Estimation of Parameters of Deleterious Mutations in Partial Selfing or Partial Outcrossing Populations and in Nonequilibrium Populations

Jian Li* and Hong-Wen Deng*†

*Osteoporosis Research Center and Department of Biological Sciences, Creighton University, Omaha, Nebraska 68131 and †College of Life Sciences, Hunan Normal University, ChangSha, Hunan 410081, People's Republic of China

Manuscript received August 27, 1999
Accepted for publication December 21, 1999

ABSTRACT

The Deng-Lynch method was developed to estimate the rate and effects of deleterious genomic mutations (DGM) in natural populations under the assumption that populations are either completely outcrossing or completely selfing and that populations are at mutation-selection (M-S) balance. However, in many plant and animal populations, selfing or outcrossing is often incomplete in that a proportion of populations undergo inbreeding while the rest are outcrossing. In addition, the degrees of deviation of populations from M-S balance are often not known. Through computer simulations, we investigated the robustness and the applicability of the Deng-Lynch method under different degrees of partial selfing or partial outcrossing and for nonequilibrium populations approaching M-S balance at different stages. The investigation was implemented under constant, variable, and epistatic mutation effects. We found that, generally, the estimation by the Deng-Lynch method is fairly robust if the selfing rate (S) is <0.10 in outcrossing populations and if S > 0.8 in selfing populations. The estimation may be unbiased under partial selfing with variable and epistatic mutation effects in predominantly outcrossing populations. The estimation is fairly robust in nonequilibrium populations at different stages approaching M-S balance. The dynamics of populations approaching M-S balance under various parameters are also studied. Under mutation and selection, populations approach balance at a rapid pace. Generally, it takes 400–2000 generations to reach M-S balance even when starting from homogeneous individuals free of DGM. Our investigation here provides a basis for characterizing DGM in partial selfing or outcrossing populations and for nonequilibrium populations.

THE genome of all organisms is subject to continuous bombardment of mutations, the majority of which are deleterious. Numerous theories based on deleterious genomic mutations (DGM) have been developed to explain some fundamental phenomena in biology (Deng 1998b; see references therein). The validity of these theories critically depends on the rate at which DGM occur per genome per generation (U) and/or their effects (mean dominance and selection coefficients fi and si). In addition, the knowledge about the mean and the variation of the effects of DGM is vital in evaluating the role of DGM in long-term survival of small populations (Lynch et al. 1993, 1995a,b; Lande 1994).

However, few estimates of the parameters of DGM are available, and even the order of magnitude of these parameters is controversial (Crow and Simmons 1983; Kondrashov 1988; Crow 1993a,b; Keightley 1996). None of the current estimation approaches (Morton et al. 1956; Bateman 1959; Mukai et al. 1972; B. Charlesworth et al. 1990; Deng and Lynch 1996) can yield unbiased results under realistic situations (Deng and Fu 1998; Deng et al. 1999a,b; W.-M. Chen, Y. Zhou and T. Kibota, unpublished results).

Furthermore, the statistical properties (sampling errors and, particularly, the direction and magnitude of bias) of the current estimation approaches are largely unknown under various biologically plausible situations. Nevertheless, our knowledge of these statistical properties is essential, at least for (1) understanding the robustness and applicability of various estimation approaches under various biology conditions; and (2) a sensible decision on the adoption of a particular estimation approach in a particular biological system. Therefore, our knowledge of these statistical properties may not only expand the range of biological systems for which the current estimation approaches can apply but may also provide a quantitative and statistical basis for correctly inferring DGM from these systems. It has been argued that robust characterization of DGM is one of the greatest challenges in the broad field of modern genetics (Crow 1993a,b; Bridges et al. 1994; Peck and Eyre-Walker 1997).

Among the available approaches, Deng and Lynch’s (1996) estimation makes use of the data on the changes of both the mean and genetic variance for fitness traits that can be acquired from inbreeding/ outbreeding in natural outcrossing/ selfing populations. Since the trait
under study for the Deng-Lynch method is fitness or its important component(s) under directional selection, it is understood hereafter that the mean and genetic variance referred to are for fitness traits. The Deng-Lynch method estimates not only $U$ but also $\bar{h}$, $\bar{s}$, and the genetic variance introduced by DGM per generation (Deng and Lynch 1996). The method was applied to Daphnia (a freshwater microcrustacean; Deng and Lynch 1997), yielding data on the characteristics of DGM that are compatible with those obtained from the time- and labor-consuming mutation-accumulation experiments (Mukai et al. 1972). In addition, Deng and Lynch’s (1996) estimation approach almost always yields the best estimation (as reflected by the mean square error, a composite index of both bias and sampling variance), when compared with other estimation approaches (Deng and Fu 1998a). However, Deng and Lynch’s (1996) estimation assumes that populations under study are either complete outcrossing or complete selfing. Although this assumption holds reasonably well in populations of many species (such as Daphnia; Deng and Lynch 1997), in populations of many other species, especially in many plant populations (Stebbins 1957; Lande and Schemske 1985; Schemske and Lande 1985; Barrett and Eckert 1990; Charlesworth 1992), this assumption is violated to various degrees. Although bias is expected for partial selfing or partial outcrossing populations, it is important to investigate the magnitude of bias, especially for highly selfing and highly outcrossing populations, as they are frequent (Lande and Schemske 1985; Schemske and Lande 1985).

