac{1}{4}\right)^{n-i-j}\left(\frac{1}{2}\right)^{i}\left(\frac{1}{4}\right)^{j} \times [i\ln(1 - h_d s_d) + j\ln(1 - s_d)]\right\}\right\},$$

$$= \frac{\bar{n}_d}{2}\ln(1 - h_d s_d) + \frac{\bar{n}_d}{4}\ln(1 - s_d), \quad \text{(A2)}$$

where

$$\binom{n}{i}\binom{n-i}{j}\left(\frac{1}{4}\right)^{n-i-j}\left(\frac{1}{2}\right)^{i}\left(\frac{1}{4}\right)^{j}$$

is the probability of obtaining i heterozygous loci for *Aa* and j homozygous loci for *aa* in a selfed offspring during random segregation upon selfing for the n heterozygous loci in the outcrossed parent. The double summation is over all possible combinations of the number of heterozygous (*Aa*) and homozygous (*aa*) loci (resulting from random segregation for dominant mutations) in a selfed offspring. The expectation of total ln(fitness) due to both dominant and overdominant mutations in the outcrossed parental generation is

$$E(\ln w_{tp}) = \sum_{n=0}^{\infty}\left\{p(n)\sum_{m=0}^{N}\sum_{l=0}^{N-m}\left\{\binom{N}{m}\binom{N-m}{l}(p^2)^{N-m-l}(2pq)^m(q^2)^l \times [n\ln(1 - h_d s_d) + m\ln(1 - h_o s_o) + l\ln(1 - s_o)]\right\}\right\}$$

$$= \bar{n}_d \ln(1 - h_d s_d) + 2pqN\ln(1 - h_o s_o) + q^2 N\ln(1 - s_o), \quad \text{(A3)}$$

where

$$\binom{N}{m}\binom{N-m}{l}(p^2)^{N-m-l}(2pq)^m(q^2)^l$$

is the probability of having m heterozygous loci for *Bb* and l homozygous loci for *bb* in an outcrossed parent given the equilibrium frequencies p for the *B* allele and q for the *b* allele. The double summation is over all possible combinations of the number of heterozygous *Bb* and homozygous *bb* loci (resulting from random combination for overdominant mutations) in an outcrossed parent. The expectation of total ln(fitness) due to both dominant and overdominant mutations in the selfed progeny generation is

$$E(\ln w_{to}) = \sum_{n=0}^{\infty}\left\{p(n)\sum_{m=0}^{N}\sum_{l=0}^{N-m}\left\{\binom{N}{m}\binom{N-m}{l}\left(p^2 + \frac{pq}{2}\right)^{N-m-l} \times (pq)^m\left(q^2 + \frac{pq}{2}\right)^{l}\left(\frac{n}{2}\ln(1 - h_d s_d) + \frac{n}{4}\ln(1 - s_d) + m\ln(1 - h_o s_o) + l\ln(1 - s_o)\right)\right\}\right\}$$

$$= \frac{\bar{n}_d}{2}\ln(1 - h_d s_d) + \frac{\bar{n}_d}{4}\ln(1 - s_d) + pqN\ln(1 - h_o s_o) + \left(q^2 + \frac{pq}{2}\right)N\ln(1 - s_o). \quad \text{(A4)}$$

With Equations A1–A4, the index α (defined in Equation 4 in the text) can be expressed in terms of the mutation parameters for outcrossing populations with constant dominant mutation effects.

