Estimation of Deleterious Genomic Mutation Parameters in Natural Populations by Accounting for Variable Mutation Effects Across Loci

Hong-Wen Deng,*,†,[1] Guimin Gao* and Jin-Long Li‡

**Osteoporosis Research Center and Department of Biological Sciences, Creighton University, Omaha, Nebraska 68131, †Laboratory of Molecular and Statistical Genetics, College of Life Sciences, Hunan Normal University, ChangSha, Hunan 410081, People's Republic of China and ‡Center for Medical Informatics, Yale University School of Medicine, New Haven, Connecticut 06520-8009*

Manuscript received April 26, 2002
Accepted for publication August 26, 2002

ABSTRACT

The genomes of all organisms are subject to continuous bombardment of deleterious genomic mutations (DGM). Our ability to accurately estimate various parameters of DGM has profound significance in population and evolutionary genetics. The Deng-Lynch method can estimate the parameters of DGM in natural selfing and outcrossing populations. This method assumes constant fitness effects of DGM and hence is biased under variable fitness effects of DGM. Here, we develop a statistical method to estimate DGM parameters by considering variable mutation effects across loci. Under variable mutation effects, the mean fitness and genetic variance for fitness of parental and progeny generations across selfing/outcrossing in outcrossing/selfing populations and the covariance between mean fitness of parents and that of their progeny are functions of DGM parameters: the genomic mutation rate U, average homozygous effect $\bar{s}$, average dominance coefficient $\bar{h}$, and covariance of selection and dominance coefficients cov(h, s). The DGM parameters can be estimated by the algorithms we developed herein, which may yield improved estimation of DGM parameters over the Deng-Lynch method as demonstrated by our simulation studies. Importantly, this method is the first one to characterize cov(h, s) for DGM.

THE genomes of all organisms are subject to deleterious genomic mutations (DGM) continuously. In spite of our increasing knowledge of the molecular underpinnings of mutations, little is known about the overall risk exerted on human health and on continuing survivability of other organisms (especially rare and endangered species) by DGM (Crow 1993a,b, 1995). To assess this overall risk correctly, we need to have a solid knowledge of the genomic mutation rate (U) at which DGM arise in the whole genome of an individual and the distribution of their effects, such as the mean selection coefficient ($\bar{s}$), the mean dominance coefficient ($\bar{h}$), and the covariance of dominance and selection coefficients of DGM [cov(h, s)]. Estimation of these parameters is also important for testing the validity of a number of evolutionary theories in genetics (Turelli and Orr 1995; and the references within Deng *et al.* 1998, 1999).

Despite the extreme importance of our knowledge of deleterious mutation parameters, few estimates are available (Simmons and Crow 1977; Crow and Simmons 1983; Kondrashov 1988; Crow 1993a,b, 1995; Bataillon 2000). Particularly, no method to estimate U is not biased by variable mutation effects, and no method to estimate cov(h, s) is important for our understanding of Haldane's rule by the dominance hypothesis (Turelli and Orr 1995). The current experimental approaches and the estimation methods of the parameters of DGM are summarized and compared (Deng and Fu 1998; Deng *et al.* 1999; Deng and Li 2001). It is concluded that under their respective assumptions of various approaches, estimation by the Deng-Lynch method (Deng and Lynch 1996, 1997) in natural populations generally results in the best statistical quality in terms of bias and sampling variance (Deng and Fu 1998). In addition, it has been shown that violation of various assumptions [including the mutation-selection (M-S) balance assumption] underlying the Deng-Lynch method does not seriously undermine its estimation robustness (Li *et al.* 1999; Li and Deng 2000; Deng and Li 2001).

As with almost all the other estimation methods (except a maximum-likelihood estimation method for mutation-accumulation experiments; Keightley 1994), the Deng-Lynch method that applies to natural outcrossing or selfing populations assumes constant fitness effects of DGM. This assumption is well recognized as biologically implausible. Although the estimation bias introduced by variable mutation effects in the Deng-Lynch estimation method by assuming constant mutation effects is not substantial (Deng *et al.* 1999), an estimation method that considers variable mutation effects may reduce estimation bias (although not necessarily always so). Most importantly, the parameters [*e.g.*, cov(h, s)] char-

[1]*Corresponding author:* Osteoporosis Research Center, Creighton University, 601 N. 30th St., STE. 6787, Omaha, NE 68131. E-mail: deng@creighton.edu

acterizing variable effects of DGM can be estimated only in statistical methods that consider variable mutation effects.

In this article, we present a method for estimating DGM parameters accounting for variable effects across loci in natural outcrossing or selfing populations at M-S balance. We investigate the statistical properties (bias and sampling variance) of this new method, using computer simulations in comparison with the Deng-Lynch method (Deng and Lynch 1996, 1997) that assumed constant mutation effects across loci.

THEORY

The assumptions are the same as those of the Morton-Charlesworth method (Morton *et al.* 1956; Charlesworth *et al.* 1990) and the Deng-Lynch method (Deng and Lynch 1996, 1997; Deng 1998b). Namely, the population is assumed to be large, randomly mating, highly selfing or outcrossing, at linkage equilibrium, and at M-S balance. In addition, the fitness function is assumed to be multiplicative, which is biologically plausible (Morton *et al.* 1956; Crow 1986; Craddock *et al.* 1995; Fu and Ritland 1996). Mutations at each locus are assumed to have constant effect s and h.

In this study, we consider variable mutation effects in the development of an estimation method for DGM parameters in natural populations. Under variable mutation effects across loci, homozygous effect s for mutations is a random variable between 0 and 1. We assume that, for a mutation, dominance coefficients h and s are functionally related so that $h = h(s)$. This assumption is supported by the limited data and theory (Simmons and Crow 1977; Kacser and Burns 1981; Crow and Simmons 1983). We divide the domain of s, [0, 1], for new mutations into T intervals with each having a width of $1/T$. Let $I_k = [k/T, (k + 1)/T]$ denote the kth interval, and define the probability

$$p_k = P(s \in I_k), \quad k = 0, 1, \ldots, T - 1.$$

When T is sufficiently large, s and h are approximately constant within each interval but are variable across various intervals. Let U_k denote the mutation rate corresponding to mutations with an effect s falling into the interval I_k, and then $U_k = Up_k$.

With the assumptions we have, in outcrossing populations, the number of mutant alleles with mutation effects s falling into an interval I_k within an individual (all in the heterozygous state; Morton *et al.* 1956; Deng and Lynch 1996) follows a Poisson distribution with an expectation

$$\bar{n}_k = U_k/h_k s_k = Up_k/h_k s_k \tag{1a}$$

(Deng and Lynch 1996, 1997). In selfing populations, the number of loci homozygous for mutant alleles with an effect s falling into an interval I_k within an individual follows a Poisson distribution with an expectation

$$\bar{n}_k = U_k/2s_k = Up_k/2s_k. \tag{1b}$$

Outcrossing populations: We illustrate our experimental design and estimation method by using populations capable of selfing. The method may be extended to outcrossing populations where selfing is not feasible as in the Deng-Lynch method (Deng 1998b). The basic data structure is outcrossed parents and multiple selfed progeny from each parent (forming selfed families). Let $\bar{W}_o$ and $\bar{W}_s$ be the mean fitness in the parental and offspring generations, respectively, σ_o^2 the genetic variance of fitness in the parental generation, σ_t^2 the total genetic variance of fitness in the selfed progeny generation, σ_s^2 the genetic variance of the mean fitness of selfed progeny in selfing families, and $\mathrm{cov}(w_p, w_s)$ the covariance between the fitness of a parent (w_p) and the mean fitness of its selfed progeny (w_s). Under the above assumption for mutation effects that are variable across various intervals at different loci, as in Deng and Lynch (1996), it can be shown that the fitness moments are related to the DGM parameters as

$$\bar{W}_o = W_{max} \exp(-U) \tag{2}$$

$$\sigma_o^2 = \bar{W}_o^2[\exp(U\overline{hs}) - 1], \tag{3}$$

$$\bar{W}_s = W_{max} \exp\{-(U/4)[2 + (1/\tilde{h})]\}, \tag{4}$$

$$\sigma_s^2 = \bar{W}_s^2 \{\exp[(U/4)(\bar{s} + \overline{hs} + \overline{s/(4h)})] - 1\}, \tag{5}$$

$$\sigma_t^2 = \bar{W}_s^2[\exp[U(\overline{hs}/2 + \overline{s/(4h)})] - 1], \tag{6}$$

$$\mathrm{cov}(w_p, w_s) = \bar{W}_s\bar{W}_o \{\exp[(1/4)U(2\overline{hs} + \bar{s})] - 1\}, \tag{7}$$

where the parameters with overbars denote arithmetic mean properties of new DGM parameters, $\tilde{h}$ is the harmonic mean dominance coefficient of new mutations, and W_{max} is the expected fitness of a mutation-free genotype in an environment where fitness measurements are taken. W_{max} serves as a scaling factor so that the fitness measurement can be on any scale instead of just from 0.0 to 1.0 and also so that mean environmental effects of experiments do not influence estimation (Deng and Lynch 1996).