Another important assumption of Deng and Lynch’s (1996) estimation is that the study population is at mutation-selection (M-S) balance. Despite extensive efforts (e.g., Houle et al. 1996; Deng 1998a; Deng et al. 1998; Charlesworth and Hughes 1999), the validity to which this assumption is reasonable in natural populations is largely unknown and is difficult to measure. It is even more difficult to measure the relative contributions of various potential mechanisms to the deviation from M-S balance (if any). The mechanisms and their relative importance in maintaining populations’ genetic variability are largely unknown. In addition, we know little about the dynamics of populations approaching M-S balance under recurrent mutations and selection. Specifically, we do not know how fast M-S balance is established and how long it takes when starting from a homogeneous population (all individuals being free of DGM) subject to deleterious mutations and selection. Li et al. (1999) investigated the robustness of Deng and Lynch’s (1996) estimation when the M-S assumption is violated by the presence of balancing selection as reflected as overdominance at some genomic loci. Drake et al. (1998) also made some qualitative predictions of the estimation bias of the Deng-Lynch method when substantial genetic variation is actively maintained by selection. However, under directional selection for DGM underlying fitness, the robustness of Deng and Lynch’s (1996) estimation in nonequilibrium populations at different stages approaching M-S balance is unknown and has never even been qualitatively discussed.

To characterize DGM in natural populations of partial selfing or outcrossing, ideally, we need to develop new methods to estimate mutation parameters by incorporating in estimation the selfing or outcrossing rates that are estimable by the methods of Ritland (1984, 1988), Milligan (1996), and Fu. Our extensive investigation (J. Li and H.-W. Deng, unpublished data) has suggested that this line of research effort may not be tenable. This is mainly because there is no accurate and simple analytical distribution to characterize the number of DGM per genome for partial selfing/outcrossing populations at M-S balance. The semianalytical distribution developed by Charlesworth et al. (1991) is not simple and requires computation iteration to approximate the number of DGM per genome for partial selfing/outcrossing populations at M-S balance. Hence, we cannot develop an analytical method for characterizing DGM based on the semianalytical distribution developed by Charlesworth et al. (1991).

Therefore, we set out to investigate the robustness and statistical properties of Deng and Lynch’s (1996) estimation in partial selfing/outcrossing populations and in populations not at M-S balance. The investigation is performed by extensive computer simulations under various situations (constant, variable, and epistatic mutation effects) and various parameters. The results on the direction and magnitude of bias and sampling variation of the estimates should provide a basis for robust characterization of DGM in natural populations of partial selfing or outcrossing and in populations under mutation and selection but not at M-S balance.

**THEORY**

The approach to characterize DGM proposed by Deng and Lynch (1996) is a fitness-moments approach, which utilizes mean and genetic variance of fitness of two generations across inbreeding or outcrossing in outcrossing and selfing populations, respectively. The Deng and Lynch (1996) estimation assumes that mating is either complete outcrossing or complete selfing, and all loci are at M-S balance for segregating polymorphism, multiplicative fitness functions, constant mutation effects, infinite population size, and free recombination (or linkage equilibrium) between loci. Under these assumptions, letting a circumflex (”) denote an estimate, in an outcrossing population,

\[ \hat{h} = \frac{1}{4\sqrt{2}\bar{h} - 2}, \quad \hat{U} = \frac{4\hat{h}}{2h - 1}, \quad \hat{s} = \frac{\hat{x}}{\hat{U}h}. \] (1)

In a selfing population,
\[
\hat{h} = \sqrt{\frac{x}{2z}}, \quad \hat{u} = \frac{y}{\hat{h} - 0.5}, \quad \hat{s} = \frac{2x}{\hat{u}}.
\]

where \(x, y, \) and \(z\) are defined, respectively, as
\[
x = \ln\left(\frac{\sigma^2_{(o)} + 1}{W_{(o)}}\right), \quad y = \ln\left(\frac{W_{(s)}}{W_{(o)}}\right), \quad z = \ln\left(\frac{\sigma^2_{(s)}}{W_{(s)}} + 1\right),
\]

where, in outcrossing populations, \(W_{(o)}\) and \(\sigma^2_{(o)}\) are, respectively, the mean and genetic variance of fitness, and \(W_{(s)}\) and \(\sigma^2_{(s)}\) are, respectively, the mean and genetic variance of fitness among selfed progeny families. In selfing populations, \(W_{(s)}\) and \(\sigma^2_{(s)}\) are, respectively, the mean and genetic variance of fitness, and \(W_{(o)}\) and \(\sigma^2_{(o)}\) are, respectively, the mean and genetic variance of fitness among outcrossed progeny.