The two terms in Equation 5 in the text can be expressed in terms of the parameters for dominant and overdominant mutations. The derivation is as follows. The expectation of the second moment of the ln(fit-

ness) in the parental generation that is due to dominant mutations only is

$$E[(\ln w_{dp})^2] = \sum_{n=0}^{\infty} p(n)[n \ln(1 - h_d s_d)]^2 = (\bar{n}_d^2 + \bar{n}_d)[\ln(1 - h_d s_d)]^2. \quad \text{(A5)}$$

By Equations A1 and A5, we have

$$\mathrm{Var}(\ln w_{dp}) = E[(\ln w_{dp})^2] - [E(\ln w_{dp})]^2 = \bar{n}_d[\ln(1 - h_d s_d)]^2. \quad \text{(A6)}$$

The expectation of the second moment of the ln(fitness) in the parental generation that is due to both dominant and overdominant mutations is

$$\begin{aligned} E[(\ln w_{tp})^2] = \sum_{n=0}^{\infty}\Bigg\{p(n)\sum_{m=0}^{N}\sum_{l=0}^{N-m}\Bigg\{\binom{N}{m}\binom{N-m}{l}(p^2)^{N-m-l}(2pq)^m(q^2)^l \\ \times [n \ln(1 - h_d s_d) + m \ln(1 - h_o s_o) \\ + l \ln(1 - s_o)]^2\Bigg\}\Bigg\} \\ = (\bar{n}_d^2 + \bar{n}_d)[\ln(1 - h_d s_d)]^2 \\ + 2pqN\,(1 + 2pq(N - 1))[\ln(1 - h_o s_o)]^2 \\ + q^2N(1 + q^2(N - 1))[\ln(1 - s_o)]^2 \\ + 2\bar{n}_d qN \ln(1 - h_d s_d) \\ \times (2p \ln(1 - h_o s_o) + q \ln(1 - s_o)) \\ + 4pq^3N(N-1)\ln(1 - h_o s_o)\ln(1 - s_o). \end{aligned} \quad \text{(A7)}$$

By Equations A3 and A7, we have

$$\begin{aligned} \mathrm{Var}(\ln w_{tp}) &= E[(\ln w_{tp})^2] - [E(\ln w_{tp})]^2 \\ &= \bar{n}_d[\ln(1 - h_d s_d)]^2 \\ &\quad + 2pq(1 - 2pq)N[\ln(1 - h_o s_o)]^2 \\ &\quad + q^2(1 - q^2)N[\ln(1 - s_o)]^2 \\ &\quad - 4pq^3N \ln(1 - h_o s_o)\ln(1 - s_o). \end{aligned} \quad \text{(A8)}$$

With Equations A6 and A8, the index β (defined in Equation 5 in the text) can be expressed in terms of the mutation parameters for outcrossing populations with constant dominant mutation effects.

Variable dominant mutation effects: Although a little more complex, the derivation is a natural extension of the cases of the constant dominant mutation effects. Let IN be number of classes of dominant mutation effects. Dominant mutation effects are constant within each such class and differ among these classes. Although dominant mutation effects are most likely variable, there is no solid knowledge on their distribution (continuous or discrete and the exact form of distribution, etc.). In addition, a discrete distribution can approximate any continuous distribution to any desired degree of accuracy. Therefore, modeling variable dominant mutation effects by a discrete distribution is appropriate,

$$\begin{aligned} E(\ln w_{dp}) &= \sum_{m_1=0}^{\infty}\cdots\sum_{m_{IN}=0}^{\infty}\left[\prod_{i=1}^{IN}p_j(m_i)\sum_{i=1}^{IN}m_i \ln(1 - h_i s_i)\right] \\ &\approx -U_d, \end{aligned} \quad \text{(A9)}$$

where m_i is the number of dominant mutations of the ith class in the genome, and $p_j(m_i)$ is the probability of having m_i mutations of the ith class in the genome. The multiple IN summations are over the IN discrete classes of mutations, including all the different combinations of the different numbers of mutations of the IN classes in the genome

$$\begin{aligned} E(\ln w_{do}) &= \sum_{m_1=0}^{\infty}\cdots\sum_{m_{IN}=0}^{\infty}\left\{\prod_{i=1}^{IN}p_i(m_i)\sum_{i=1}^{IN}m_i\left[\frac{1}{2}\ln(1 - h_i s_i) + \frac{1}{4}\ln(1 - s_i)\right]\right\} \\ &\approx -\frac{U_d}{2} - \frac{U_d}{4\tilde{h}}, \end{aligned} \quad \text{(A10)}$$