Among Equations 2–7, there are only five independent equations containing six unknown parameters. By assuming one of the six parameters known in the estimation, estimators of the other parameters can be derived. This is the strategy employed in the likelihood characterization of DGM parameters when variable mutation effects are considered in estimation (Keightley 1994; Deng *et al.* 1999; Deng and Li 2001). Here we assume that U is known in the estimation for the time being. Alternatively, an initial value of U may be estimated from other approaches (Deng *et al.* 1999) or may be estimated by the current experimental design and data with the Deng-Lynch method (Deng and Lynch 1996; see below). (If we assume that one of the parameters $\tilde{h}$, $\bar{s}$, and $\overline{hs}$ is known, similar estimation procedures can be derived for U and the rest of the other parameters. $\tilde{h}$ can

be estimated by methods such as that of Deng 1998a.) Solving these equations jointly yields estimators of $\tilde{h}$, $\overline{hs}$, and $\bar{s}$ and as

$$\tilde{h} = \frac{1}{2-4(y/U)}, \quad \bar{s} = \frac{4b - 2x}{U}, \quad \overline{hs} = \frac{x}{U}, \tag{8}$$

where

$$x = \ln\left(\frac{\sigma_o^2}{\overline{W}_o^2} + 1\right), \quad y = \ln\left(\frac{\overline{W}_s}{\overline{W}_o}\right),$$
$$z = \ln\left(\frac{\sigma_s^2}{\overline{W}_s^2} + 1\right), \quad b = \ln\left(\frac{\mathrm{cov}(w_p, w_s)}{\overline{W}_o \overline{W}_s} + 1\right). \tag{9}$$

From these estimates, other composite parameters of DGM, such as the mean number of mutations per genome $\overline{n}$, mutational variance V_m per generation, and mean mutation effects on fitness $U\overline{hs}$, can be derived (Deng and Lynch 1996). The covariance of h and s for mutations cov(h, s) can be approximated, or at least an upper bound can be estimated, as

$$\mathrm{cov}(h, s) = \overline{hs} - \bar{h}\bar{s} \le \overline{hs} - \tilde{h}\bar{s}. \tag{10}$$

This is because for any distribution, $\bar{h} \ge \tilde{h}$. Let $\mathbf{cov(\boldsymbol{h}, \boldsymbol{s})} = \overline{hs} - \bar{s}\tilde{h}$, where **cov(*h*, *s*)** denotes an upper bound of cov(h, s). This offers us *the first opportunity* to quantify the magnitude and the sign of cov(h, s). It would be impossible to come up with analytical estimators for DGM parameters such as cov(h, s) if in the analytical derivation, variable mutation effects are not considered. This is simply because these parameters such as cov(h, s) would be zero and meaningless in an analytical estimation developed under constant mutation effects.

Selfing populations: Random pairs of highly selfing and homozygous parental genotypes (denoted as P generation) are crossed to obtain outcrossed progeny (denoted as F_1 generation). Let $\overline{W}_p$ and σ_p^2 be the mean fitness and genetic variance of fitness in the P generation, respectively, $\overline{W}_{F_1}$ and $\sigma^2_{F_1}$ be the mean fitness and genetic variance of fitness in the F_1 generation, respectively, and cov($\overline{P}$, F_1) be the covariance between the mean fitness of the two parents and the fitness of their F_1 progeny. Under variable mutation effects across loci, the fitness moments are related to the DGM parameters as follows:

$$\overline{W}_p = W_{max} \exp(-U/2), \tag{11}$$
$$\sigma_p^2 = \overline{W}_p^2[\exp(U\bar{s}/2) - 1], \tag{12}$$
$$\overline{W}_{F_1} = W_{max} \exp(-U\bar{h}), \tag{13}$$
$$\sigma_{F_1}^2 = \overline{W}_{F_1}^2(\exp(U\overline{h^2 s}) - 1), \tag{14}$$
$$\mathrm{cov}(\overline{P}, F_1) = \overline{W}_p \overline{W}_{F_1}[\exp(U\overline{hs}/2) - 1]. \tag{15}$$

It should be noted that the derivation for Equations 2–7 and 11–15 assumes mutation effects that are variable. The strategy is to divide the range of variable selection coefficient s (from zero to one) into infinitely small intervals so that s can be treated as constant within each of the intervals but varying across intervals in our analytical derivation. Again, there are six unknowns (U, $\bar{h}$, $\overline{hs}$, $\bar{s}$, $\overline{h^2 s}$, and W_{max}) in the above five equations. By assuming or estimating one of the six parameters, estimators of the other five parameters can be derived. Here, as earlier for outcrossing populations, we assume that U is known in the estimation for illustration. Alternatively, an initial value of U may be estimated from other approaches (Deng *et al.* 1999) or may be estimated with the Deng-Lynch method from the same data and experimental design as the current estimation method (Deng and Lynch 1996). Solving these equations jointly yields estimators of $\bar{h}$, $\overline{hs}$, and $\bar{s}$,

$$\bar{h} = 0.5 - (y/U), \quad \bar{s} = \frac{2x}{U}, \quad \overline{hs} = \frac{2b}{U}, \tag{16}$$

where

$$x = \ln\left(\frac{\sigma_p^2}{\overline{W}_p^2} + 1\right), \quad y = \ln\left(\frac{\overline{W}_{F_1}}{\overline{W}_p}\right),$$
$$z = \ln\left(\frac{\sigma_{F_1}^2}{\overline{W}_{F_1}^2} + 1\right), \quad b = \ln\left(\frac{\mathrm{cov}(\overline{P}, F_1)}{\overline{W}_p\ \overline{W}_{F_1}} + 1\right). \tag{17}$$

In selfing populations, we can use Equation 10 to estimate cov(h, s) by the above estimates of $\bar{h}$, $\overline{hs}$, and $\bar{s}$, which are unbiased under variable mutation effects with a known correct U. The estimators for $\bar{h}$ and $\bar{s}$ when assuming U is known are the same as those in Deng and Lynch (1996) for selfing populations.

The above estimation developed herein does not assume any specific functional relationship between s and h and any specific distribution form for the selection coefficient s. Therefore, the estimates are robust to different unknown forms of the distribution of s and the functional relationship between s and h. This is true despite that we assume specific distributions of s and a functional relationship between s and h in the following simulation studies to investigate the statistical properties of our estimation.

SIMULATIONS AND RESULTS

As with Keightley (1994), we assume that s for mutations follows a gamma distribution, with a density function

$$g(s) = \alpha^\beta s^{\beta-1} e^{-\alpha s}/\Gamma(\beta),$$

where $\Gamma(\beta) = \int_0^\infty y^{\beta-1} e^{-y}\, dy$. α and β are the scale and shape parameters, respectively. $\bar{s} = \beta/\alpha$ and $\sigma_s^2 = \beta/\alpha^2$. As in Deng and Lynch (1996), we let $h = h(s) = e^{-As}/2$, where $A = 13$, which is in rough accordance with the few available data (Gregory 1965; Mackay *et al.* 1992; Deng and Lynch 1996; Deng and Fu 1998). With these assumptions, the parameters $\bar{h}$, $\tilde{h}$, $\overline{hs}$, and cov(h, s) can be derived as

$$\bar{h} = \alpha^{\beta}/[2(A + \alpha)^{\beta}],$$

$$\tilde{h} = \begin{cases} (\alpha - A)^{\beta}/[2\alpha^{\beta}] & \text{when } \alpha - A > 0 \\ 0 & \text{when } \alpha - A \le 0, \end{cases}$$

$$\overline{hs} = \beta\alpha^{\beta}/[2(A + \alpha)^{\beta+1}],$$

$$\text{cov}(h, s) = -A\beta\alpha^{\beta-1}/[2(A + \alpha)^{\beta+1}].$$

These DGM parameters can be used for comparison to examine the estimated values with our estimation methods in simulations.

