Estimating Within-Locus Nonadditive Coefficient and Discriminating Dominance Versus Overdominance as the Genetic Cause of Heterosis

Hong-Wen Deng

Osteoporosis Research Center, Department of Medicine, Creighton University, Omaha, Nebraska 68131

Manuscript received June 26, 1997
Accepted for publication December 18, 1997

ABSTRACT

Testing (over)dominance as the genetic cause of heterosis and estimating the (over)dominance coefficient (h) are related. Using simulations, we investigate the statistical properties of Mukai's approach, which is intended to estimate the average ($\bar{h}$) of h_i across loci by regression of outcrossed progeny on the sum of the two corresponding homozygous parents. A new approach for estimating $\bar{h}$ is also developed, utilizing data on families formed by multiple selfed genotypes from each outcrossed parent, thus not requiring constructing homozygotes. Assuming constant mutation effects, h can be estimated accurately by both approaches under dominance. When rare alleles have low frequencies at any polymorphic locus, Mukai's approach can estimate h accurately under over(under)dominance. Therefore, the (over)dominance hypothesis for heterosis can be tested by estimating h, under either dominance or overdominance at all genomic loci. However, this is invalid with more plausible mixed dominance and overdominance at different loci. Estimating the variance of h_i across loci is also investigated. In self-compatible outcrossing populations with mutations of variable effects and lethals, our new approach is better than Mukai's, not only because of not requiring homozygotes but also because of the better statistical performance reflected by the smaller mean square errors of the estimates.

INBREEDING depression results from mating among relatives, and outbreeding enhancement results from mating among usually inbreeding lines or isolated populations. Both phenomena are widely observed (*e.g.*, Wright 1977; Charlesworth and Charlesworth 1987; Falconer 1989; Crow 1993; Lynch and Walsh 1997). For simplicity, hereafter we will refer to both phenomena collectively as heterosis. The magnitude of heterosis has implications in many areas, such as the evolution of self-incompatibility systems in plants (Lande and Schemske 1985; Schemske and Lande 1985; Charlesworth and Charlesworth 1987), the evolution of dispersal mechanisms for inbreeding avoidance in animals (Shields 1982), the biological conservation of rare and endangered species (Soule 1986), the improvement of agricultural production (Falconer 1989), and the protection of human welfare (Cavalli-Sforza and Bodmer 1971).

There are two main rival genetic hypotheses concerning individual loci to explain heterosis. One is the dominance hypothesis (Davenport 1908; Crow 1952), which argues that heterosis is caused by an enhanced expression of deleterious genes when homozygosity is increased and the heterozygote performance is somewhere (but not exactly) in between the two corresponding homozygotes. The other is the overdominance hypothesis (East 1908; Shull 1908; Crow 1952), which argues for heterozygote superiority relative to both homozygotes. Although neither dominance nor overdominance is necessary for heterosis (Richey 1942; Minvielle 1987; Schnell and Cockerham 1992), here we concentrate on studying just these two mechanistic causes (see discussion).

Although most experimental 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). At present, the evidence for functional overdominance does not seem to be very convincing, and most cited examples are compatible with associated overdominance [an artifact of linked deleterious recessive genes (*cf.* Houle 1989, 1994; Crow 1993)]. The debate over the relative importance of the two hypotheses continues (Sprague 1983; Wallace 1989; Crow 1993; Mitton 1993) and is unlikely to be settled until an unambiguous test is devised (H.-W. Deng, Y.-X. Fu and M. Lynch, unpublished results).

Testing dominance *vs.* the overdominance hypothesis is important for discerning mechanisms for the maintenance of genetic variability (Crow 1993). Some fairly recent studies tested the (over)dominance hypotheses and inferred nonadditivity of within-locus mutation effects by estimating the average ($\bar{h}$, the arithmetic mean) of the within-locus nonadditive coefficients (h_i) across loci (*e.g.*, Eanes *et al.* 1985; Crow 1993; Johnston and Schoen 1995). Although very appealing and intuitive, the validity of the method used in the above tests and

Address for correspondence: Osteoporosis Research Center, Department of Medicine, Creighton University, 601 N. 30th St., Suite 6787, Omaha, NE 68131. E-mail: deng@creighton.edu

inferences remains to be investigated. Additionally, estimating the harmonic mean of h_i under the dominance hypothesis is important for estimating genomic mutation rates—an important parameter in many evolutionary theories (Charlesworth *et al.* 1990; Deng and Lynch 1996a; H.-W. Deng and Y.-X. Fu, unpublished results) and for testing theories of transition from haploidy to diploidy (Perrot *et al.* 1991). However, partly because the often-used method (Mukai *et al.* 1972) of estimating $\bar{h}$ requires constructing homozygous lines, estimates of $\bar{h}$ are still very scarce, especially in outcrossing populations. Therefore, there is a need to develop alternative approaches to estimate $\bar{h}$ in outcrossing populations and to understand their statistical properties.

This study is purported to (1) develop a new approach to estimate $\bar{h}$ and the variance (σ_h^2) of h_i across loci in outcrossing populations that are capable of selfing; (2) examine the statistical properties of the newly developed approach and an earlier one that has been used extensively; and (3) examine the validity of testing the hypotheses for heterosis and inferring within locus (non)-additivity by estimating $\bar{h}$. First, we review an existing approach and develop a new one to estimate $\bar{h}$ and σ_h^2. Next, using computer simulations we study the statistical properties of the two approaches under the pure mode (*i.e.*, all loci are overdominant or all are dominant) and then under the mixed modes of dominance and overdominance (*i.e.*, some loci are overdominant and others dominant). The investigation of the statistical properties is important because they have never been formally investigated even for the widely applied Mukai's approach (Mukai *et al.* 1972). Then we study the estimation properties under variable mutation effects with and without lethals. Finally, we discuss the limitations of the practice of applying Mukai's approach in discriminating (over)dominance as the cause for heterosis and inferring (non)additivity of within-locus allelic effects.

TWO APPROACHES TO ESTIMATE $\bar{h}$ AND σ_h^2

Hypothesis testing and parameter estimation in statistics are two highly related topics. Unbiased and efficient estimation (estimation with a small sampling error) of a parameter generally forms a basis for a powerful test concerning that parameter. For a locus with the two alleles **A** and **a**, let the three genotypic values be, respectively:

AA	**Aa**	**aa**
1	$1 - h_i s_i$	$1 - s_i$

Then $h_i < 0$ implies overdominance, $h_i = 0.5$ additivity, $0 \le h_i \le 1$ ($h_i \neq 0.5$) dominance, and $h_i > 1$ implies underdominance. Note that throughout we use "dominant" or "dominance" to refer to the cases of $0 \le h_i \le 1$ ($h_i \neq 0.5$), which include both complete dominant or recessive, and partial dominant or recessive. For individual locus of quantitative traits, h_i is almost impossible to estimate, and thus $\bar{h}$ is normally estimated. Even for $\bar{h}$ (the arithmetic mean of h_i across loci), there currently is no method to directly estimate it. Often, $\bar{h}$ is approximately estimated by Mukai's method (Mukai *et al.* 1972), which is closely related to the earlier methods of Comstock and Robinson (1949) and Hayman (1954). Because the arithmetic mean is the most often used measure of the average, the investigation of the bias of Mukai's and our new method will be relative to the true $\bar{h}$. Throughout, h_i will refer to dominance coefficient at individual loci, $\bar{h}$ refers to the arithmetic mean of h_i across loci, and h refers to the general term dominance coefficient or the dominance coefficient with constant h_i across loci.