where $\tilde{h}$ is the harmonic mean of the dominant mutation effects across loci

$$\begin{aligned} E(\ln w_{tp}) = \sum_{m_1=0}^{\infty}\cdots\sum_{m_{IN}=0}^{\infty}\Bigg\{\Bigg[\prod_{i=1}^{IN}p_i(m_i)\Bigg]\sum_{r=0}^{N}\sum_{s=0}^{N-r} \\ \times\left\{\binom{N}{r}\binom{N-r}{s}(p^2)^{N-r-s}(2pq)^r(q^2)^s \right. \\ \times\Bigg[\sum_{i=1}^{IN}m_i\ln(1 - h_i s_i) \\ \left. + r \ln(1 - h_o s_o) + s \ln(1 - s_o)\Bigg]\right\}\Bigg\}, \\ \approx -U_d + 2pqN \ln(1 - h_o s_o) + q^2N \ln(1 - s_o) \end{aligned} \quad \text{(A11)}$$

$$\begin{aligned} E(\ln w_{to}) = \sum_{m_1=0}^{\infty}\cdots\sum_{m_{IN}=0}^{\infty}\Bigg\{\Bigg[\prod_{j=1}^{IN}p_j(m_i)\sum_{r=0}^{N}\sum_{s=0}^{N-r} \\ \times\left\{\binom{N}{r}\binom{N-r}{s}\left(p^2 + \frac{pq}{2}\right)^{N-r-s} \right. \\ \times (pq)^r\left(q^2 + \frac{pq}{2}\right)^s \\ \times\Bigg\{\sum_{i=1}^{IN}m_i\Bigg[\frac{1}{2}\ln(1 - h_i s_i) \\ \left. + \frac{1}{4}\ln(1 - s_i)\Bigg] + r \ln(1 - h_o s_o) \right. \\ \left. + s \ln(1 - s_o)\Bigg\}\right\}\Bigg\} \\ \approx -\frac{U_d}{2} - \frac{U_d}{4\tilde{h}} + pqN \ln(1 - h_o s_o) \\ + \left(p^2 + \frac{pq}{2}\right)N \ln(1 - s_o). \end{aligned} \quad \text{(A12)}$$

With Equations A9–A12, the index α can be expressed in terms of the mutation parameters for outcrossing populations with variable dominant mutation effects

$$E[(\ln w_{dp})^2] = \sum_{m_1=0}^{\infty}\cdots\sum_{m_{IN}=0}^{\infty}\left\{\prod_{i=1}^{IN}p_j(m_i)\left[\sum_{i=1}^{IN}m_i\ln(1-h_is_i)\right]^2\right\}$$
$$= \sum_{i=1}^{IN}(\overline{n}_i^2+\overline{n}_i)[\ln(1-h_is_i)]^2 + \sum\sum_{1\le i<j\le IN}2\overline{n}_i\overline{n}_j\ln(1-h_is_i)\ln(1-h_js_j), \quad (A13)$$

$$E[(\ln w_{tp})^2] = \sum_{m_1=0}^{\infty}\cdots\sum_{m_{IN}=0}^{\infty}\left\{\left[\prod_{i=1}^{IN}p_j(m_i)\right]\sum_{r=0}^{N}\sum_{s=0}^{N-r} \times\left\{\binom{N}{r}\binom{N-r}{s}(p^2)^{N-r-s}(2pq)^r(q^2)^s \times\left[\sum_{i=1}^{IN}m_i\ln(1-h_is_i) + r\ln(1-h_os_o) + s\ln(1-s_o)\right]^2\right\}\right\}. \quad (A14)$$

By Equations A9 and A13, we have

$$\mathrm{Var}(\ln w_{dp}) = E[(\ln w_{dp})^2] - [E(\ln w_{dp})]^2 \approx U_d\overline{hs}. \quad (A15)$$

By Equations A11 and A14, we have

$$\mathrm{Var}(\ln w_{tp}) = E((\ln w_{tp})^2) - [E(\ln w_{tp})]^2 \approx U_d\overline{hs} + 2pq(1-2pq)N[\ln(1-h_os_o)]^2 + q^2(1-q^2)N[\ln(1-s_o)]^2 - 4pq^3N\ln(1-h_os_o)\ln(1-s_o). \quad (A16)$$

With Equations A15–A16, the index β can be expressed in terms of the mutation parameters for outcrossing populations with variable dominant mutation effects.