The simulation procedures are the same as those that have been documented extensively earlier (Deng and Lynch 1996; Deng 1998b) and are thus not elaborated here. In simulations, we assume that the fitnesses of various genotypes can be measured with little error, which is justifiable in the investigation of estimation bias and comparison of various estimation methods (Deng *et al.* 1999). Under the assumptions for the analytical development of our estimation methods, the number of mutant alleles corresponding to an interval I_k per individual follows the Poisson distributions (Equations 1a and 1b) with p_k being determined as

$$p_k = P(s \in I_k) = \frac{1}{\Gamma(\beta)} \int_{I_k} \alpha^{\beta} s^{\beta-1} e^{-\alpha s} ds, \quad k = 0, 1, \ldots, T - 1.$$

It can be shown that

$$p_k = -\exp\left(-\alpha\frac{k+1}{T}\right) + \exp\left(-\alpha\frac{k}{T}\right), \quad \text{when } \beta = 1;$$

$$p_k = -\left[\exp\left(-\alpha\frac{k+1}{T}\right)\right]\left(\alpha\frac{k+1}{T} + 1\right) + \left[\exp\left(-\alpha\frac{k}{T}\right)\right]\left(\alpha\frac{k}{T} + 1\right),$$

when $\beta = 2$

and when $\beta = 0.5$,

$$p_k = Erf\left(\sqrt{\alpha\frac{k+1}{T}}\right) - Erf\left(\sqrt{\alpha\frac{k}{T}}\right),$$

where $Erf(x) = (2/\sqrt{\pi})\int_0^x e^{-t^2}\,dt$ $(x > 0)$. $Erf(x)$ can be approximated as

$$Erf(x) \approx 1 - (1 + \sum_{i=1}^{6} a_i x^i)^{-16} \qquad (18)$$

(Gao 1995), where $a_1 = 0.0705230784$, $a_2 = 0.0422820123$, $a_3 = 0.0092705272$, $a_4 = 0.0001520143$, $a_5 = 0.0002765672$, $a_6 = 0.0000430638$.

To evaluate the performance of our estimation in outcrossing populations in simulations, for each set of parameters U, α, and β, K parents were sampled from the parental generation, and from each of these, M selfed progeny were produced. The fitness of an individual from the parental generation is

$$W_o = W_{\max}\prod_{k=1}^{T}(1 - h_k s_k)^{n_k},$$

where n_k is the number of mutation-bearing loci with their effects falling into the interval I_k in an individual, obtained by random sampling from the Poisson distribution defined above. The fitness of each selfed offspring was obtained by allowing the n_k heterozygous loci of a parent to segregate randomly into the *AA*, *Aa*, and *aa* genotypes with respective probabilities of 1/4, 1/2, and 1/4. Letting n_{1k} and n_{2k} $(k = 1, \ldots, T)$ be the numbers of heterozygous and homozygous loci containing mutations with effects falling into the interval I_k in a selfed offspring, the fitness of the selfed progeny is

$$W_s = W_{\max}\prod_{k=1}^{T}(1 - h_k s_k)^{n_{1k}}(1 - s_k)^{n_{2k}}.$$

Unless otherwise specified, for each set of parameters (U, α, β, K, M), we performed 1000 simulations. We let $W_{\max} = 1$ throughout, as the value of $W_{\max}$ does not influence DGM parameter estimation.

For selfing populations, the fitness of an individual from the parental generation is

$$W_p = W_{\max}\prod_{k=1}^{T}(1 - s_k)^{n_k},$$

where n_k is the number of mutation-bearing loci with mutation effects falling into the interval I_k in an individual, and it is obtained by random sampling from the Poisson distribution defined earlier. Each parent mates with another random parent (not in the original set of K) to produce a total of K progeny (one per family) with fitness

$$W_{F_1} = W_{\max}\prod_{k=1}^{T}(1 - h_k s_k)^{n_{1k}+n_{2k}},$$

where n_{1k} and n_{2k} $(k = 1, \ldots, T)$ are the numbers of homozygous mutant loci in interval I_k in the two parents, respectively.

In the estimation Equations 8 or 16, U must be known, assumed, or estimated with other approaches first. In simulations, we experimented and examined two methods to estimate U: (1) by the Deng-Lynch method (Deng and Lynch 1996) and (2) by an empirical regression procedure introduced here. We simulated parents and their children according to variable effects for each set of given parameter values of U, α, and β, and obtained the estimates $\hat{U}_1$, $\hat{s}_1$, and $\hat{h}_1$ by the Deng-Lynch method (Deng and Lynch 1996). (A circumflex indicates an estimated value throughout.) We found a strong linear relationship between the parameter values of U and the estimates $\hat{U}_1$ and $\hat{s}_1$ under any fixed β. Through a series of simulations, we obtained samples under various parameter values of U, α, and fixed β-values, and we obtained estimates $\hat{U}_1$ and $\hat{s}_1$ with the Deng-Lynch method under various fixed β-values. Then we fit a multiple regression model under each specific β-value,

$$\hat{U} = \hat{a}_1 + \hat{b}_1\hat{U}_1 + \hat{c}_1\hat{s}_1, \qquad (19)$$

where $\hat{U}$ estimates U with little bias when β is correctly assumed as shown by our simulation results not presented

here. The empirical estimation is useful only when the shape parameter β can be estimated using other methods and experimental data (*e.g.*, Keightley 1994).

The simulation results are represented by the data in Tables 1–4. The ranges of the values for the parameters (such as U, $\bar{h}$, and $\bar{s}$) generally cover those reported earlier from classical empirical experiments (*e.g.*, Mukai *et al.* 1972; Lynch *et al.* 1999). Three general conclusions emerge from our simulation studies under variable mutation effects. First, when U is set to equal true values or when the estimates of U are obtained via Equation 19 by assuming a correct β-value, application of Equation 8 or 16 to both obligate selfing or outcrossing populations yields nearly unbiased estimates for the DGM parameters with small standard deviation. The estimates of U by Equation 8 have smaller mean square error despite larger standard deviation when U is set equal to the estimates obtained by regression Equation 19 than those obtained by the Deng-Lynch method. The larger standard deviation may be partly due to the fact that Equation 19 is established by empirical regression procedures that involve an additional level of sampling error for the final estimation. The estimates of $\bar{s}$ by Equation 8 in outcrossing populations have smaller sampling variance and smaller bias than those obtained directly by the Deng-Lynch method, *e.g.*, by comparison of the estimates in rows 1 and 3 for each parameter set in Table 1. This is true even when no prior assumption is made about the magnitude of U, when U is first estimated directly with the Deng-Lynch method, and then the estimate of U is used in the current estimation method, (Equation 8) for the other DGM parameters. The estimates of $\tilde{h}$ by Equation 8 have smaller or comparable sampling variance than those obtained directly by the Deng-Lynch method for $\bar{h}$ (for each parameter set, compare the estimates of the second to fourth rows with that of the first row in Table 1). The comparison of the estimation quality between the current estimation method and the Deng-Lynch method changes little with the parameter values (Table 1). When $\beta = 0.5$, the bias of the estimates of the parameters is larger than that when $\beta = 1$ and 2. This may be due to the approximation formula 18 used to compute $p_k = P(s_i \in I_k)$ when $\beta = 0.5$, while the computation of $p_k = P(s_i \in I_k)$ when $\beta = 1$ and 2 is exact.

Second, when U is set equal to the estimates ($\hat{U}_1$) that were obtained by the Deng-Lynch method (Deng and Lynch 1996) and that are downwardly biased, the estimates of the other DGM parameters by Equations 8 and 16 are biased with small sampling variance (Tables 1 and 2). For outcrossing populations, the estimation Equation 8 yields less biased estimates with smaller standard deviation for $\bar{s}$ than for the Deng-Lynch method (Table 1), and the estimates of $\bar{s}$, $\overline{hs}$, **cov(*h, s*)** are upwardly biased and estimates of $\tilde{h}$ are downwardly biased. The result can be understood from Equation 8, since $\hat{U}_1$ is downwardly biased as estimated by the Deng-Lynch method. In selfing populations, Equation 16 yields the same estimates for $\bar{s}$ and $\bar{h}$ as those obtained by the Deng-Lynch method (Table 2), which is expected as pointed out earlier. The estimates of $\bar{s}$, $\overline{hs}$, and cov(h, s) are upwardly biased and estimates of $\bar{h}$ are downwardly biased because $\hat{U}_1$ is downwardly biased, which can be understood from Equation 16.