Mukai's approach: This approach was developed under the assumptions that dominance is the sole mode of within-locus genetic effects, the frequency of deleterious allele is very small, the population is at Hardy-Weinberg equilibrium, and mutation effects across loci are additive. It approximately estimates $\bar{h}$ by the slope of the regression of the outcrossed-progeny fitness (x) on the fitness sum (y) of the two corresponding parental homozygotes (Mukai *et al.* 1972; Simmons and Crow 1977; Crow and Simmons 1983)

$$\bar{h} \approx \frac{Cov(x,y)}{Var(y)}. \tag{1a}$$

σ_h^2 can be approximately estimated as (Mukai and Yamaguchi 1974)

$$\sigma_h^2 \approx \frac{Var(x)}{Var(y)} - \left(\frac{Cov(x,y)}{Var(y)}\right)^2. \tag{1b}$$

This approach is readily applicable to highly selfing populations (Johnston and Schoen 1995), where construction of homozygotes is easy. In most outcrossing populations, however, construction of pure homozygotes is usually difficult, except in a few organisms (such as Drosophila) where it can be relatively easily achieved by special chromosomal constructs (*e.g.*, Mukai *et al.* 1972; Eanes *et al.* 1985). Therefore, the following approach is developed to estimate $\bar{h}$ and σ_h^2 in self-compatible outcrossing populations without homozygous lines.

New approach: A wide variety of outcrossing plants and invertebrates are capable of selfing. In such populations, if we self a random sample of genotypes and obtain a number of selfed progeny from each parent to form selfed families, then $\bar{h}$ and σ_h^2 can be estimated. The data needed are the genotypic value of the parent (w) and the mean genotypic value (z) of the selfed progeny within each selfed family. Under the same assumptions as Mukai's approach outlined above, for a diallelic locus, we define the one-locus genetic effects as in Table 1. It can be easily seen from Table 1 that,

The variance of t': $$Var(t') \cong 2pqs_i^2 \tag{2a}$$

The variance of w': $$Var(w') \cong 2pqs_i^2h_i^2 \tag{2b}$$

TABLE 1

Distribution of mutation effects within selfed families in an outcrossing population

Parental genotypes	***AA***		***Aa***		***aa***
Frequencies	p^2		$2pq$		q^2
Mutation effect (w')	0		hs		s
Selfed offspring genotypes	***AA***	***AA***	***Aa***	***aa***	***aa***
Within-family frequencies	1	1/4	1/2	1/4	1
Mean mutation effects (z')	0		$h_is_i/2 + s_i/4$		s_i
$t' = 4z' - 2w'$	0		s_i		$2s_i$

The covariance of w' and t': $Cov(w',t') \cong 2pqs_i^2h_i$ (2c)

Then the expected regression of the mutation effects of outcrossed parent (w') on the quantity [4*(mean mutation effects of selfed family z') − 2*w'] is given by the ratio of the covariance to the variance, summed over all the relevant loci. This gives an approximate estimate of $\bar{h}$

$$\frac{Cov(w',t')}{Var(t')} = \frac{\Sigma p_iq_is_i^2h_i}{\Sigma p_iq_is_i^2} \approx \bar{h}. \quad (3a)$$

As with Mukai's approach, $\bar{h}$ is approximately estimated by the average of h_is at individual loci weighted by the genetic variance of the homozygotes. Although the above derivation is based on the mutational effects for ease of derivation, it can be easily shown (because $w = 1 - w'$ and $z = 1 - z'$) that the estimation can be performed on the original trait values

$$\frac{Cov(w,t)}{Var(t)} = \frac{\Sigma p_iq_is_i^2h_i}{\Sigma p_iq_is_i^2} \approx \bar{h} \quad (3b)$$

where $t = 4z - 2w$. Under constant mutation effects,

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

Additionally, we note that the ratio of the variance of w' to that of t' is

$$\frac{Var(w')}{Var(t')} = \frac{Var(w)}{Var(t)} = \frac{\Sigma p_iq_is_i^2h_i^2}{\Sigma p_iq_is_i^2} \approx \overline{h^2}. \quad (4a)$$

As with the estimates of $\bar{h}$, $\overline{h^2}$ is approximately estimated by the average of h_i^2s at individual loci weighted by the genetic variance of the homozygotes. Thus, an approximate σ_h^2 estimate is

$$\sigma_h^2 = \overline{h^2} - \bar{h}^2 \approx \frac{Var(w)}{Var(t)} - \left(\frac{Cov(w,t)}{Var(t)}\right)^2. \quad (4b)$$

In the above derivation, note that the assumption of Hardy-Weinberg equilibrium is not necessary; in fact, neither is it in Mukai's approach. One key assumption, though, is that the frequency of the rarer allele at any locus is low. This is essential in order for the approximation of Equation 2, a–c, to hold reasonably well, and also is true in deriving Equation 1, a and b (Mukai *et al.* 1972).

SIMULATIONS

The above derivations make a number of assumptions, for example, the within-locus nonadditive genetic effects are dominant and genetic effects across loci are additive. However, there is good evidence that genes for fitness or its components usually act multiplicatively (Morton *et al.* 1956; Crow 1986; Fu and Ritland 1996). To investigate the robustness of the above approaches under overdominance and the more reasonable mixed modes of dominance and overdominance at different loci, and to investigate their statistical properties, simulations are performed in which fitnesses are assumed to be multiplicative. To concentrate on studying the robustness of the approaches and the influence of the genetic process (selfing outcrossed individuals or outcrossing homozygous lines) on the estimation, all genotypic values are assumed to be measured accurately. In reality, this would require that each genotype be clonally replicated and assayed a very large number of times. Ignoring measurement error of genotypic values due to random environmental and developmental processes will likely inflate the sampling error of estimation, but unlikely bias the estimation and comparison of the two approaches under the same sample sizes of genotypes (Deng and Lynch 1996a).

This section is organized according to the presentation of different mutation effects across the genome, starting from the simplest case of constant effects, progressing to biologically more complex and plausible situations. Simulations will be described for each situation, respectively, and some necessary analytical results will be developed.

Constant mutation effects with dominance: Dominance (h_i) and selection (s_i) coefficients across loci are the same, that is, $h_i = h$, $s_i = s$.

Mukai's approach in outcrossing populations: Assume some random pairs of homozygotes are established from natural populations [such as with a special chromosome construct in Drosophila (Mukai *et al.* 1972)]. Their fit-

ness is $W = (1 - s)^n$, where n is the number of mutations randomly determined from a Poisson distribution with mean $U/(2hs)$, where U is the genomic mutation rate. This is because the mean number of mutations per genome in an outcrossing population is $U/(hs)$, and nearly all mutations are heterozygous in state (Deng and Lynch 1996a), and homozygous lines established are expected to carry only about one-half of the total genomic mutations of the outcrossed generation. Because genome size is usually very big and the mutation rate (μ) of each locus is very small, the mutant allele frequency (μ/hs at mutation-selection balance; Crow and Kimura 1970) is also very small. It is unlikely that the homozygotes established have mutations at the same loci (Charlesworth *et al.* 1990; Deng and Lynch 1996a), as corroborated by our computer simulations (H.-W. Deng, unpublished data). Therefore, throughout for dominant loci under mutation-selection equilibrium, the number of mutations in an outcrossed progeny is the sum of the number of mutations of its two parental homozygotes (n_m and n_f), but all at heterozygous state. Thus, the fitness of an outcrossed progeny is $W = (1 - hs)^{n_m + n_f}$. Because

$$\ln W = \ln(1 - s)^n \approx -ns, \qquad (s << 1.0) \quad (5)$$

$\ln(W)$ therefore approximates the fitness reduction under the additive fitness function. Thus, throughout, logarithmic transformation is performed for fitness, and then Equation 1a is employed to estimate h. Throughout, in simulating Mukai's approach, 20 pairs of homozygotes and their hybrids are used.