APPENDIX B: DETERMINATION OF α AND β IN SELFING POPULATIONS

Constant dominant mutation effects: The derivation for selfing populations is relatively straightforward. Noting that the number of dominant loci in a selfed progeny is the sum of those in its selfed parents, we have the four expectation terms for the ln(fitness) in the selfed parental and outcrossed offspring generations that are due to pure dominance (d) or both dominance and overdominance (t):

$$E(\ln w_{dp}) = \sum_{n=0}^{\infty}p(n)n\ln(1-s_d) = \overline{n}_d\ln(1-s_d), \quad (B1)$$

$$E(\ln w_{do}) = \sum_{n=0}^{\infty}p(n)n\ln(1-h_ds_d) = 2\overline{n}_d\ln(1-h_ds_d), \quad (B2)$$

$$E(\ln w_{tp}) = \sum_{n=0}^{\infty}\left[p(n)\sum_{m=0}^{\infty}p(m)[n\ln(1-s_d) + m\ln(1-s_o)]\right] = \overline{n}_d\ln(1-s_d) + \overline{n}_o\ln(1-s_o), \quad (B3)$$

$$E(\ln w_{to}) = \sum_{n=0}^{\infty}\left[p(n)\sum_{m=0}^{\infty}p(m)[n\ln(1-h_ds_d) + m\ln(1-h_os_o)]\right] = 2\overline{n}_d\ln(1-h_ds_d) + 2\overline{n}_o\ln(1-h_os_o). \quad (B4)$$

With Equations B1–B4, the index α can be expressed in terms of the mutation parameters for selfing populations with constant dominant mutation effects. The expectations of the second moment of the ln(fitness) in the selfed parental and outcrossed offspring generations that are due to pure dominance (d) or both dominance and overdominance (t) can also be relatively easily derived:

$$E((\ln w_{dp})^2) = \sum_{n=0}^{\infty}p(n)[n\ln(1-s_d)]^2 = (\overline{n}_d^2+\overline{n}_d)[\ln(1-s_d)]^2, \quad (B5)$$

$$E((\ln w_{tp})^2) = \sum_{n=0}^{\infty}\left[p(n)\sum_{m=0}^{\infty}p(m)(n\ln(1-s_d) + m\ln(1-s_o))^2\right] = (\overline{n}_d^2+\overline{n}_d)[\ln(1-s_d)]^2 + (\overline{n}_o^2+\overline{n}_o)[\ln(1-s_o)]^2 + 2\overline{n}_d\overline{n}_o\ln(1-s_d)\ln(1-s_o), \quad (B6)$$

$$\mathrm{Var}(\ln w_{dp}) = E((\ln w_{dp})^2) - [E(\ln w_{dp})]^2 = \overline{n}_d[\ln(1-s_d)]^2, \quad (B7)$$

$$\mathrm{Var}(\ln w_{tp}) = E((\ln w_{tp})^2) - [E(\ln w_{tp})]^2 = \overline{n}_d[\ln(1-s_d)]^2 + \overline{n}_o[\ln(1-s_o)]^2. \quad (B8)$$

With Equations B7–B8, the index β can be expressed in terms of the mutation parameters for selfing populations with constant dominant mutation effects.