The experimental procedures in outcrossing and selfing populations are detailed previously (Deng and Lynch 1996, 1997; Deng and Fu 1998; Deng 1998b). Briefly, in outcrossing populations, (1) a sample of genotypes are selfed to obtain a number of selfed progenies from each parent to form selfed families; (2) parental genotypes are cloned to estimate \(W_{(o)}\) and \(\sigma^2_{(o)}\) are, respectively, the mean and genetic variance of fitness, and \(W_{(s)}\) and \(\sigma^2_{(s)}\) are, respectively, the mean and genetic variance of fitness among outcrossed progeny.

SIMULATIONS

By a deterministic method that was developed by Kon-drashov (1985) and later modified by D. Charlesworth et al. (1990), incomplete selfing or outcrossing populations are constructed by computer simulations in which mutation effects are constant, variable, or epistatic across loci as in Deng and Lynch (1996). Except in the case of epistatic mutation effects, fitness function is assumed to be multiplicative across loci, which is biologically plausible (Deng 1998b and references within). Populations are simulated starting from homogeneous individuals free of DGM. The dynamics of the populations approaching M-S balance is recorded. The estimation of Deng and Lynch (1996) is applied to the simulated nonequilibrium populations at different stages approaching the M-S balance and the populations that have already reached M-S balance for fitness or its important components that are normally quantitative traits, the phenotypic values are determined not only by genotypic values but also by random environmental sources of variation (Lynch and Walsh 1998). Since the focus here is on the robustness of the assumption under partial selfing/outcrossing and in nonequilibrium populations, the estimation bias is of major interest. Therefore, genotypic values of fitness or its important components are assumed to be measured without error throughout in simulations. In reality, this would require that each genotype be clonally replicated and assayed a very large number of times.

Ignoring measurement error for genotypic values will probably reduce the sampling error of estimates but is unlikely to bias the estimation. This is supported by our previous investigation of measurement error for mutation parameter estimation by the Deng-Lynch technique (Deng and Lynch 1996; Deng and Fu 1998; Deng et al. 1999).

Population construction: Assuming that the population size is infinite and the number of loci of each individual is very large, all new mutations may be considered as only occurring on wild-type homozygous loci. Hence, new mutations only change the number of heterozygous loci. The number of new mutations per genome per generation follows a Poisson distribution with mean \(\lambda\). To focus on the effects of partial selfing/outcrossing and M-S disequilibrium on the estimation, we assume that all loci are unlinked. The situation of linkage disequilibrium among DGM at different loci is being investigated (H.-W. Deng and J. Li, unpublished results).

For each simulation cycle, starting from adult individuals, mutation (in the \(t\)th generation), mating [to generate zygotes for the \((t + 1)\)th generation], and selection [in the \((t + 1)\)th generation] are simulated sequentially. In each simulation, a population free of any mutation is employed as the starting population. In each generation, the population includes two subpopulations, in one of which mating is outcrossing and in the other selfing. The mutation-selection process is performed with the two subpopulations separately. Then they are combined to generate zygote pools, with proportional contributions being \(1 - S\) and \(S\), respectively, for the outcrossing subpopulation and for the selfing subpopulation. \(S\) is the selfing rate in the whole population. Then the cycle is repeated for the next generation. Our data (J. Li and H.-W. Deng, unpublished results) show that this simulation is essentially identical to simulating a large population with each individual having a probability of \(S\) for selfing and a probability of \((1 - S)\) for outcrossing. M-S balance is considered to be reached when the relative changes of both mean and genetic
variance of fitness of the whole population in two contiguous generations are \(< 1 \times 10^{-8}\). Selfing is employed as a special inbreeding form for partial outcrossing populations to demonstrate the qualitative effects of inbreeding on the estimation in outcrossing populations. The effects demonstrated should be an upper bound for inbreeding under the same inbreeding rate in natural populations, as selfing is the most extreme form of inbreeding.

Following Kondrashov (1985) and D. Charlesworth et al. (1990), the simulation procedure is detailed below. Denote \( p_{ij(A)} \), \( p_{ij(A)}' \), and \( p_{ij(A)}'' \), respectively, as the frequencies of individuals with \( i \) heterozygous and \( j \) homozygous mutations before mutation, recombination, and selection in \( A \), where \( A \) denotes the whole population (T), the selfing (S) subpopulation, or the outcrossing (O) subpopulation. In the initial population, \( p_{ij(T)} \) is 1 for \( i, j = 0 \) and is 0 for all others. In each generation after mutation, under the assumption that new mutations only occur on wild-type homozygous loci, we have individual frequencies in outcrossing and selfing subpopulations, respectively, as

\[
\begin{align*}
p_{ij(O)} &= e^{-u} \sum_{k=0}^{i} p_{ik(O)} \frac{U^{i-k}}{(i-k)!}, \quad i, j = 0, 1, \ldots, \\
p_{ij(S)}' &= e^{-u} \sum_{k=0}^{i} p_{ik(S)} \frac{U^{i-k}}{(i-k)!}, \quad i, j = 0, 1, \ldots.
\end{align*}
\]