Mukai's approach in selfing populations: Homozygous lines are readily obtainable. Simulations are performed as above, except that n in the fitness function $W = (1 - s)^n$ is randomly determined from a Poisson distribution with mean $U/(2s)$ (Charlesworth *et al.* 1990; Deng and Lynch 1996a).

New approach: A number of randomly outcrossed parents are sampled, with each having fitness $W = (1 - hs)^n$, where n is the number of mutations randomly determined from the Poisson distribution with mean $U/(hs)$ (Deng and Lynch 1996a). A number of selfed progeny genotypes for each parent is obtained, with each selfed genotype being determined by allowing the n parental heterozygous loci to segregate randomly into **AA**, **Aa**, and **aa** classes with respective probabilities of 0.25, 0.5, and 0.25. Letting n_1 and n_2 (both resulting from random segregation) be the numbers of heterozygous and homozygous loci containing mutations in a selfed offspring, its fitness is $W(n_1,n_2) = (1 - hs)^{n_1} (1 - s)^{n_2}$. Equation 3c is used to estimate h. For the new approach in outcrossing populations 20 parents are employed throughout, each having 50 selfed progeny genotypes.

Constant mutation effects with over(under)dominance: Under over(under)dominance, the key assumption for estimating h, that the frequency of one homozygote at any polymorphic locus is low, is unlikely to be valid in outcrossing populations; however, in at least two situations, it may hold and Mukai's approach should be applicable regardless of h value. The first is in highly selfing populations, where overdominance for mutations is unlikely to be responsible for the maintenance of genetic variability (Kimura and Ohta 1971). The second is when lines are obtained by generations of mutation accumulation from homozygous replicate lines. If the mutation rate per locus is low, overdominant mutants are unlikely to achieve high frequencies.

TABLE 2

Estimating h by Mukai's approach is not influenced by different distributions of the number of loci with dominant or overdominant mutations per genome

Distributions	$\hat{h}$ dominance ($h = 0.36$, $s = 0.03$)	$\hat{h}$ overdominance ($h = -0.2$, $s = 0.03$)
Poisson	0.357 (0.000)	−0.196 (0.000)
Uniform	0.357 (0.000)	−0.196 (0.000)
Exponential	0.357 (0.000)	−0.196 (0.000)
Normal	0.357 (0.000)	−0.196 (0.000)

The simulations are performed under constant mutation effects. In simulations, the means of the Poisson, exponential, and normal distributions are all 25, the variance for the normal distributions is 25, the range of the uniform distribution is from 0 to 25. Simulations are also performed for a range of other parameters, and results not shown here indicate similar conclusions.

Generally, the distribution of the number of loci per genome with overdominant mutations is not clear. Simulations are performed for different distributions of the number of loci having (over)dominance mutations. The simulation procedures are the same as before for the dominance case, except that the distribution of the number of loci under (over)dominance is different. The results (Table 2) indicate little influence on the estimation by Mukai's approach under very different simulated distributions of the number of loci per genome having (over)dominant alleles. This is not unexpected, because the derivation of the approach is based on the one-locus results and within-family data and extended to multiple loci under additive mutation effects across loci. No assumption was made as to the distribution of the number of loci having (over)dominant alleles per genome. Therefore, as before (*i.e.*, Poisson distribution of the polymorphic loci is used) except that we let the parameter $h < 0$ (overdominance) or $h > 1$ (underdominance), simulations are performed for Mukai's approach for selfing populations. For Mukai's approach using mutation accumulation lines, the principle is the same and the simulation and results are similar, and hence not presented.

Mixed dominance and overdominance: *Lines from highly selfing populations or derived by mutation accumulations:* It is possible that both dominance and overdominance underlie heterosis and that new mutations of either nature can

occur in the genome. Some interesting questions are the following: In highly selfing populations, what is the major cause of heterosis? In the genome, what is the major type of new mutations for heterosis? Can we answer these questions from $\hat{h}$? Throughout, a circumflex (^) indicates an estimated value.

With both dominant and overdominant loci present, if fitness effects of individual loci are independent, as is the case for the multiplicative fitness function, the heterosis due to dominance and overdominance is independent. Let the mutation rate to deleterious alleles and constant mutational effects under dominance be U_1, h_1, and s_1, respectively. In large highly selfing populations, the expected heterosis (the ratio of the mean fitness of the outcrossed offspring generation W_o to that of the homozygous parental generation W_p), which is due to dominance (δ_d), is then (Charlesworth *et al.* 1990; Deng and Lynch 1996a)

$$\delta_d = \exp(-U_1 h_1)/\exp(-U_1/2) = e^{U_1(0.5 - h_1)} \quad (6a)$$

New mutational occurrence in the genome most likely follows a Poisson distribution, whether it involves dominant or overdominant mutations. Throughout, mutations fitting the (over)dominance hypothesis will be referred to as (over)dominant mutations. In highly selfing populations, mutant alleles will be maintained by mutation-selection balance, regardless of their (over)dominance. This is because within a selfing line, frequent selfing will quickly bring any polymorphic locus into homozygous state. Under obligate selfing, different selfing lines are essentially reproductively isolated from each other, and thus overdominance will not contribute to the maintenance of genetic variability. Hence, as in the dominance case, we assume that the number of loci with overdominant mutants (n) (all in homozygous state) per genome in selfing populations is Poisson distributed with mean $\bar{n}$ and constant effects h_2 and s_2. If the genomic mutation rate to overdominant (but less fit) allele **a** is U_2, it can be easily shown that at mutation-selection equilibrium, $\bar{n} = U_2/(2s_2)$ (Charlesworth *et al.* 1990; Deng and Lynch 1996a). The expected heterosis due to overdominance (δ_o) is then

$$\begin{aligned}
\delta_o &= \left(\sum_{n=0}^{\infty}(1 - h_2 s_2)^n 2(\bar{n})^n e^{-2\bar{n}}/n!\right)/\left(\sum_{n=0}^{\infty}(1 - s_2)^n(\bar{n})^n e^{-\bar{n}}/n!\right)\\
&= \left(\sum_{n=0}^{\infty}(2\bar{n} - 2h_2 s_2 \bar{n})^n e^{-2\bar{n}}/n!\right)/\left(\sum_{n=0}^{\infty}(\bar{n} - s_2\bar{n})^n e^{-\bar{n}}/n!\right)\\
&= \frac{\exp(-2h_2 s_2\bar{n})}{\exp(-s_2\bar{n})}\left(\sum_{n=0}^{\infty}(2\bar{n} - 2h_2 s_2\bar{n})^n e^{-(2\bar{n} - 2h_2 s_2\bar{n})}/n!\right)/\\
&\qquad \left(\sum_{n=0}^{\infty}(\bar{n} - s_2\bar{n})^n e^{-(\bar{n} - s_2\bar{n})}/n!\right)\\
&= \exp(-2h_2 s_2\bar{n})/\exp(-s_2\bar{n})\\
&= e^{(1 - 2h_2)s_2\bar{n}} = e^{U_2(1 - 2h_2)/2}. \qquad (6b)
\end{aligned}$$

The total heterosis is $\delta = E(W_o)/E(W_p) = \delta_d\delta_o$. The contribution to heterosis from dominance relative to overdominance can then be measured by the index

$$\alpha = \frac{\delta_d}{\delta_o} = \exp\left(U_1(0.5 - h_1) - (1 - 2h_2)s_2\bar{n}\right). \quad (7)$$

α is an important index that will be used more later. If $\alpha > 1$, the primary cause of heterosis is dominance; if $\alpha < 1$, it is overdominance; otherwise, dominance and overdominance contribute about equally to heterosis. The smaller the α, the larger the contribution to heterosis from overdominance.