Third, in outcrossing populations, the **cov(*h, s*)** is correctly estimated to be an upper bound of cov(h, s); however, the sign of **cov(*h, s*)** can sometimes be estimated to be different from that of cov(h, s). In selfing populations, cov(h, s) can always be estimated with correct sign and small estimation bias.

ROBUSTNESS ANALYSIS

In the estimation of the DGM parameters, we need a prior estimate of one of the six parameters (such as U as investigated here) based on some external knowledge obtained from other estimation approaches. The estimation bias of this parameter or the bias of an assumed value will cause estimation bias of the other parameters. Hence, we investigate the sensitivity of estimators to the departures of U from true value, using computer simulations (Figures 1 and 2). We define a relative bias rate (RBR), (estimate − true value)/(true value), to measure the sensitivity of estimators to an incorrectly assumed or estimated U value. In examining the robustness of the estimator for **cov(*h, s*)**, the true value used is the parameter value of **cov(*h, s*)** as defined after Equation 10 and not cov(h, s).

In simulations for the investigation of the robustness of our current estimation of the other DGM parameters, U is set equal to a given value (denoted as U_{given}), which ranges from $0.5U_0$ to $1.5U_0$ (U_0 is the true value of U). This range of the estimate of U investigated is reasonable given the magnitude of bias that is normally found with the method such as that of Deng and Lynch (1996). The changes in the mean relative bias rates (MRBR) of the estimates of the parameter values in 1000 simulations are shown in Figures 1 and 2. It can be seen that when U_{given} ranged from $0.7U_0$ to $1.5U_0$ (which means that the departure of U_{given} from U_0 ranged from $-0.3U_0$ to $0.5U_0$), the MRBR of the estimates of the parameter values changed smoothly and changed little in both outcrossing and selfing populations. When U_{given} ranged from $0.9U_0$ to $1.2U_0$, the absolute values of the MRBR of the estimates of parameters [except **cov(*h, s*)** for outcrossing populations when $\alpha = 20$] are <0.185 in both outcrossing and selfing populations. For outcrossing populations, when $\alpha = 20$, if $U_{\text{given}} \leq 0.9U_0$ or $U_{\text{given}} \geq 1.1U_0$, the absolute values of the MRBR of **cov(*h, s*)** are >1.0 (Figure 1, b and d). (Note the scale difference of the y-axis in Figure 1, b and d, with the other plots in Figures 1 and 2.) Thus, even when U is estimated with some bias, if the magnitude is similar to that obtained by methods such as that of Deng and Lynch

TABLE 1

Parameter estimates under variable mutational effects in outcrossing populations ($\beta = 1.0$)

U	$\bar{s}$	$\bar{h}$	α	$\overline{hs}$	$\tilde{h}$	$\mathrm{cov}(h, s)$	$\hat{U}$	$\hat{\bar{s}}$	$\hat{\tilde{h}}$	$\hat{\bar{h}}$	$\hat{\overline{hs}}$	$\mathbf{cov}(\hat{h}, s)$
1.50	0.05	0.30	20	0.0092	0.175	−0.0060	0.803 ± 0.071	0.1550 ± 0.0211		0.112 ± 0.007		
								0.0498 ± 0.0076	0.175 ± 0.002		0.0092 ± 0.0009	0.0004 ± 0.0010
								0.0940 ± 0.0180	0.112 ± 0.007		0.0172 ± 0.0016	0.0067 ± 0.0007
							1.495 ± 0.142	0.0506 ± 0.0101	0.174 ± 0.011		0.0092 ± 0.0010	0.0005 ± 0.0007
	0.01	0.44	100	0.0039	0.435	−0.0005	0.754 ± 0.067	0.0209 ± 0.0029		0.383 ± 0.008		
								0.0102 ± 0.0010	0.434 ± 0.006		0.0040 ± 0.0004	−0.0004 ± 0.0001
								0.0205 ± 0.0028	0.383 ± 0.008		0.0080 ± 0.0010	0.0002 ± 0.0000
							1.461 ± 0.129	0.0105 ± 0.0015	0.432 ± 0.005		0.0041 ± 0.0005	−0.0004 ± 0.0001
0.50	0.05	0.30	20	0.0092	0.175	−0.0060	0.266 ± 0.022	0.1560 ± 0.0188		0.111 ± 0.007		
								0.0501 ± 0.0068	0.175 ± 0.003		0.0092 ± 0.0009	0.0004 ± 0.0009
								0.0949 ± 0.0157	0.111 ± 0.007		0.0173 ± 0.0015	0.0067 ± 0.0007
							0.477 ± 0.048	0.0533 ± 0.0101	0.169 ± 0.011		0.0096 ± 0.0010	0.0007 ± 0.0006
	0.01	0.44	100	0.0039	0.435	−0.0005	0.252 ± 0.023	0.0209 ± 0.0029		0.383 ± 0.008		
								0.0102 ± 0.0001	0.434 ± 0.001		0.0040 ± 0.0004	−0.0004 ± 0.0001
								0.0204 ± 0.0028	0.383 ± 0.008		0.0080 ± 0.0010	0.0002 ± 0.0000
							0.509 ± 0.045	0.0101 ± 0.0014	0.435 ± 0.005		0.0040 ± 0.0005	−0.0004 ± 0.0001

For each set of parameters given in the first seven columns, 1000 simulations were performed. Each simulation is based on 200 parents and dividing the range of s_k (0–1) into 200 equal intervals each with a width of 0.005. For outcrossing populations, 40 selfed progeny were evaluated per parent. Reported values for $\hat{U}$, $\hat{\bar{s}}$, $\hat{\tilde{h}}$, $\hat{\bar{h}}$, $\hat{\overline{hs}}$, and $\mathbf{cov}(\hat{h}, s)$ are the mean ± SD. For each set of parameters, we give four sets of estimates: The estimates in the first row are obtained by the Deng-Lynch method (Deng and Lynch 1996). The estimates in the second, third, and fourth rows are obtained by Equation 8; U is set equal to the true value in the second row, U is estimated by the Deng-Lynch method in the third row, and U is estimated by an empirical equation Equation 19 in the fourth row ($\hat{U} = 0.03926 + 1.898\ \hat{U}_1 - 0.436\ \hat{s}_1$; $R = 0.998$, $P < 0.001$).

TABLE 2

Parameter estimates under variable mutational effects in selfing populations ($\beta = 1.0$)

U	$\bar{s}$	$\bar{h}$	α	$\overline{hs}$	cov(h, s)	$\hat{U}$	$\hat{\bar{s}}$	$\hat{\bar{h}}$	$\hat{\overline{hs}}$	**cov($\hat{h}$, s)**
1.50	0.05	0.30	20	0.0092	−0.0060	1.051 ± 0.056	0.0713 ± 0.0086	0.218 ± 0.012		
							0.0498 ± 0.0049	0.303 ± 0.008	0.0092 ± 0.0010	−0.0059 ± 0.0013
							0.0713 ± 0.0086	0.218 ± 0.012	0.0131 ± 0.0013	−0.0024 ± 0.0009
						1.550 ± 0.107	0.0486 ± 0.0069	0.309 ± 0.012	0.0089 ± 0.0010	−0.0060 ± 0.0011
	0.01	0.44	100	0.0039	−0.0005	0.914 ± 0.242	0.0178 ± 0.0050	0.398 ± 0.021		
							0.0102 ± 0.0010	0.441 ± 0.003	0.0040 ± 0.0004	−0.0005 ± 0.0004
							0.0178 ± 0.0046	0.398 ± 0.021	0.0068 ± 0.0013	−0.0002 ± 0.0004
						1.516 ± 0.389	0.0107 ± 0.0028	0.439 ± 0.012	0.0041 ± 0.0008	−0.0005 ± 0.0004
0.50	0.05	0.30	20	0.0092	−0.0060	0.349 ± 0.023	0.0722 ± 0.0090	0.218 ± 0.014		
							0.0503 ± 0.0059	0.303 ± 0.014	0.0092 ± 0.0010	−0.0060 ± 0.0014
							0.0722 ± 0.0090	0.218 ± 0.014	0.0132 ± 0.0014	−0.0024 ± 0.0009
						0.458 ± 0.056	0.0560 ± 0.0114	0.283 ± 0.025	0.0102 ± 0.0016	−0.0053 ± 0.0009
	0.01	0.44	100	0.0039	−0.0005	0.300 ± 0.071	0.0179 ± 0.0046	0.398 ± 0.020		
							0.0102 ± 0.0010	0.441 ± 0.005	0.0040 ± 0.0004	−0.0005 ± 0.0004
							0.0179 ± 0.0046	0.398 ± 0.025	0.0069 ± 0.0012	−0.0002 ± 0.0004
						0.563 ± 0.123	0.0095 ± 0.0024	0.446 ± 0.010	0.0037 ± 0.0007	−0.0006 ± 0.0004