Then, mating is simulated within each of the two subpopulations. In the selfing subpopulation, the individual frequencies are changed to

\[
p_{ij(S)}'' = \sum_{k} \sum_{l} p_{ij(S)}'' s_{ij}(k, l), \quad i, j, k, l = 0, 1, \ldots,
\]

where \( s_{ij}(k, l) \) is the probability that an individual having \( k \) heterozygous and \( l \) homozygous mutations produces an offspring with \( i \) heterozygous and \( j \) homozygous mutations by selfing,

\[
s_{ij}(k, l) = \begin{cases} 
k\binom{k-i}{j} \binom{l-1}{k-l-1} & \text{if } k, l \in [i, k + l - i]. \\
0 & \text{otherwise}
\end{cases}
\]

In the outcrossing subpopulation, assuming that there is no shared mutant allele for any two individuals, the outcrossed progeny will have no homozygous mutational loci. The frequencies of different individuals are

\[
p_{ij(O)}'' = \sum_{g} \sum_{h} \sum_{k} \sum_{l} p_{gk(O)}'' p_{hj(O)}'' b(g, h; i - h - l),
\]

where \( b(x, y; z) \) is the frequency of producing an offspring with \( z \) heterozygous mutational loci by two parents having \( x \) and \( y \) heterozygous mutational loci, respectively. It is easy to show that

\[
b(x, y; z) = \begin{cases} (0.5)^{(x+y-z)} \frac{(x+y)!}{(x+y-z)!}, & z \leq x + y, \\
0, & z > x + y.
\end{cases}
\]

After mating, selection starts to operate. Relative fitness is used to determine the frequencies after selection, i.e.,

\[
p_{ij(O)}'' = p_{ij(O)}'' w_{ij}, \quad i = 0, 1, \ldots
\]

where \( w_{ij} \) is the fitness of an individual with \( i \) heterozygous and \( j \) homozygous mutations; \( w_{ij} = \frac{\sum_{i} \sum_{j} p_{ij(O)}'' w_{ij}}{\sum_{i} \sum_{j} p_{ij(O)}''} \), which is the mean fitness of the selfing subpopulation before selection; and \( w_{ij} = \frac{\sum_{i} \sum_{j} p_{ij(O)}'' w_{ij}}{\sum_{i} \sum_{j} p_{ij(O)}''} \), which is the mean fitness of the outcrossing subpopulation before selection.

Before the next cycle begins, the two subpopulations are merged to obtain the frequencies of individuals with different mutations in the whole population for the next generation:

\[
w_{ij} = \left( 1 - S \right) w_{ij} + S w_{ij},
\]

where \( w_{ij} \) is the fitness of an individual having \( i \) heterozygous and \( j \) homozygous mutations is

\[
w_{ij} = (1 - hs) (1 - g).
\]

In populations that have reached M-S balance, \( K \) individuals are sampled as parents. Progenies are produced by outcrossing (in populations with \( S > 0.5 \)) or selfing (in populations with \( S < 0.5 \)) the parents. For populations with \( S > 0.5 \) (predominantly selfing populations), one outcrossed progeny per parent is obtained and the Deng-Lynch method (Deng and Lynch 1996) for selfing populations is applied for estimating parameters of DGM. For populations with \( S < 0.5 \) (predominantly outcrossing populations), \( M \) selfed progeny are obtained per parent. The Deng-Lynch method for outcrossing populations is applied for the estimation. When \( S = 0.5 \), the Deng-Lynch methods for outcrossing and selfing populations are both applied for the estimation. The sampling and the estimation procedures for the following two mutation effects are the same. Throughout, unless otherwise, for each parameter set, 500 simulations are performed, \( K = 200 \) and \( M = 40 \).

**Variable mutation effects:** Mutation effects are unlikely to be constant across loci (Caballero and Keightley 1994; Deng and Lynch 1996). We utilized the model described by Deng and Lynch (1996), in
which the distribution of homozygous effects is roughly exponential with density distribution,

\[ p(s) = \frac{1}{s} \exp(-s/s), \]

and there is an inverse relationship between \( h \) and \( s \)

\[ h = \frac{e^{-1s}}{2}. \]

These are consistent with our few data on the distribution of mutation effects (Deng and Lynch 1996). For an individual with \( i \) loci heterozygous and \( j \) loci homozygous for mutations, its fitness is

\[ w_{ij} = \prod_{i=1}^{i}(1 - h_i s_i) \prod_{m=i+1}^{j} (1 - s_m), \]

where \( s_i \) and \( s_m \) come from the above exponential distribution with mean \( s \).