Using Equation 7, we can determine the number of overdominant loci (N_o) when dominance and overdominance contribute equally to heterosis

$$N_o = \frac{U_1(0.5 - h_1)}{s_2(1 - 2h_2)}. \quad (8)$$

$\bar{n} > N_o$, $\bar{n} < N_o$, $\bar{n} = N_o$, respectively, correspond to $\alpha < 1$, $\alpha > 1$, $\alpha = 1$.

In simulations, the genome contains both dominant and overdominant loci, all at mutation-selection equilibrium. In the parental generation, the number of dominant loci in each individual is sampled from the Poisson distribution of mean $U/(2s_1)$ (Charlesworth *et al.* 1990; Deng and Lynch 1996a), and that for the overdominant loci is determined from a Poisson distribution with mean $\bar{n}$ [$= U_2/(2s_2)$]. Simulations are then performed as before for Mukai's approach.

Homozygous lines constructed from outcrossing populations: In outcrossing populations, with overdominance the key assumption that the rarer allele at any polymorphic locus is of a low frequency is often invalid. Thus, both Mukai's and our new approaches should not be used. However, there have been some practice and data on estimating $\bar{h}$ by Mukai's approach using homozygous lines constructed from outcrossing populations such as Drosophila (*e.g.*, Mukai *et al.* 1972; Mukai and Yamaguchi 1974; Eanes *et al.* 1985). Therefore, under mixed dominance and overdominance, we investigate whether those estimates are of any use and whether they can discern the predominant mode of genetic effects for heterosis.

Let U_1, h_1, s_1, h_2, s_2, δ_d, δ_o, α, and N_o be defined as before. In large outcrossing populations, upon selfing (Deng and Lynch 1996a)

$$\begin{aligned}
\delta_d &= \exp(-U_1)/\exp(-U_1(h_1 + 0.5)/2h_1)\\
&= e^{U_1(1 - 2h_1)/(4h_1)}. \qquad (9a)
\end{aligned}$$

Overdominant alleles are mainly maintained by balancing selection in outcrossing populations. If there are n overdominant loci with constant effects h_2 and s_2

$$\delta_o = \left(\frac{2(2h_2 - 1)^2 - 2h_2^2 s_2(2h_2 - 1)}{2(2h_2 - 1)^2 - s_2 h_2(2h_2^2 + h_2 - 1)}\right)^n. \quad (9b)$$

Then

$$\alpha = \frac{\delta_d}{\delta_o} = \left(\frac{2(2h_2 - 1)^2 - s_2h_2(2h_2^2 + h_2 - 1)}{2(2h_2 - 1)^2 - 2h_2^2s_2(2h_2 - 1)}\right)^n \exp\left(\frac{U_1(1 - 2h_1)}{4h_1}\right) \tag{10}$$

By Equation 9, a and b, we can determine N_o if we set $\alpha = 1$

$$N_o = \frac{U_1(1 - 2h_1)}{4h_1 * \ln\left(\frac{2(2h_2 - 1)^2 - 2h_2^2s_2(2h_2 - 1)}{2(2h_2 - 1)^2 - s_2h_2(2h_2^2 + h_2 - 1)}\right)}. \tag{11}$$

Again, $n > N_o$, $n < N_o$, $n = N_o$, respectively, are equivalent to $\alpha < 1$, $\alpha > 1$, $\alpha = 1$, which correspond to situations where dominance contributes to heterosis less than, more than, and equal to overdominance, respectively.

In the simulations, the homozygous lines constructed from large populations at mutation-selection equilibrium contain both dominant and overdominant loci. The number of loci homozygous for deleterious alleles (under the dominance hypothesis) is determined from a Poisson distribution of mean $U/(2h_1s_1)$, as explained before. The alleles at the n overdominant loci are determined from random uniform variables (ξs) (from 0 to 1) with allele **A** being chosen if $\xi \leq h_2 - 1/2h_2 - 1$ [where $h_2 - 1/2h_1 - 1$ is the equilibrium frequency of **A** allele (Crow 1986)], and allele **a** chosen otherwise. This simulation procedure allows identity by state for mutant alleles at the overdominant loci in different inbred lines. In outcrossed progeny, the number of the loci (all in heterozygous state) having dominant mutations is the sum of the number of mutations of its two homozygous parents; the genotypes at the n overdominant loci are determined by the overdominant alleles at these loci in its two homozygous parents. Other aspects being the same as before, simulations for Mukai's approach are then performed.

Variable mutation effects under dominance: Deleterious mutation effects across loci (s_i and h_i) are not constant. The few available data suggest that s_i has a roughly exponential distribution (Gregory 1965; Mackay *et al.* 1992; Keightley 1994). As in Deng and Lynch (1996a), we use the exponential distribution to model s_i

$$f(s_i) = \frac{1}{\bar{s}} \exp(-s_i/\bar{s}), \qquad (1 > s > 0) \tag{12a}$$

where $\bar{s}$ is the mean of s_i. Little information exists on the distribution of h_i, but biochemical arguments suggest an inverse relationship between s_i and h_i (under dominance hypothesis), mutant alleles with larger effects tending to be more recessive (Kacser and Burns 1981). The few available data are consistent with this idea (Crow and Simmons 1983). Therefore, as in Deng and Lynch (1996a), the following function is adopted to approximate the inverse relationship between s_i and h_i

$$h_i = \frac{1}{2} \exp(-13s_i), \tag{12b}$$

which is in rough accordance with the few available data (Crow and Simmons 1983; Deng and Lynch 1996a). Under the above distribution and function, the mean and the variance of h_i is

$$\bar{h} = \frac{1}{(26\bar{s} + 2)} \qquad \sigma_h^2 = \frac{169_s^{-2}}{(26\bar{s} + 1)(13\bar{s} + 1)^2} \tag{12c}$$

To evaluate how Equations 1b and 4b perform for estimating σ_h^2 and how Equations 1a and 3b behave in estimating $\bar{h}$ under variable mutation effects, simulations are conducted under dominance. As in Deng and Lynch (1996a), we use a discrete version of the exponential distribution of Equation 12a by dividing the entire range of s_i (0, 1) into 200 classes of width 0.005. Each parental individual in a simulation is then randomly assigned a number of mutations from each of these classes by drawing from a Poisson distribution with means of $Up_i/(h_is_i)$ in outcrossing and of $Up_i/(2s_i)$ in selfing populations, respectively, where p_i is the density of the mutation distribution in the ith class.

Because the estimates are usually biased under the variable mutation effects, we compute their *MSE* (mean square error) for comparison: $MSE = E(\hat{x} - E(x))^2 = Var(\hat{x}) + (\bar{\hat{x}} - E(x))^2$, where $\bar{\hat{x}}$ stands for the estimated mean. Note that when $\bar{\hat{x}}$ is unbiased, *MSE* is simply the variance of $\hat{x}$.