For each set of parameters given in the first six columns, 1000 simulations were performed. Each simulation is based on 200 parents and dividing the range of s_i (0–1) into 200 equal intervals. For selfing populations, 200 random outcrossed progeny were evaluated per simulation (one for each of the original 200 parents, randomly outcrossed). Values reported for $\hat{U}$, $\hat{s}$, $\hat{h}$, $\hat{\overline{hs}}$, and **cov($\hat{h}$, s)** are the mean ± SD. For each set of parameters, we give four sets of estimates: The estimates in the first row are obtained by the Deng-Lynch method (Deng and Lynch 1996). The estimates in the second, third, and fourth rows are obtained by Equation 16; U is given equal to the true value in the second row, U is estimated by the Deng-Lynch method in the third row, and U is estimated by an empirical regression equation (Equation 19) in the fourth row ($\hat{U} = 0.157 + 1.551 \times \hat{U}_1 - 3.332 \times \hat{s}_1$; $R = 0.998$, $P < 0.001$).

TABLE 3

Parameter estimates under variable mutational effects in outcrossing populations

β	U	$\bar{s}$	$\bar{h}$	α	$\overline{hs}$	$\tilde{h}$	cov(h, s)	$\hat{U}$	$\hat{\bar{s}}$	$\hat{\tilde{h}}$	$\hat{\bar{h}}$	$\hat{\overline{hs}}$	cov($\hat{h}$, s)
0.5	1.50	0.010	0.45	50	0.0035	0.43	−0.0009	0.571 ± 0.055	0.0294 ± 0.0037		0.345 ± 0.010		
									0.0106 ± 0.0011	0.427 ± 0.009		0.0038 ± 0.0004	−0.0007 ± 0.0001
								1.570 ± 0.137	0.0102 ± 0.0014	0.430 ± 0.005		0.0037 ± 0.0004	−0.0007 ± 0.0002
	0.50	0.010	0.45	50	0.0035	0.43	−0.0009	0.185 ± 0.018	0.0300 ± 0.0045		0.342 ± 0.011		
									0.0105 ± 0.0011	0.427 ± 0.001		0.0037 ± 0.0004	−0.0007 ± 0.0002
								0.588 ± 0.047	0.0090 ± 0.0012	0.436 ± 0.005		0.0032 ± 0.0004	−0.0007 ± 0.0001
2.0	1.50	0.01	0.44	200	0.0041	0.44	−0.0003	0.926 ± 0.095	0.0168 ± 0.0024		0.404 ± 0.008		
									0.0100 ± 0.0012	0.437 ± 0.005		0.0041 ± 0.0005	−0.0003 ± 0.0001
								1.458 ± 0.145	0.0104 ± 0.0016	0.435 ± 0.005		0.0043 ± 0.0006	−0.0002 ± 0.0001
	0.50	0.01	0.44	200	0.0041	0.44	−0.0003	0.316 ± 0.034	0.0166 ± 0.0023		0.406 ± 0.008		
									0.0102 ± 0.0010	0.436 ± 0.001		0.0042 ± 0.0004	−0.0002 ± 0.0001
								0.524 ± 0.048	0.0096 ± 0.0014	0.439 ± 0.005		0.0040 ± 0.0005	−0.0003 ± 0.0001

The simulation conditions are as described in the note to Table 1, except that 200 simulations were performed for each set of parameters. For each set of parameters we give three sets of estimates: The estimates in the first row are obtained by the Deng-Lynch method (Deng and Lynch 1996); the estimates in the second and third rows are obtained by Equation 8; U is given equal to the true values in the second row and the U is estimated by empirical regression equation (Equation 19) in the third row [when β = 0.5, $\hat{U} = 0.109 + 2.619 \times -0.404 \times \hat{s}_1$ ($R = 0.998$, $P < 0.001$); when β = 2.0, $\hat{U} = 0.03926 + 1.898 \times -0.436 \times \hat{s}_1$; $R = 0.998$, $P < 0.001$].

TABLE 4

Parameter estimates with variable mutational effects in selfing populations

β	U	$\bar{s}$	$\bar{h}$	α	$\overline{hs}$	cov(h, s)	$\hat{U}$	$\hat{\bar{s}}$	$\hat{\bar{h}}$	$\hat{\overline{hs}}$	cov($\hat{h}$, s)
0.5	1.50	0.010	0.45	50	0.0035	−0.0009	0.704 ± 0.123	0.0233 ± 0.0052	0.373 ± 0.019		
								0.0105 ± 0.0011	0.442 ± 0.003	0.0038 ± 0.0004	−0.0009 ± 0.0004
							1.458 ± 0.244	0.0113 ± 0.0024	0.439 ± 0.009	0.0040 ± 0.0006	−0.0009 ± 0.0005
	0.50	0.010	0.45	50	0.0035	−0.0009	0.234 ± 0.047	0.0234 ± 0.0053	0.373 ± 0.021		
								0.0106 ± 0.0011	0.442 ± 0.005	0.0038 ± 0.0004	−0.0009 ± 0.0004
							0.558 ± 0.097	0.0099 ± 0.0023	0.446 ± 0.009	0.0035 ± 0.0005	−0.0009 ± 0.0005
2.0	1.50	0.01	0.44	200	0.0041	−0.0003	1.135 ± 0.426	0.0145 ± 0.0043	0.415 ± 0.021		
								0.0099 ± 0.0010	0.441 ± 0.003	0.0041 ± 0.0004	−0.0003 ± 0.0004
							1.461 ± 0.716	0.0114 ± 0.0035	0.434 ± 0.016	0.0046 ± 0.0010	−0.0003 ± 0.0004
	0.50	0.01	0.44	200	0.0041	−0.0003	0.379 ± 0.126	0.0144 ± 0.0045	0.416 ± 0.021		
								0.0100 ± 0.0010	0.441 ± 0.005	0.0041 ± 0.0004	−0.0003 ± 0.0004
							0.528 ± 0.170	0.0103 ± 0.0031	0.440 ± 0.015	0.0041 ± 0.0009	−0.0003 ± 0.0004

The simulation conditions are as described in the footnote to Table 2. For each set of parameters, we give three sets of estimates: The estimates in the first row are obtained by the Deng-Lynch method (Deng and Lynch 1996); the estimates in the second and third rows are obtained by Equation 16; U is given equal to the true values in the second row and the estimate ($\hat{U}$) of U by regression Equation 19 in the third row [when β = 0.5, $\hat{U} = 0.170 + 1.940 \times \hat{U}_1 - 2.859 \times \hat{s}_1$ ($R = 0.997$, $P < 0.001$); when β = 2.0, $\hat{U} = 0.00734 + 1.256 \times \hat{U}_1 - 1.443 \times \hat{s}_1$ ($R = 0.998$, $P < 0.001$)].