**Epistatic mutation effects**: Although fitness or its component most likely acts multiplicatively (Deng 1998b and references within), synergistically epistatic mutation effects cannot be ruled out entirely. Hence, the robustness of the Deng-Lynch method is also tested under epistatic mutation effects. The model we consider was described by Charlesworth (1990) and employed earlier by us (Deng and Lynch 1996; Deng 1998b). In this model, the fitness of an individual having \( i \) heterozygous and \( j \) homozygous mutations (with constant effects) is

\[ w_{ij} = \exp(-\alpha n - \beta n^2/2), \]

where \( n \) is the effective number of mutations, \( n = h_i + j \). \( \alpha = s \) in a selfing population. The parameter \( \beta \) provides a measure of the synergistic effects of deleterious alleles and the ratio \( \beta n^2 / (2\alpha) \) provides a measure of the relative contribution of synergistic effects to mean fitness. Charlesworth (1990) provided more details on this epistatic model of mutation effects and the details on the relationship of the parameters in this model with \( h \) and \( s \) in outcrossing and selfing populations.

**Estimation for nonequilibrium populations**: All the above estimations are conducted for populations that have already reached M-S balance. To examine the robustness of the estimation of Deng and Lynch (1996) for nonequilibrium populations, we applied the estimation to populations at different stages of approaching M-S balance when they start from homogeneous populations free of DGM. In simulations, we obtain samples every 25 generations during the course of the populations approaching M-S balance. The investigation was conducted for complete outcrossing/selfing populations. These samples are used as the parental generations and corresponding matings (selfing or outcrossing) are performed to yield outcrossed (for selfing populations) or selfed (for outcrossed populations) progenies. Then Deng and Lynch’s (1996) estimation is applied to estimate mutation parameters.

**Dynamics of populations approaching M-S balance**: During the simulations, under various values of \( S \), \( U \), \( h \), and \( s \), starting from homogeneous populations free of DGM, the dynamics of populations are recorded. The dynamics of populations include the mean and genetic variance of populations at different stages approaching M-S balance since starting from a homogeneous population free of DGM and the number of generations taken to reach M-S balance.

Throughout, unless otherwise specified in tables or figures, in simulations reported, \( h \) (or \( \bar{h} \)) is 0.36 and \( s \) (or \( \bar{s} \)) is 0.03, and reported values are the means ± standard deviation (SD). The selected values of \( h \) (or \( \bar{h} \)) = 0.36 and \( s \) (or \( \bar{s} \)) = 0.03 are consistent with those summarized from the extensive experiments in Drosophila (Crow 1993a;b; Lynch et al. 1995a,b).

**RESULTS**

**Estimation under constant mutation effects** *(Table 1)*: Estimated values are unbiased for complete outcrossing \( (S = 0) \) or selfing \( (S = 1.0) \) populations and become biased when the population deviates from complete outcrossing or selfing. The degree of bias increases with increasing deviation from complete outcrossing or selfing, i.e., when \( S \) is getting closer to 0.5. \( \bar{h} \) is always upwardly biased for partial selfing/outcrossing \( (0 < S < 1) \). This is because the ratio \( x/z \) in populations that are critical for \( \bar{h} \) estimation (Equations 1 and 2) increases when \( S \) deviates from 0 or 1 and approaches 0.5 (Figure 1). Recall (Equation 3) that \( x \) and \( z \) are related to the squared coefficient of variation of genotypic values of fitness in the outcrossed or selfed generations, respectively. Generally, the accuracy of \( \bar{U} \) and \( \bar{s} \) decreases when populations deviate more from \( S = 0 \) or \( S = 1 \), as reflected by the increased bias and/or standard deviation. For the same magnitude of deviation from outcrossing or selfing, the biases are larger in predominantly outcrossing populations than in predominantly selfing populations. For example, when \( U \sim 0.5 \), \( h \sim 0.36 \) and \( s \sim 0.03 \), \( \bar{U} = 1.89 \), \( \bar{h} = 0.467 \), and \( \bar{s} = 0.009 \) when \( S = 0.3 \); and \( \bar{U} = 0.70 \), \( \bar{h} = 0.417 \), and \( \bar{s} = 0.027 \) when \( S = 0.7 \). When \( S = 0.5 \), application of the estimation method for selfing populations yields less biased estimates than the application of the estimation method for outcrossing populations. This conclusion also holds under variable and epistatic mutation effects (Tables 2 and 3). When \( S < 0.05 \) and when \( S > 0.9 \), the Deng and Lynch (1996) estimation yields little bias. As is apparent from Table 1, the direction of bias of \( \bar{U} \) and \( \bar{s} \) may be upward or downward, depending upon the different parameter values of \( U \) and \( S \). \( \bar{U} \) can be negative values with smaller SD when \( U = 1.0 \) and \( S \) approaches 0.5 starting from intermediate \( S \) values (\( \sim 0.2 \)). This can be explained as follows. Since \( \bar{y} \) is always negative as revealed in our simulations, the sign of \( \bar{S} \) depends on whether \( \bar{h} > 0.5 \) (Equation 1); if \( \bar{h} < 0.5 \), \( \bar{U} > 0 \) and if \( \bar{h} > 0.5 \), \( \bar{U} < 0 \).
always be obtained, although the estimation is biased values are also obtained for

when \( h \approx 0.5 \) (additive mutation effects, for which there is no crossing populations (\( z = 0 \)). Therefore, \( \hat{U} \) can always be obtained, although the estimation is biased with the bias getting larger when \( h \) is closer to 0.5.