The effects of lethals: The above study for variable mutation effects assumes that the genome contains no lethal mutations. This is a good assumption for selfing populations, where lethal mutations cannot survive for more than a few generations, due to frequent exposure to selection in homozygous state. In outcrossing populations, this assumption does not hold (Simmons and Crow 1977; Crow and Simmons 1983), as lethals are usually shielded from selection in heterozygous state by their low degree of dominance. With Mukai's approach, lethals will not appear in final homozygous lines constructed. During generations of inbreeding to construct homozygotes, lines homozygous for lethals will be immediately lost. Therefore, Mukai's approach gives the estimates only for mildly deleterious mutations. Hence, we only evaluate our new approach under variable mutation effects with lethals in outcrossing populations.

Lethals ($s_L = 1$) compose approximately 1% of the genomic mutations, and h_L for lethals is estimated to be about 0.02 (Simmons and Crow 1977; Crow and Simmons 1983). The simulations in outcrossing populations with lethals and variable mutation effects are identical in all respects to those in the previous section, but with an additional low genomic mutation rate (1%) to lethals ($s_L = 1$, $h_L = 0.02$). In simulations, selfed offspring homozygous for lethals will be excluded from analysis, as is most likely the practice in actual experi-

TABLE 3

Estimating *h* under constant mutation effects

	Mukai's approach		New approach
	Selfing populations	Outcrossing populations	Outcrossing populations
h	$\hat{h}$	$\hat{h}$	$\hat{h}$
Dominance			
0.2	0.198 (0.000)	0.198 (0.000)	0.193 (0.004)
0.4	0.396 (0.000)	0.396 (0.000)	0.386 (0.016)
0.6	0.596 (0.000)	0.596 (0.000)	0.576 (0.030)
0.8	0.798 (0.000)	0.798 (0.000)	0.763 (0.039)
Overdominance			
−0.2	−0.196 (0.000)		
−0.4	−0.392 (0.000)		
Underdominance			
1.2	1.204 (0.000)		
1.4	1.409 (0.000)		

In all the simulations, the selection coefficient s equals 0.03, and the genomic mutation rate U equals 1, which approximates those from mutation accumulation experiments (Mukai *et al.* 1972; Kibota and Lynch 1996) without assuming dominance or overdominance. Throughout, the numbers reported are the means and standard deviations (SDs, numbers within parentheses) over 100 simulations; a SD of 0.000 indicates a SD <0.0005.

ments. The selfed family means are for those offspring that do not contain homozygous lethals.

RESULTS

Constant mutation effects with either dominance or (under)overdominance (Table 3): The bias and sampling variance of estimates by both approaches is very small, especially for Mukai's approach. The very small bias is because the logarithmic transformation of the multiplicative fitness function is used to approximate the additive fitness function assumed by the derivations (Deng and Fu 1997). In selfing and outcrossing populations, estimates by Mukai's approach yield nearly identical results, except for the undetectable difference in the sampling errors. The new approach has slightly larger bias and larger sampling variance. This is partly because the estimate is subject to more sampling error, as the mean of each selfed family is estimated by a limited number of progeny. In highly selfing populations or in lines from mutation accumulations where the key assumption holds, Mukai's approach can estimate h accurately under over(under)dominance.

Mixed dominance and overdominance: *Lines from highly selfing populations or derived by mutation accumulations (Table 4):* When the contributions to heterosis from dominance and overdominance are about the same ($\alpha \approx 1$), the $\hat{h}$s are always positive with small sampling errors (Table 4), which favors the dominance hypothesis. Only with relatively large overdominance effects ($h < -0.1$) and overwhelming contributions from overdominant loci ($\alpha \sim 0.05$), is $\hat{h} < 0$. Simulation results not shown

TABLE 4

Estimating *h* under mixed dominance and overdominance in highly selfing populations by MUKAI's approach

h_2	$\bar{n}$	α	$\hat{h}$
−0.01	0	1.15	0.357 (0.000)
	4	1.02	0.285 (0.041)
	5	0.99	0.270 (0.039)
	10	0.85	0.231 (0.047)
	100	0.05	0.039 (0.029)
	200	0.02	0.013 (0.023)
−0.1	0	1.15	0.357 (0.000)
	3	1.03	0.268 (0.051)
	4	1	0.268 (0.051)
	10	0.80	0.200 (0.058)
	100	0.03	−0.037 (0.036)
−0.3	0	1.15	0.357 (0.000)
	2	1.04	0.290 (0.043)
	3	1	0.264 (0.050)
	10	0.71	0.133 (0.084)
	100	0.01	−0.207 (0.051)

U_1, h_1, and s_1 are the parameters of dominant mutations, while h_2 and s_2 are the parameters of overdominant mutations. $\bar{n}$ is the mean number of overdominant loci per genome. α is a measure of relative contribution of dominance and overdominance to heterosis (see text for a detailed definition). Simulations for data were performed under $U_1 = 1$, $h_1 = 0.36$, $s_1 = 0.03$, and $s_2 = 0.03$, which approximate previous experimental data (Mukai *et al.* 1972; Kibota and Lynch 1996). Simulation results not shown here indicate no influence on the conclusion by employing different parameters of U_1, h_1, and s_1. Mutation effects are assumed to be constant for dominant and overdominant mutations.

TABLE 5

Estimating *h* under mixed dominance and overdominance in outcrossing populations by MUKAI's approach

h_2	n	α	$\hat{h}$
−0.01	0	1.21	0.357 (0.000)
	1308	1	0.255 (0.041)
	2600	0.83	0.193 (0.046)
	5000	0.58	0.119 (0.035)
−0.1	0	1.21	0.357 (0.000)
	141	1	0.247 (0.052)
	280	0.83	0.187 (0.057)
	500	0.61	0.149 (0.064)
	1000	0.31	0.105 (0.072)
	2500	0.04	0.088 (0.084)
−0.3	0	1.21	0.357 (0.000)
	53	1	0.261 (0.063)
	100	0.84	0.222 (0.079)
	500	0.20	0.086 (0.099)
	1000	0.03	0.051 (0.107)

U_1, h_1, s_1, h_2, s_2, and α are denoted in the legend to Table 4. n is the number of overdominant loci per genome. Simulations for data in this table are performed under $U_1 = 1$, $h_1 = 0.36$, $s_1 = 0.03$, and $s_2 = 0.03$. Simulation results not shown here indicate no influence on the conclusion by employing different parameters of U_1, h_1, and s_1. Mutation effects are assumed to be constant for dominant and overdominant mutations.

here indicate similar conclusions for lines from mutation accumulations.

Homozygous lines constructed from outcrossing populations (Table 5): With $\alpha \approx 1$, $\hat{h}$s are always positive with small sampling errors. Unlike estimates from selfing populations or mutation accumulation where the key assumption holds, $\hat{h}$ is always positive when employing Mukai's approach in outcrossing populations. Even with pure overdominant genetic effects, $\hat{h}$ is almost always greater than 0. For example, if the genome contains only 200 overdominant loci with $h_2 = -0.1$ and $s_2 = 0.03$, the equilibrium frequency for allele **A** is 0.917 and **a** 0.083. Applying Mukai's approach, we obtain $\hat{h}$ = 5.84E-4 (1 SD = 0.05738). This is because the key assumption of the approach that rarer alleles at all loci are of low frequencies is violated at overdominant loci.