Variable dominant mutation effects: The derivations and the notations are similar to and simpler than those in outcrossing populations, thus will not be elaborated for selfing populations. The expectations of the first moment of ln(fitness) in the selfed parental and outcrossed offspring generations that are due to dominance alone (d) or both dominance and overdominance (t) are, respectively,

$$E(\ln w_{dp}) = \sum_{m_1=0}^{\infty}\cdots\sum_{m_{IN}=0}^{\infty}\left\{\left[\prod_{j=1}^{IN}p_j(m)\right]\prod_{i=1}^{IN}m_i\ln(1-s_i)\right\} \approx -\frac{U_d}{2}, \quad (B9)$$

$$E(\ln w_{do}) = \sum_{m_1=0}^{\infty}\cdots\sum_{m_{IN}=0}^{\infty}\left\{\left[\prod_{j=1}^{IN}p_j(m)\right]\sum_{i=1}^{IN}m_i\ln(1-h_is_i)\right\} \approx -U_d\overline{h}, \quad (B10)$$

$$E(\ln w_{tp}) = \sum_{m_1=0}^{\infty} \cdots \sum_{m_{IN}=0}^{\infty} \left\{ \left[\prod_{j=1}^{IN} p_j(m) \right] \sum_{k=0}^{\infty} p(k) \times \left[\sum_{i=1}^{IN} m_i \ln(1 - s_i) + k \ln(1 - s_o) \right] \right\}$$

$$\approx -\frac{U_d}{2} + \bar{n}_o \ln(1 - s_o), \quad \text{(B11)}$$

$$E(\ln w_{to}) = \sum_{m_1=0}^{\infty} \cdots \sum_{m_{IN}=0}^{\infty} \left\{ \left[\prod_{j=1}^{IN} p_j(m) \right] \sum_{k=0}^{\infty} p(k) \times \left[\sum_{i=1}^{IN} m_i \ln(1 - h_d s_i) + k \ln(1 - h_o s_o) \right] \right\}$$

$$\approx -U_d \bar{h} + 2\bar{n}_o \ln(1 - h_o s_o). \quad \text{(B12)}$$

With Equations B9–B12, the index a can be expressed in terms of the mutation parameters for selfing populations with variable dominant mutation effects. The expectations of the second moment of ln(fitness) in the selfed parental and outcrossed offspring generations that are due to dominance alone (d) or both dominance and overdominance (t) are, respectively,

$$E((\ln w_{dp})^2) = \sum_{m_1=0}^{\infty} \cdots \sum_{m_{IN}=0}^{\infty} \left\{ \left[\prod_{j=1}^{IN} p_j(m) \right] \left[\sum_{i=1}^{IN} m_i \ln(1 - s_i) \right]^2 \right\}$$
$$= \sum_{i=1}^{IN} (\bar{n}_i^2 + \bar{n}_i)[\ln(1 - s_i)]^2 + \sum\sum_{1 \le j < j \le IN} 2\bar{n}_i \bar{n}_j \ln(1 - s_i) \ln(1 - s_j), \quad \text{(B13)}$$

$$E((\ln w_{tp}))^2 = \sum_{m_1=0}^{\infty} \cdots \sum_{m_{IN}=0}^{\infty} \left\{ \left[\prod_{j=1}^{IN} p_j(m) \right] \sum_{k=0}^{\infty} p(k) \left[\sum_{i=1}^{IN} m_i \ln(1 - s_i) + k \ln(1 - s_o) \right]^2 \right\}, \quad \text{(B14)}$$

$$\mathrm{Var}(\ln w_{dp}) = E((\ln w_{dp})^2) - [E(\ln w_{dp})]^2 \approx \frac{U_d \bar{s}}{2}. \quad \text{(B15)}$$

$$\mathrm{Var}(\ln w_{tp}) = E((\ln w_{tp})^2) - [E(\ln w_{tp})]^2 \approx \frac{U_d \bar{s}}{2} + \bar{n}_o [\ln(1 - s_o)]^2. \quad \text{(B16)}$$

With Equations B15–B16, the index β can be expressed in terms of the mutation parameters for selfing populations with variable dominant mutation effects.