**Estimation under variable mutation effects (Table 2):**

The estimates are generally biased, but not always. When \( 0 < S < 0.5 \) and when \( S \) increases, \( \hat{s} \) decreases from \( \sim 2s \) to negative values. \( \hat{U} \) and \( \overline{h} \) change from downwardly biased to upwardly biased. Therefore, in some intermediate selfing rates, \( \hat{s} \), \( \hat{U} \), and \( \overline{h} \) may be unbiased and this depends on mutation effects, \( \hat{s} \), and mutation parameters that are generally unknown. For some \( S \) values, when \( S \) approaches 0.5 from 0, \( \overline{h} \) approaches values \( \sim 0.5 \), yielding large bias and sampling error for \( \hat{U} \). The biases can be very large in predominantly outcrossing populations (e.g., \( U = 0.2 \) and \( U = 0.10 \). Therefore, the estimates are not that bad. When \( 0.5 < S < 1 \) and when \( S \) decreases, \( \hat{s} \) increases, from \( \sim 2s \) to \( \sim 3.2s \) (for \( U = 0.5 \)) or 4.8s (for \( U = 1.0 \)). \( \hat{U} \) decreases only slightly. Therefore, the estimation under various \( S \) in predominantly selfing populations is more robust than the estimation under various \( S \) in predominantly outcrossing populations. Again, the estimation is fairly robust when \( S < 0.10 \) and when \( S > 0.8 \).

**Estimation under epistatic effects (Table 3):** Under the epistatic mutation model investigated, except under complete selfing, there is no \( s \) parameter for a mutation as the mutation effects are epistatic and depend on other mutations. Although the Deng-Lynch method still yields an \( s \) estimate, it does not make sense to discuss the bias for \( s \) estimation, as the \( s \) parameter does not exist for a single DGM under epistatic effects. Therefore, we only summarize estimation for \( \hat{U} \) and \( \overline{h} \). The estimates are generally biased, but not always. For complete outcrossing populations, \( \hat{U} \) and \( \overline{h} \) are underestimated. For complete selfing populations, \( \hat{U} \) is fairly accurate with little bias. When the population deviates from complete outcrossing (\( S = 0 \)) or complete selfing (\( S = 1 \)) and \( S \) approaches 0.5, \( h \) increases and \( \hat{U} \) has increasingly larger sampling errors, for the same reason outlined.
before. It can be seen that when populations deviate slightly from $S = 0$, the effects of epistasis and partial selfing cancel each other to a certain extent, and the estimates of the mutation parameters become less biased, then unbiased, and then the bias gets larger in the other direction. For example, when $0 < S < 0.5$ and when $S$ increases, $\hat{U}$ is first biased downwardly, then unbiased, and then biased upwardly. Again, the estimation is fairly robust when $S < 0.10$ and when $S > 0.8$.

Robustness of the Deng-Lynch estimation in nonequilibrium populations (Tables 4 and 5): The performance of Deng and Lynch’s (1996) estimation is fairly robust in nonequilibrium populations at different stages approaching M-S balance. Under constant mutation effects (Table 4), $\hat{h}$ and $\hat{s}$ are almost always unbiased for complete outcrossing or selfing populations. $\hat{U}$ is downwardly biased when $S < \sim 200$ (in outcrossing populations) or $S < \sim 75$ (in selfing populations) generations of mutation and selection are experienced by populations starting from homogeneous individuals free of any DGM. After a moderately large number of generations ($>200$ and 75 generations, respectively, in outcrossing and selfing populations), the estimation is generally unbiased for constant mutation effects.

The robust performance of the estimation in nonequilibrium populations can be explained by the detailed analyses of the dynamics of $x$, $y$, and $z$ (Equation 3) in nonequilibrium populations during the course of approaching M-S balance (Figure 2). $x$, $y$, and $z$ change monotonously with the number of generations experienced under mutation and selection. The change is relatively large only at the initial generations. However, the ratios $x/z$ and $x/y$ (or $z/y$) that are important for the estimation of $\hat{U}$, $\hat{h}$, and $\hat{s}$ (Equations 1-3) change little and remain essentially the same after the few initial generations. Since $\hat{h}$ is the function of $x/z$, it will approximately reach a constant after a few generations after the mutation and selection starts to operate. Furthermore, $\hat{s}$ is the function of $\hat{h}$ and $x/y$ (in outcrossing populations) or $z/y$ (in selfing populations). Therefore, $\hat{h}$ and $\hat{s}$ quickly become the same as those in populations at M-S balance. $\hat{U}$ is a function of $\hat{h}$ and $\hat{y}$ (Equations 1 and 2). When $\hat{h}$ remains relatively constant and the absolute value of $\hat{y}$ increases with the more generations of mutation and selection experienced, $\hat{U}$ increases and becomes stable only when $\hat{y}$ approaches equilibrium values.