In summary, with mixed dominance and overdominance jointly causing heterosis, Mukai's approach cannot be employed to distinguish dominance *vs.* overdominance. On the other hand, it is encouraging to see that the presence of overdominant loci does not greatly bias the estimation of *h* for the dominant alleles. Even with $\alpha \approx 1$ (*i.e.*, equal contribution of dominance and overdominance to heterosis), $\hat{h}$ is about 70% of the true *h* value for the dominant alleles (Table 5). Therefore, the $\hat{h}$s estimated by Mukai's approach in outcrossing populations such as Drosophila (*e.g.*, Mukai *et al.* 1972; Mukai and Yamaguchi 1974; Eanes *et al.* 1985) may represent values that are underestimated (but not to a great extent) for dominant loci. These results may be better understood if we recall that, by derivation (Mukai *et al.* 1972), Mukai's method estimates an average of h_is at individual loci weighted by the genetic variance of the homozygotes. This can lead to very peculiar results as reflected by the simulation results here. In an extreme case involving loci of symmetrical overdominance ($s_2 = 0$), even though the loci contribute to inbreeding depression, they are assigned a zero weight in estimating the average of h_is across loci. This reflects the fact that the contributions to inbreeding depression and to the estimation of the average dominance coefficient from individual loci are not exactly the same. Therefore, if dominant and overdominant loci coexist in the genome, inferring which is the major cause for heterosis by the sign of $\hat{h}$s is invalid.

Variable mutation effects under dominance (Table 6): Under variable mutation effects without lethals, $\bar{\hat{h}}$s are always biased (underestimated compared to $\bar{h}$, the arithmetic mean of h_i across loci). The biases of $\bar{\hat{h}}$s in selfing populations using Mukai's approach are the smallest, and those in outcrossing populations with Mukai's approach are slightly smaller than with our new approach. The bias increases with an increasing σ_h^2. The bias patterns are not unexpected. For example, the smaller the σ_h^2, the more constant is h_i across loci, and less bias should be expected for $\bar{\hat{h}}$ because as it has been shown that under constant *h*, there is no estimation bias.

A similar bias pattern is observed for $\hat{\sigma}_h^2$; however, $\hat{\sigma}_h^2$ may not always be biased. The ratio $\hat{\sigma}_h^2/\sigma_h^2$ decreases with an increasing σ_h^2 so that $\hat{\sigma}_h^2$ is upwardly biased when σ_h^2 is relatively small, and downwardly biased when σ_h^2 is relatively large.

The biases may have come from at least two sources: (1) The logarithmic transformation of the multiplicative fitness function is employed to approximate the additive fitness function (Equation 5). (2) The definitions of the estimates of $\bar{\hat{h}}$, $\hat{\sigma}_h^2$ (Equations 1, a and b, 3b, and 4b) are not the usual statistical definitions of $\bar{h}$ and σ_h^2. Currently, there is no method to estimate $\bar{h}$ and $\overline{h^2}$ directly, and thus they are approximately estimated by the averages of h_is or h_i^2s at individual loci both weighted by the genetic variance of the homozygotes, instead of by their frequencies as usual. Despite this, it is encouraging that, with the most likely parameters ($U = 1$, $\bar{s} = 0.03$, $\bar{h} = 0.36$, Mukai *et al.* 1972; Lynch *et al.* 1996), the biases for $\bar{\hat{h}}$ and $\hat{\sigma}_h^2$ are reasonably small. With Mukai's approach, $\hat{\sigma}_h^2/\sigma_h^2 = 1.07$, 1.32 and $\bar{\hat{h}}/\bar{h} = 0.71$, 0.48, respectively, in selfing and outcrossing populations; with our new approach, $\hat{\sigma}_h^2/\sigma_h^2 = 1.52$ and $\bar{\hat{h}}/\bar{h} = 0.47$ in outcrossing populations.

The effects of lethals in outcrossing populations (Table 7): $\bar{\hat{h}}$ and $\hat{\sigma}_h^2$ are usually biased, but the bias is much smaller than when lethals are absent. The bias of $\bar{\hat{h}}$s

TABLE 7

Estimating mean and variance of h_is in outcrossing populations by the new approach with mutations of variable effects and lethals

$\tilde{s}$	$\tilde{h}$	$\tilde{\sigma}_h^2$	$\bar{s}$	$\bar{h}$	σ_h^2	$\hat{\bar{h}}$	$\hat{\sigma}_h^2$
0.02	0.439	4.09E-3	0.01	0.441	2.63E-3	0.388 [0.056]	8.98E-3 [4.81E-3]
0.038	0.359	1.13E-2	0.03	0.360	1.11E-2	0.218 [0.144]	1.25E-2 [0.66E-2]
0.057	0.303	1.68E-2	0.05	0.303	1.69E-2	0.109 [0.196]	0.99E-2 [0.89E-2]

$\tilde{s}$, $\tilde{h}$, and $\tilde{\sigma}_h^2$ are the parameters (means and variance), including lethals. $\bar{s}$, $\bar{h}$, and σ_h^2 are the parameters for the mildly deleterious mutations only (lethals excluded). $\hat{\bar{h}}$ and $\hat{\sigma}_h^2$ are estimates for all mutations, including lethals. Lethal mutation rate is 0.01 per genome per generation, and the mean number of lethals per genome in outcrossing parental generations at mutation-selection equilibrium is 0.5. To facilitate comparisons with the data in Table 6 and also because the main interests are to estimate the parameters for mildly deleterious mutations in most evolutionary genetics theory, *MSE*'s in this table are computed based on the parameters for the mildly deleterious mutations.

increases with an increasing $\tilde{\sigma}_h^2$ ($\sim$ indicates the parameter for all mutations including lethals). A similar bias pattern is observed for $\hat{\sigma}_h^2$ and $\hat{\sigma}_h^2/\tilde{\sigma}_h^2$ decreases with an increasing $\tilde{\sigma}_h^2$ so that $\hat{\sigma}_h^2$ is upwardly biased when $\tilde{\sigma}_h^2$ is relatively small, and downwardly biased when $\tilde{\sigma}_h^2$ is relatively large. Again, it is encouraging that with the most likely parameters ($U = 1$, $\tilde{s} = 0.038$, $\tilde{h} = 0.359$), the biases for $\hat{\bar{h}}$ and $\hat{\sigma}_h^2$ are reasonably small ($\hat{\sigma}_h^2/\tilde{\sigma}_h^2 = 1.10$ and $\hat{\tilde{h}}/\tilde{h} = 0.607$). Actually, they are much smaller than those by Mukai's approach when lethals are absent in outcrossing populations (Tables 5 and 6), and nearly as small as in selfing populations using Mukai's approach. The same conclusion holds for comparison with *MSE* that includes both bias and sampling variance.

DISCUSSION

In this study, we develop a new approach to estimate $\bar{h}$ in self-compatible outcrossing populations. It does not require the construction of homozygous lines, which is usually difficult. We also investigate the statistical properties of our new approach and the widely used Mukai's approach. Under the assumption of constant effects and that either dominance or overdominance is the genetic cause of heterosis, h can be estimated satisfactorily with small sampling errors. This may then, indeed, form a basis for powerful tests to discriminate between the overdominance and dominance hypotheses by the sign of $\hat{h}$s. However, if dominant and overdominant mutations coexist in the genome, which may be more plausible, then inferring which is the dominant cause for heterosis by the sign of $\hat{h}$s is misleading. Estimates of $\bar{h}$ and σ_h^2 depend on the parameter values. Close estimation of σ_h^2 is possible with the most likely parameters of $\bar{s}$ and $\bar{h}$ and the likely relationship between s_i and h_i. In self-compatible outcrossing populations with mutations of variable effects and lethals, our new approach is better than Mukai's, not only because no homozygous lines need to be constructed but also because of the better statistical performance reflected by the smaller bias and *MSE* of the estimates.