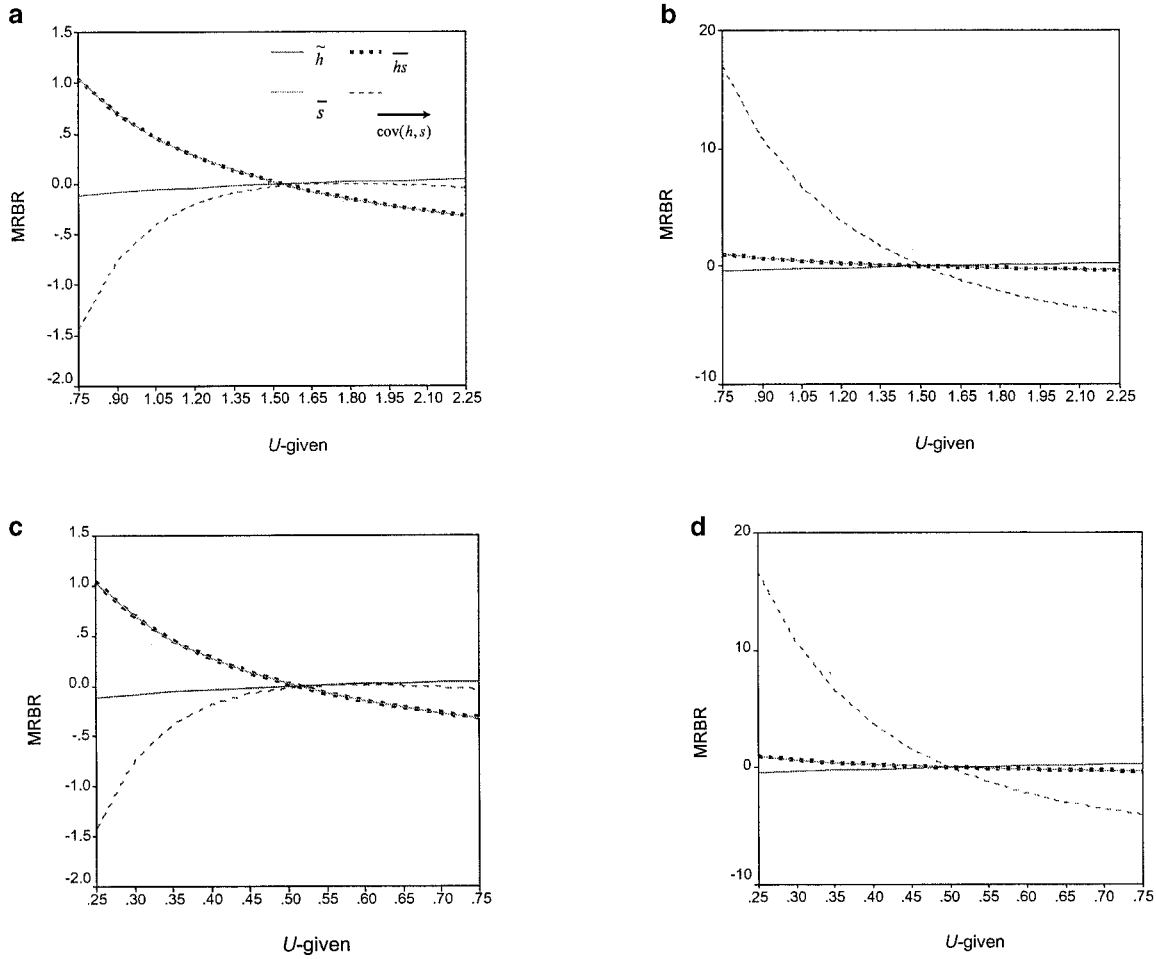

FIGURE 1.—The changes in RBR of the estimates of $\bar{s}$, $\tilde{h}$, $\overline{hs}$, **cov(*h*, *s*)** obtained by Equation 8 in outcrossing populations when U were given equal to the values that ranged from $0.5U_0$ to $1.5U_0$. Each data point was the mean in 1000 simulations with the following sets of parameters and $\beta = 1.0$: (a) $U_0 = 1.5$, $\bar{s} = 0.01$, and $\alpha = 100$; (b) $U_0 = 1.5$, $\bar{s} = 0.05$, and $\alpha = 20$; (c) $U_0 = 0.5$, $\bar{s} = 0.01$, and $\alpha = 100$; and (d) $U_0 = 0.5$, $\bar{s} = 0.05$, and $\alpha = 20$.

(1996), our current estimation method can generally still yield relatively robust estimates of DGM parameters (except **cov(*h*, *s*)** for outcrossing populations when α is as small as 20). In outcrossing populations, the MRBR changed the sign in the robustness investigation of **cov(*h*, *s*)** when $\bar{s} = 0.01$ and 0.05, respectively. This is because the parameter value **cov(*h*, *s*)** changed the sign from negative to zero and then to positive values under the functions assumed when $\bar{s}$ changes from 0.047 to 0.048.

DISCUSSION

We have developed a method in this study for considering variable mutation effects across loci in the estimation. The method may yield improved estimation over that of DENG and LYNCH (1996) as shown by employing additional and independent information (such as the covariance between mean fitness of parents and that of their progeny) to that employed in DENG and LYNCH (1996), although the experimental design is the same. Importantly, cov(h, s) for DGM can be estimated (Equation 10) from an experiment for the first time. Previously, a negative correlation between h and s has long been conjectured from theory only (KACSER and BURNS 1981) and from limited data (SIMMONS and CROW 1977; CROW and SIMMONS 1983). There has been no formal statistical analysis and experimental design to characterize cov(h, s).

Characterization of cov(h, s) is important, for example, for testing the validity of the dominance hypothesis (TURELLI and ORR 1995) in explanation of Haldane's rule. Haldane's rule states that when one sex is inviable or sterile in the hybrids of two different animal races, that sex is often the heterogametic sex. The dominance hypothesis (TURELLI and ORR 1995) states that alleles decreasing hybrid fitness are partially recessive. For the dominance hypothesis to explain Haldane's rule, it is necessary that cov(h, s) is <0. Hence, our estimation method here may offer the first opportunity to test the validity of the dominance hypothesis in explaining Hal-

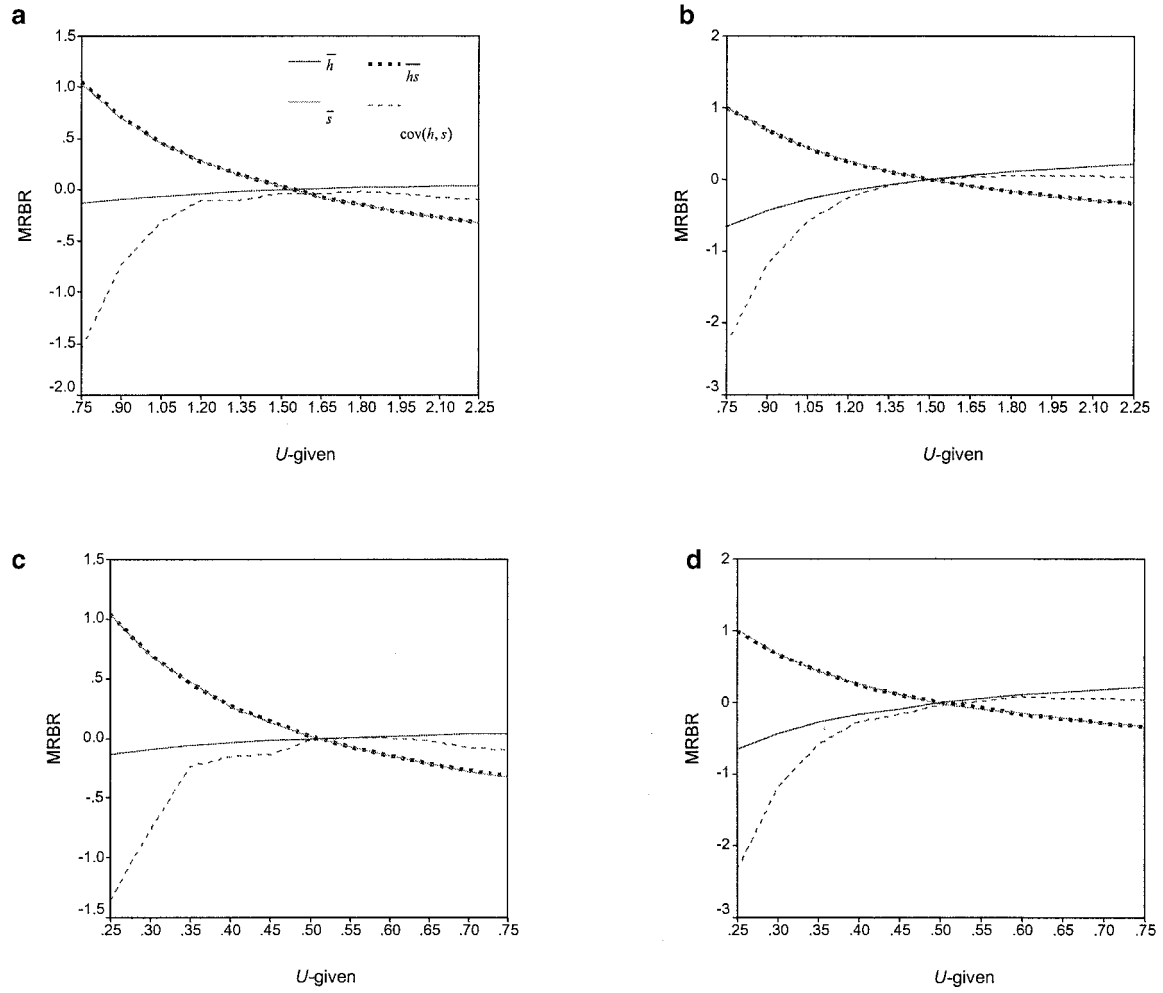

FIGURE 2.—The changes in MRBR of the estimates of $\bar{s}$, $\bar{h}$, $\overline{hs}$, cov(h, s) obtained by Equation 16 in selfing populations when U were given equal to the values that ranged from $0.5U_0$ to $1.5U_0$. Each data point was the mean in 1000 simulations with the following sets of parameters and $\beta = 1.0$: (a) $U_0 = 1.5$, $\bar{s} = 0.01$, and $\alpha = 100$; (b) $U_0 = 1.5$, $\bar{s} = 0.05$, and $\alpha = 20$; (c) $U_0 = 0.5$, $\bar{s} = 0.01$, and $\alpha = 100$; and (d) $U_0 = 0.5$, $\bar{s} = 0.05$, and $\alpha = 20$.

dane's rule by characterizing the sign of cov(h, s). Although it would be nice and significant to have estimators for the other DGM parameters as well, such as variance of s, the observable phenotypic moments of fitness do not relate to other DGM parameters (including the variance of s) in our analytical derivation that considers mutation effects in Equations 2–7 and 11–15.