Similar conclusions hold under variable mutation effects (Table 5). With an increasing number of generations of mutation and selection experienced, the degree of the estimation bias in nonequilibrium populations will quickly approach that in populations at M-S balance. If starting from a homogeneous population free of DGM, it generally takes $\sim 200$ generations (in outcrossing populations) and $\sim 75$ generations (in selfing populations) to reach asymptotic degrees of bias expected for populations at M-S balance.

### Table 2

<table>
<thead>
<tr>
<th>$S$ (outcrossing)</th>
<th>$U = 0.5$</th>
<th>$U = 10$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{U}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{h}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{s}$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Dynamics of nonequilibrium populations approaching
M-S balance (Figures 3 and 4): $S$ has significant effect on the generations needed to reach M-S balance (Figure 3a). All else being the same, the larger the $S$, the less the generations are needed to reach M-S balance. This is mainly because as $S$ increases, the mean number of mutations accumulated in the genome necessary to reach M-S balance gets smaller, as is found in our simulations and those of D. Charlesworth et al. (1990). With an increasing $U$, the number of generations to reach M-S balance increases slightly (Figure 3c). This is because when $U$ increases, it takes more mutations accumulated in the genome and thus more generations to reach M-S balance. On the other hand, with an increasing $U$, mutations occur faster in the population, reducing the number of generations needed to reach M-S balance. These two effects cancel each other to a certain extent, causing the number of generations needed to reach M-S balance to increase only slightly with an increasing $U$. $h$ and $s$ have similar effects (Figure 3, b and d) in determining the number of generations needed to reach M-S balance. Small values of $h$ and $s$ will result in more generations needed to reach M-S balance. This is largely because selection against DGM is not effective with small $h$ and $s$ and more accumulated mutations are needed to reach M-S balance.

For populations at M-S balance, we found that, as
expected, h and s have small effects on mean of fitness and S and U have relatively large effects on mean of fitness (Figure 4). With increasing S and/or decreasing U, mean fitness at M-S balance increases. This is due to the increasing efficiency against DGM with an increasing S and mutation pressure decreasing with a decreasing U. For the genetic variance of fitness, increasing S and s and decreasing h result in larger genetic variance at M-S balance. Genetic variance at M-S balance generally increases with an increasing U when S is large (S = 0.8) and there is an intermediate maximum genetic variance with an intermediate U value when S is small (S = 0.3). This should not be surprising. As expected from the theory (Equations 1b and 10b, Deng and Lynch 1996), under complete selfing, the genetic variance at M-S balance is a monotonic function of U; under complete outcrossing, it is not and there is an intermediate maximum level of genetic variance with an intermediate level of U.

**DISCUSSION**

In this article, the robustness and the statistical properties of the Deng and Lynch (1996) method are extensively investigated under incomplete outcrossing/selfing and in nonequilibrium populations at different stages approaching M-S balance. It is demonstrated that when S < 0.1 or when S > 0.8, the Deng and Lynch method is generally robust and applicable to yield reasonably reliable estimates. For 0.1 < S < 0.8, relatively larger bias will result and the bias generally increases when S approaches 0.5. Under partial outcrossing and variable or epistatic mutation effects, depending on the unknown mutation effects, selfing rate, and mutation parameters (see results), the estimates of the mutation parameters may not always be biased. The Deng and Lynch estimation in nonequilibrium populations at different stages approaching M-S balance is fairly robust and is similar to the performance in populations at M-S balance. The underlying mechanisms are investigated and given. The results for only a limited number of mutation parameters are given in this article. Our results not shown for other mutation parameters (U = 0.1–10; s = 0.01–0.10; h = 0.2–0.4) demonstrated similar results and indicated that s and h also influence the estimation bias, although the magnitude is generally small relative to the effects of S and U. Our investigation here provides a basis for applying the Deng and Lynch method to characterize DGM in partial selfing or outcrossing populations and for nonequilibrium populations.

Our results on the robustness of the estimation in partial selfing/outcrossing populations are consistent with that of B. Charlesworth et al. (1990) in testing the robustness of their inbreeding depression method to estimate U. The estimation method of B. Charlesworth et al. (1990) only employs information of change in mean fitness traits upon outcrossing in selfing populations and depends on a prior knowledge of h that is generally not known (Deng and Fu 1998). By computer simulations, B. Charlesworth et al. (1990) found that their inbreeding depression method is fairly robust in populations with selfing rate S > 0.8. These results are consistent in that the estimation methods (B. Charlesworth et al. 1990; Deng and Lynch 1996) developed for complete outcrossing and selfing populations may also be applicable to partial selfing/outcrossing populations, provided that S > 0.8 in predominantly selfing populations or S < 0.1 in predominantly outcrossing populations. There are many populations that are partially outcrossing/selfing (Stebbins 1957; Lande and Schenske 1985; Schenske and Lande 1985; Barrett and Eckert 1990; Charlesworth 1992). In addition, S in many of these populations is either high and close to one or low and close to zero (Lande and Schenske 1985).