We developed a methodology in natural outcrossing

TABLE 6

Estimating mean and variance of variable h_is across loci under dominance

			Mukai's approach				New approach	
			Selfing populations		Outcrossing populations		Outcrossing populations	
$\bar{s}$	$\bar{h}$	σ_h^2	$\hat{\bar{h}}$	$\hat{\sigma}_h^2$	$\hat{\bar{h}}$	$\hat{\sigma}_h^2$	$\hat{\bar{h}}$	$\hat{\sigma}_h^2$
0.01	0.44	2.63E-3	0.391 [0.052]	4.23E-3 [2.71E-3]	0.379 [0.063]	5.35E-3 [3.74E-3]	0.375 [0.067]	946E-3 [8.20E-3]
0.03	0.36	1.11E-2	0.256 [0.109]	1.19E-2 [0.64E-2]	0.172 [0.190]	1.47E-2 [0.74E-2]	0.169 [0.194]	1.50E-2 [0.87E-2]
0.05	0.30	1.69E-2	0.170 [0.134]	1.62E-2 [0.83E-2]	0.049 [0.252]	0.68E-2 [1.07E-2]	0.043 [0.258]	0.60E-2 [1.13E-2]

$U = 1$ for all the parameters simulated. The simulation conditions are the same as those in Table 2. Numbers within brackets in Tables 6 and 7 are the square root of *MSE* of the estimates, which are reduced to SD when estimates are unbiased.

and selfing populations to estimate $\bar{h}$ as well as the genomic mutation rate and the mean homozygous effects $\bar{s}$ (Deng and Lynch 1996a, 1997). The approach utilizes the information of the mean and genetic variance before and after selfing/outcrossing in outcrossing/selfing populations. Because more information from the data is used, more parameters concerning genomic mutations can be estimated by our previous approach (Deng and Lynch 1996a, 1997). Additionally, when estimating several genomic mutation parameters simultaneously under the same sample size, it is generally superior, statistically, to other available methods as revealed by our computer simulations (Deng and Fu 1997). However, our newly developed method here utilizes the information that was not considered by Deng and Lynch (1996a, 1997), that is, the covariance between the parent and the mean of the selfed progeny. As in the method of Deng and Lynch (1996a, 1997), our newly developed method also only requires two generations of husbandry in the laboratory. Our newly developed method requires multiple selfed progeny from each parent, while this requirement in Deng and Lynch (1996a) is eliminated by our later work (Deng and Lynch 1997). However, replications of genotypes, which are not feasible for many outcrossing populations, are not required for our newly developed method, although they are currently required in the methods of Deng and Lynch (1996a, 1997). Furthermore, estimation of σ_h^2 is possible by the methods investigated here but is not possible by our earlier methods (Deng and Lynch 1996a, 1997). Different methods developed (such as those discussed here and that in H.-W. Deng and Y.-X. Fu, unpublished results) may supplement each other in different situations, so that estimating the parameters of deleterious genomic mutations can be accomplished in a wider range of taxa.

We concentrate on studying the most plausible multiplicative mutation effects. Although epistatic mutation effects have been speculated and may be possible, their detection is a very difficult empirical problem, and little convincing information exists on the subject. We therefore do not study their effects here. The effects of synergistic mutation have been investigated for estimating genomic mutation rate U (Charlesworth *et al.* 1990) and U, $\bar{h}$, $\bar{s}$ (Deng and Lynch 1996a), indicating that the effects of even strong putative synergism are slight. Linkage equilibrium for quantitative traits has been subject to extensive studies for a long period of time (*e.g.*, Lewontin 1985; Houle 1989; Zapata and Alverez 1992, 1993; Lynch and Deng 1994; Deng and Lynch 1996b). Given the inconsistent empirical results from the long time studies, we choose to study in detail its potential effects on the estimation in our later studies. Mutation rate at any locus is generally small; hence, the key assumption that the rare allele has low frequency is likely to be valid in populations at mutation-selection balance, whether it influences fecundity or viability. However, for populations not at equilibrium, such as those resulting from recent admixture, care needs to be taken to see that the key assumption holds reasonably well before applying the methods investigated here.

Due to the lack of knowledge of the statistical properties, it has been a practice (even until fairly recently) to discriminate dominance and overdominance hypotheses by estimated $\hat{h}$s with Mukai's approach (Crow 1993; Johnston and Schoen 1995). However, our studies show that this may be valid only when one mode of constant genetic effects (either dominance or overdominance) underlies heterosis or mutations. Although it may be more plausible that both dominance and overdominance underlie heterosis or new mutations, the conclusions from this approach should be treated with caution. Estimated $\hat{h}$s with Mukai's approach have also been used to infer additivity or nonadditivity of within-locus mutation effects (Eanes *et al.* 1985; Johnston and Schoen 1995). This would not only need to assume that either dominant or overdominant alleles underlie heterosis or new mutations but also that mutation effects across loci are constant, so that h can be estimated with little bias. Our simulation results indicate that under more reasonable conditions (variable mutation effects and/or mixed dominance and overdominance), $\bar{h}$ cannot be estimated without bias by Mukai's approach. Thus, inferring (non)additivity of within-locus mutation effects by $\hat{h}$s with Mukai's approach may also be invalid. In outcrossing populations, the estimation of $\bar{h}$ must be based on the assumption that overdominance does not contribute to heterosis. Otherwise, the key assumption does not hold. As shown, applying Mukai's approach in outcrossing populations will result in an underestimation of $\bar{h}$ for the dominant alleles, if loci with overdominant alleles exist elsewhere in the genome.

We concentrate on studying the estimation of $\bar{h}$ and σ_h^2 under the hypothesis that dominance and overdominance are the only two genetic causes of heterosis. This is not necessarily true (Richey 1942; Minvielle 1987; Schnell and Cockerham 1992). If a population is in gamatic phase disequilibrium, fitness is multiplicative, and loci do not recombine freely, then change of mean upon inbreeding can happen even under pure additive genetic effects. However, the contribution to heterosis from this kind of genetic source (additive-by-additive epistatic effects) is generally of minor importance, compared with those from dominance/overdominance (Schnell and Cockerham 1992). Therefore, the conclusions from our simulations, ignoring such genetic sources of heterosis, are unlikely to be much affected. Because the arithmetic mean is the most often-used measure of the average, we compare the bias of $\hat{h}$ relative to $\bar{h}$. However, it needs to be keep in mind that sometimes other measures of the average (such as the harmonic mean) of h_i may be preferred (Deng and Lynch 1996a).