In the estimation of the DGM parameters, we need a prior estimate of one of the six parameters based on some external knowledge or based on the estimates obtained from alternative approaches or from the same experimental design by using the Deng-Lynch method as demonstrated here. We provided the estimators of the other DGM parameters by using Equations 8 and 16 when assuming that U is known or estimated via other approaches. If we assume that one of the parameters $\bar{s}$, $\tilde{h}$ ($\bar{h}$), or $\overline{hs}$ is known or estimated from other approaches, estimators of the other DGM parameters can be obtained. Among the parameters, $\bar{s}$ and $\overline{hs}$, $\tilde{h}$ (h) can be estimated individually with the analysis methods already developed (MUKAI *et al.* 1972; DENG 1998a) or with the Deng-Lynch method. We present in the APPENDIX the estimators of other DGM parameters when $\bar{h}$ is assumed or estimated and some representative simulation results.

It can be seen from Equations 1a and 1b that the mean of h for the Charlesworth technique (CHARLESWORTH *et al.* 1990) in estimating U in selfing populations should be the arithmetic mean $\bar{h}$, and the mean for the Morton technique (MORTON *et al.* 1956) in outcrossing populations should be the harmonic mean $\tilde{h}$. This has seldom, if ever, been pointed out because the Morton-Charlesworth technique was derived under constant mutation effects. To our knowledge, there has been no method for estimating either $\tilde{h}$ or $\bar{h}$. Our proposed estimation methods here are able to, again for the first time, allow estimates of $\tilde{h}$ and $\bar{h}$ with relatively small bias under variable mutation effects.

The majority of earlier estimation methods for DGM assume constant mutation effects. The only exception is the maximum-likelihood estimation developed for analyses of mutation-accumulation experiments (KEIGHTLEY

1994, 1996). Like our current estimation method, Keightley's maximum-likelihood estimation also needs to assume a parameter value of DGM to estimate the other DGM parameters in his model. Our results (Deng and Li 2001) suggest that a method that accounts for variable mutation effects does not necessarily always yield better estimation than a method that assumes constant mutation effects even under variable mutation effects. In our current estimation, the covariance between mean fitness of parents and that of their progeny is independent of the other measurable experimental data (such as the means and genetic variance of fitness of the two generations across inbreeding/outcrossing) that are used in the Deng-Lynch estimation (Deng and Lynch 1996). This additional and independent information contributes to the improved estimation of our current method in quality and to our ability to estimate additional DGM parameters that could be estimated earlier.

For our methods that are applicable to natural outcrossing populations and selfing-fertilizing populations, M-S balance is assumed to be the mechanism maintaining variation for fitness. Alternatives to M-S balance, such as functional overdominance or overdominance induced by fluctuating selection, may, in principle, maintain polymorphisms. Most evidence suggests dominance as heterozygous mutation effects and thus is compatible with M-S balance (Houle 1989, 1994; Houle *et al.* 1996; Deng *et al.* 1998). However, mechanisms responsible for the maintenance of genetic variance are complex and may differ among populations. If any other mechanism, such as balancing selection or migration, leads to the maintenance of genetic variation (Drake *et al.* 1998; Keightley 1998), our methods may result in biased estimation. Using approaches (Li *et al.* 1999; Li and Deng 2000; H.-W. Deng and J. Li, unpublished results) that we have used to investigate the robustness of the Deng-Lynch method in the presence of violation of the M-S balance assumption, we can and we will pursue in our future studies investigation of how robust the current method is with different degrees of violation of M-S balance assumption.

We are grateful to anonymous reviewers for their constructive comments that greatly helped to improve the manuscript. This work is partially supported by grant R01 GM60402-01A1 from National Institutes of Health (NIH). H.W. Deng was also partially supported by grants from Health Future Foundation, NIH K01 grant AR02170-01, NIH R01 grant AR45349-01, grants from State of Nebraska Cancer and Smoking Related Disease Research Program (LB598), Nebraska Tobacco Settlement Fund (LB692), NIH grant P01 DC01813-07, U.S. Department of Energy grant DE-FG03-00ER63000/A00, grants (30025025 and 30170504) from National Science Foundation of China, and grants from HuNan Normal University and the Ministry of Education of China.

LITERATURE CITED

Bataillon, T., 2000 Estimation of spontaneous genome-wide mutation rate parameters: Whither beneficial mutation? Heredity **84:** 497–501.

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

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

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

Crow, J. F., 1993b Mutation, mean fitness, and genetic load, pp. 3–42 in *Oxford Surveys in Evolutionary Biology*, Vol. 9, edited by D. J. Futuyma and J. Antonovics. Oxford University Press, Oxford, New York.

Crow, J., 1995 Spontaneous mutations as risk factors. Exp. Clin. Immunogenet. **12:** 121–128.

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/New York.

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 Characterization of deleterious mutation rate and properties in outcrossing populations. Genetics **150:** 945–956.

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

Deng, H.-W., and J. Li, 2001 Comparison of two estimation methods for mutation accumulation experiments: maximum likelihood and method of moments. Life Sci. Res. **5:** 189–201.

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, 1998 Inferring the major genomic mode of dominance and overdominance. Genetica **102/103:** 559–567.

Deng, H.-W., J. Li and J.-L. Li, 1999 On the experimental designs and data analyses of mutation accumulation experiments. Genet. Res. **73:** 147–164.

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

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

Gao, H.-X., 1995 *Statistical Computation.* Peking University Press, Beijing.

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

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.

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.

Keightley, P. D., 1998 Inference of genome-wide mutation rates and distributions of mutation effects for fitness traits: a simulation study. Genetics **150:** 1283–1293.

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

Li, J., and H.-W. Deng, 2000 Estimation of parameters of deleterious mutations in partial selfing or partial outcrossing populations and in nonequilibrium populations. Genetics **154:** 1893–1906.

Li, J.-L., J. Li and H.-W. Deng, 1999 The effects of overdominance on characterizing deleterious genomic mutations in natural populations. Genetics **151:** 895–913.

Lynch, M., J. Blanchard, T. Kibota, S. Schultz, L. Vassilieva *et al.*,

1999 Perspective: spontaneous deleterious mutation. Evolution **53:** 645–663.

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:** 315–332.

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., 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.

Simmons, M. J., and J. F. Crow, 1977 Mutations affecting fitness in *Drosophila* populations. Annu. Rev. Genet. **11:** 49–78.

Turelli, M., and A. Orr, 1995 The dominance theory of Haldane's rule. Genetics **140:** 389–402.

Communicating editor: Z-B. Zeng

APPENDIX: ESTIMATION OF OTHER DGM PARAMETERS WHEN $\bar{h}$ IS ASSUMED OR ESTIMATED AND SOME REPRESENTATIVE SIMULATION RESULTS

If $\tilde{h}$ (in outcrossing populations) or $\bar{h}$ (in selfing populations) is known by other estimation methods or assumed at particular values on the basis of some external knowledge, based on Equations 2–7 and 11–15, we have estimators for other DGM parameters as follows, the notations being the same as in the text, in outcrossing populations,

$$U = \frac{4y}{2 - (1/\tilde{h})}, \quad \bar{s} = \frac{4b - 2x}{U}, \quad \overline{hs} = \frac{x}{U}. \tag{A1}$$

and in selfing populations,

$$U = \frac{y}{(0.5 - \bar{h})}, \quad \bar{s} = \frac{2x}{U}, \quad \overline{hs} = \frac{2b}{U}. \tag{A2}$$

Simulations are performed similar to that described in the text and with the above estimation for other DGM parameters when $\tilde{h}$ (in outcrossing populations) or $\bar{h}$ (in selfing populations) is known or estimated. The simulation and the experimental procedures, when $\tilde{h}$ (in outcrossing populations) and $\bar{h}$ (in selfing populations) are estimated by the methods of Deng (1998a) or Mukai *et al.* (1972), are detailed in Deng *et al.* (1998) and thus are not elaborated here.

Some representative results are presented in Tables A1 and A2. It can be seen that, relative to the Deng-Lynch method, the new method developed here can estimate more parameters, such as cov(h, s) and its sign. In an outcrossing population, the sign of cov(h, s) cannot be reliably estimated. However, in selfing populations, if the $\bar{h}$ is estimated first by the Deng-Lynch method and then used in the current method, the sign of cov(h, s) can be characterized correctly.

TABLE A1

Parameter estimates under variable mutational effects in outcrossing populations when $\tilde{h}$ is known or estimated ($\beta = 1.0$)

U	$\bar{s}$	$\bar{h}$	α	$\overline{hs}$	$\tilde{h}$	$\mathrm{cov}(h, s)$	$\hat{U}$	$\hat{\bar{s}}$	$\hat{\bar{h}}$	$\hat{\overline{hs}}$	$\mathbf{cov}(\hat{h}, s)$
1.50	0.05	0.30	20	0.0092	0.175	−0.0060	0.803 ± 0.076	0.1541 ± 0.0209	0.112 ± 0.008		
							1.496 ± 0.020	0.0501 ± 0.0079		0.0092 ± 0.0010	0.0004 ± 0.0011
							0.803 ± 0.076	0.0944 ± 0.0185		0.0171 ± 0.0016	0.0066 ± 0.0007
							0.244 ± 0.036	0.3093 ± 0.0383	0.040 ± 0.006	0.0570 ± 0.0073	0.0446 ± 0.0080
	0.01	0.44	100	0.0039	0.435	−0.0005	0.753 ± 0.067	0.0207 ± 0.0026	0.383 ± 0.008		
							1.527 ± 0.018	0.0099 ± 0.0009		0.0039 ± 0.0004	−0.0004 ± 0.0001
							0.753 ± 0.067	0.0202 ± 0.0026		0.0079 ± 0.0009	0.0002 ± 0.0000
							0.572 ± 0.045	0.0265 ± 0.0028	0.357 ± 0.008	0.0104 ± 0.0009	0.0009 ± 0.0001
0.50	0.05	0.30	20	0.0092	0.175	−0.0060	0.265 ± 0.022	0.1557 ± 0.0182	0.111 ± 0.007		
							0.501 ± 0.011	0.0494 ± 0.0069		0.0091 ± 0.0009	0.0004 ± 0.0009
							0.265 ± 0.022	0.0940 ± 0.0156		0.0172 ± 0.0014	0.0068 ± 0.0007
							0.079 ± 0.014	0.3171 ± 0.0433	0.039 ± 0.006	0.0587 ± 0.0097	0.0465 ± 0.0103
	0.01	0.44	100	0.0039	0.435	−0.0005	0.250 ± 0.024	0.0210 ± 0.0029	0.383 ± 0.009		
							0.510 ± 0.010	0.0100 ± 0.0011		0.0039 ± 0.0004	−0.0004 ± 0.0001
							0.250 ± 0.024	0.0205 ± 0.0028		0.0080 ± 0.0010	0.0019 ± 0.0000
							0.190 ± 0.016	0.0269 ± 0.0031	0.356 ± 0.009	0.0105 ± 0.0010	0.0010 ± 0.0001

For each set of parameters given in the first seven columns, 200 simulations were performed. Each simulation is based on 200 parents and dividing the range of s_i (0–1) into 200 equal intervals each with a width of 0.005. For outcrossing populations, 40 selfed progeny were evaluated per parent. Reported values for $\hat{U}$, $\hat{\bar{s}}$, $\hat{\bar{h}}$ $\hat{\bar{h}}$, $\hat{\overline{hs}}$, and $\mathbf{cov}(\hat{h}, s)$ are the mean ± SD. For each set of parameters, we give four sets of estimates; the estimates in the first row are obtained by the Deng-Lynch method (Deng and Lynch 1996). The estimates in the second, third, and fourth rows are obtained by Equation A1; $\tilde{h}$ is set equal to its true parameter value in the second row; $\tilde{h}$ is equal to $\bar{h}$, which is estimated by the Deng-Lynch method in the third row; and $\tilde{h}$ is equal to $\bar{h}$, which is estimated by the Deng method in the fourth row (Deng 1998a).

TABLE A2

Parameter estimates under variable mutational effects in outcrossing populations when $\bar{h}$ is known or estimated ($\beta = 1.0$)

U	$\bar{s}$	$\bar{h}$	α	$\overline{hs}$	$\mathrm{cov}(h, s)$	$\hat{U}$	$\hat{\bar{s}}$	$\hat{\bar{h}}$	$\hat{\overline{hs}}$	$\mathbf{cov}(\hat{h}, s)$
1.50	0.05	0.30	20	0.0092	−0.0060	1.051 ± 0.056	0.0713 ± 0.0086	0.218 ± 0.012		
						1.501 ± 0.060	0.0498 ± 0.0047		0.0092 ± 0.0011	−0.0059 ± 0.0013
						1.051 ± 0.056	0.0713 ± 0.0086		0.0131 ± 0.0013	−0.0024 ± 0.0009
						0.899 ± 0.041	0.0832 ± 0.0083	0.171 ± 0.010	0.0153 ± 0.0016	0.0012 ± 0.0014
	0.01	0.44	100	0.0039	−0.0005	0.914 ± 0.242	0.0178 ± 0.0047	0.398 ± 0.021		
						1.541 ± 0.078	0.0099 ± 0.0011		0.0039 ± 0.0004	−0.0005 ± 0.0004
						0.914 ± 0.242	0.0178 ± 0.0047		0.0068 ± 0.0013	−0.0002 ± 0.0004
						0.802 ± 0.053	0.0191 ± 0.0023	0.389 ± 0.005	0.0075 ± 0.0009	0.00005 ± 0.0007
0.50	0.05	0.30	20	0.0092	−0.0060	0.349 ± 0.023	0.0722 ± 0.0090	0.218 ± 0.014		
						0.500 ± 0.036	0.0503 ± 0.0047		0.0093 ± 0.0011	−0.0060 ± 0.0015
						0.349 ± 0.023	0.0722 ± 0.0090		0.0132 ± 0.0014	−0.0024 ± 0.0009
						0.299 ± 0.022	0.0841 ± 0.0089	0.171 ± 0.012	0.0155 ± 0.0018	0.0011 ± 0.0016
	0.01	0.44	100	0.0039	−0.0005	0.300 ± 0.071	0.0179 ± 0.0046	0.398 ± 0.020		
						0.513 ± 0.045	0.0100 ± 0.0012		0.0039 ± 0.0005	−0.0005 ± 0.0004
						0.300 ± 0.071	0.0179 ± 0.0046		0.0069 ± 0.0012	−0.0002 ± 0.0004
						0.268 ± 0.026	0.0192 ± 0.0025	0.390 ± 0.005	0.0075 ± 0.0010	0.00003 ± 0.0007

For each set of parameters given in the first six columns, 1000 simulations were performed. Each simulation is based on 200 parents and dividing the range of s_i (0–1) into 200 equal intervals. For selfing populations, 200 random outcrossed progeny were evaluated per simulation (1 for each of the original 200 parents, randomly outcrossed). Values reported for $\hat{U}$, $\hat{\bar{s}}$, $\hat{\bar{h}}$ $\hat{\overline{hs}}$, and $\mathbf{cov}(\hat{h}, s)$ are the mean ± SD. For each set of parameters, we give four sets of estimates: The estimates in the first row are obtained by the Deng-Lynch method (Deng and Lynch 1996). The estimates in the second, third, and fourth rows are obtained by Equation A2; $\bar{h}$ is given equal to the true value in the second row, $\bar{h}$ is estimated by the Deng-Lynch method in the third row, and $\bar{h}$ is estimated by the Mukai method (Mukai *et al.* 1972).