Our results indicate that estimating $\bar{h}$ in selfing popu-

lations may be slightly better than in outcrossing populations, in terms of the degree of bias and sampling variance. However, the parameter $\bar{h}$ in outcrossing populations may not be the same as that in selfing populations due to the mating system difference. Hence, there is a need to estimate $\bar{h}$ in outcrossing populations. Although in outcrossing populations Mukai's approach is slightly better than our new approach in terms of bias and sampling variance in the absence of lethals, it requires construction of homozygous lines. Even in the most extreme case of inbreeding—selfing, the genomic homozygosity is only expected to be reduced by one-half each generation (Crow and Kimura 1970). After 10 generations, the residual genomic heterozygosity is $\sim$0.1% of its original value before selfing. How the residual genomic heterozygosity would affect the estimation of $\bar{h}$ is not clear. Furthermore, with lethals present in the genome in outcrossing populations, which more often is the case, our new approach is actually better than Mukai's in terms of sampling variance and bias, and nearly as good as Mukai's approach with selfing populations. Therefore, our new approach is better than Mukai's in self-compatible outcrossing populations.

When implementing the two methods here, some practical issues need to be considered. The discussion of these practical issues cannot possibly be exhaustive here because different situations have their peculiar practical problems. An important common problem is the intergenerational environmental change. In selfing populations, homozygous parental genotypes can be cloned by further selfing, so that parents and outcrossed progeny can be assayed side by side with a randomized design in a single environment. In outcrossing populations where cloning of genotypes is possible, such as in cyclical parthenogens (Deng 1995), parents and selfed progeny can also be assayed in the same environment. Cloning of genotypes can essentially eliminate the problem of intergenerational environmental change (Deng 1995, 1997). In outcrossing populations where cloning of genotypes is not possible, the problem can be minimized by making the assay environments of the parent and progeny as similar as possible; additionally, large controls can be raised so that the values of the parents and offspring can be adjusted by the controls to be comparable.

Estimating $\bar{h}$ and σ_h^2 is important. Few estimates are available. The present study may have opened a door to estimating $\bar{h}$ and σ_h^2 in a relatively inexpensive fashion and to correctly interpreting the estimates. Estimates of the variability of selection coefficient s_i across loci are few and usually inferred by the difficult mutation accumulation experiments by assuming the values of other genomic mutation parameters such as genomic mutation rate (Keightley 1994). However, if s_i and h_i are correlated in some way, estimates of σ_h^2 by the simple methods presented here may also shed some light on the variability of s_i across loci. Theoretical investigation is needed, as are data.

We thank Drs. D. Charlesworth, D. Houle, M. Lynch, M. Uyenoyama and two anonymous reviewers for very helpful comments on the manuscript. H.-W. Deng thanks Dr. M. Lynch for years of advice and Dr. D. Hedgecock for providing support to attend the conference "The Genetic and Physiological Bases of Heterosis," which greatly benefited the development of this work. The work was partially supported by a FIRST AWARD from the National Institutes of Health to Y.-X. Fu. H.-W. Deng was also supported by a Health Future Foundation grant to Dr. R. Recker when preparing this article.

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.

Cavalli-Sforza, L. L., and W. F. Bodmer, 1971 *The Genetics of Human Populations.* W. H. Freeman, New York.

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

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.

Comstock, R. E., and H. F. Robinson, 1949 The components of genetic variance in populations of biparental progenies and their use in estimating the average degree of dominance. Biometrics **4:** 254–266.

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, NY.

Crow, J. F., 1993 Mutation, mean fitness, and genetic load, in *Oxford Surveys in Evolutionary Biology*, Vol. 9. Oxford University Press, Oxford, England.

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.

Crow, J., and M. Kimura, 1970 *An Introduction to Population Genetics Theory*. Harper & Row, New York.

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

Deng, H.-W., 1995 Sexual reproduction in Daphnia: its control and genetic consequences. Ph. D. Thesis. University of Oregon, Eugene, OR.

Deng, H.-W., 1997 Increase of developmental instability upon inbreeding in *Daphnia*. Heredity **78:** 182–189.

Deng, H.-W., and M. Lynch, 1996a Estimation of the genomic mutation parameters in natural populations. Genetics **144:** 349–360.

Deng, H.-W., and M. Lynch, 1996b Change of genetic architecture in response to sex. Genetics **143:** 203–212.

Deng, H.-W., and M. Lynch, 1997 Inbreeding depression and inferred deleterious-mutation parameters in Daphnia. Genetics **147:** 147–155.

Deng, H.-W., and Y.-X. Fu, 1997 On the three methods for estimating deleterious genomic mutation parameters. Genet. Res. (in press).

Eanes, W. F., J. Hey and D. Houle, 1985 Homozygous and hemizygous variability variation on the X chromosome of *Drosophila melanogaster*. Genetics **111:** 831–844.

East, E. M., 1908 Inbreeding in corn. Rept. Conn. Agrc. Exp. Stn. (1907): 419–428.

Falconer, D. S., 1989 *Introduction to Quantitative Genetics*. Longman, New York.

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

Gregory, W. C., 1965 Mutation frequency, magnitude of change

and the probability of improvement in adaptation. Radiat. Bot. **5** (Suppl.)**:** 429–441.
Hayman, B. I., 1954 The theory and analysis of diallele crosses. Genetics **39:** 789–809.
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.
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.
Kibota, T. T., and M. Lynch, 1996 Estimates of the genomic mutation rate deleterious to overall fitness in *E. coli*. Nature **381:** 694–696.
Kimura, M., and T. Ohta, 1971 *Theoretical Topics in Population Genetics*. Princeton University Press, Princeton, NJ.
Lande, R., and D. W. Schemske, 1985 The evolution of self-fertilization and inbreeding depression in plants. I. Genetic models. Evolution **39:** 24–40.
Lewontin, R. C., 1985 Population genetics. Annu. Rev. Genet. **19:** 81–102.
Lynch, M., and H.-W. Deng, 1994 Genetic slippage in response to sex. Am. Natur. **144:** 242–261.
Lynch, M., and B. Walsh, 1998 *Genetics and Data Analysis of Quantitative Traits*. Sinauer Associates, Sunderland, MA (in press).
Lynch, M., J. Conery and R. Burger, 1996 Mutation meltdowns in sexual populations. Evolution **49:** 1067–1080.
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.
Minvielle, F., 1987 Dominance is not necessary for heterosis: a two-locus model. Genet. Res. **49:** 245–247.
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., 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. 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.
Perrot, V., S. Richerd and M. Valero, 1991 Transition from haploidy to diploid. Nature **351:** 315–317.
Richey, F. D., 1942 Mock-dominance and hybrid vigor. Science **96:** 280–281.
Schemske, D. W., and R. Lande, 1985 The evolution of self-fertilization and inbreeding depression in plants. II. Empirical observations. Evolution **39:** 41–52.
Schnell, F. W., and C. C. Cockerham, 1992 Multiplicative *vs*. arbitrary gene action in heterosis. Genetics **131:** 461–469.
Shields, W. M., 1982 *Philotary, Inbreeding, and the Evolution of Sex*. SUNY Press, Albany, NY.
Shull, G. H., 1908 The composition of a field of maize. Rpt. 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.
Soule, M., 1986 *Conservation Biology*. Sinauer Associates, Sunderland, MA.
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.
Wright, S., 1977 *Evolution and the Genetics of Populations*. Vol. 3. *Experimental Results and Evolutionary Deductions*. The University of Chicago Press, Chicago.
Zapata, C., and G. Alverez, 1992 The detection of gametic disequilibria between allozyme loci in natural populations of *Drosophila*. Evolution **46:** 1900–1917.
Zapata, C., and G. Alverez, 1993 On the detection of nonrandom association between DNA polymorphisms in natural populations of *Drosophila*. Mol. Biol. Evol. **10:** 823–841.

Communicating editor: M. K. Uyenoyama