### TABLE 5

<table>
<thead>
<tr>
<th>Generation</th>
<th>U = 0.5</th>
<th>U = 1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \hat{U} )</td>
<td>( \hat{h} )</td>
</tr>
<tr>
<td>50</td>
<td>0.12 ± 0.01</td>
<td>0.218 ± 0.009</td>
</tr>
<tr>
<td>200</td>
<td>0.29 ± 0.02</td>
<td>0.219 ± 0.009</td>
</tr>
<tr>
<td>500</td>
<td>0.36 ± 0.02</td>
<td>0.219 ± 0.008</td>
</tr>
<tr>
<td>1000</td>
<td>0.36 ± 0.02</td>
<td>0.219 ± 0.008</td>
</tr>
<tr>
<td>Balance</td>
<td>0.36 ± 0.02</td>
<td>0.218 ± 0.008</td>
</tr>
<tr>
<td>25</td>
<td>0.23 ± 0.02</td>
<td>0.219 ± 0.014</td>
</tr>
<tr>
<td>75</td>
<td>0.38 ± 0.02</td>
<td>0.214 ± 0.012</td>
</tr>
<tr>
<td>150</td>
<td>0.41 ± 0.03</td>
<td>0.213 ± 0.011</td>
</tr>
<tr>
<td>300</td>
<td>0.42 ± 0.03</td>
<td>0.213 ± 0.012</td>
</tr>
<tr>
<td>Balance</td>
<td>0.42 ± 0.03</td>
<td>0.213 ± 0.012</td>
</tr>
</tbody>
</table>
Deleterious Mutation Rate and Effects

Figure 3.—The number of generations needed to reach M-S balance under different parameters of constant mutation effects. (a) The effect of selfing rate $S$; in simulations, $h = 0.36$ and $s = 0.03$. (b) The effect of $h$; in simulations, $U = 1.0$ and $s = 0.03$. (c) The effect of $U$; in simulations, $h = 0.36$ and $s = 0.03$. (d) The effect of $s$; in simulations, $U = 1.0$ and $h = 0.36$.

Figure 2.—Dynamics of $x$, $y$, $z$, $x/z$, $x/y$, and $z/y$ with the number of generations experienced under mutation and selection when starting from homogeneous populations free of DGM under different parameters. Mutation effects are constant. Solid lines, $U = 0.5$; dashed lines, $U = 1.0$. Longer lines, $S = 0$; shorter lines, $S = 1$. For d, e, and f, there are only two lines, because for the same selfing rate $S$, different $U$'s have almost the same lines.
fectiveness of selection against DGM. These can be further demonstrated by a theoretical approximation. By the work of Lande and Schlomke (1985), for partial selfing population at M-S balance, consider a locus with two alleles (wild-type B and mutant allele b), and the frequencies of B and b are, respectively,

\[ p = 1 - q, \quad q = \mu / \mathbb{E}[f + (1 - f)h] \]

where \( \mu \) is the mutation rate from B to b per locus and f is Wright's inbreeding coefficient (Lynch and Walsh 1998). The frequencies of genotypes BB, Bb, and bb are, respectively,

\[ D = fp + (1 - f) p^2, \quad 2H = 2(1 - f)pq, \]
\[ R = fq + (1 - f)q^2. \]

Ignoring the term \( q^2 \) and letting \( f \equiv S / (2 - S + 2s) \) (for relatively large S so that \( f > h \); Lande and Schlomke 1985), the numbers of heterozygous and homozygous mutations per individual are then, respectively,

\[ n_{het} \approx U \cdot \frac{2(1 - S + Ss)}{s[S + 2(1 - S + Ss)h]}, \]
\[ n_{hom} \approx U \cdot \frac{S}{2} \cdot \frac{s[S + 2(1 - S + Ss)h]}. \]

From the above equations, the qualitative effects of S, U, h, and s on the number of mutations per individual in partial selfing/outcrossing populations at M-S balance can be easily seen.

In this article, for nonequilibrium populations at various stages approaching M-S balance, we examine the dynamics of y, z, and \( \pi \) and their ratios, all of which are functions of the mean and genetic variance of fitness in populations. These detailed examinations provide an explanation for the surprisingly robust performance of the Deng and Lynch estimation in nonequilibrium populations. Although we did not formally investigate the performance of the inbreeding depression approximation (Morton et al. 1956; Charlesworth et al. 1990) in our simulated nonequilibrium populations, their performance may not be good. This was suggested by the fact that the inbreeding depression \( f \) alone changes dramatically before M-S balance is reached (Figure 2). The robust performance of the Deng and Lynch (1996) estimation lies in the fact that it employs changes of mean and genetic variance of fitness. The ratios of the mean and genetic variance that are critical for the Deng-Lynch estimation remain fairly stable in nonequilibrium populations at different stages approaching M-S balance (Figure 2). Generally, it requires only \( \sim 200 \) or \( \sim 75 \) generations, respectively, in complete outcrossing and selfing populations for these ratios to be stable since the populations experience mutation and selection starting from a homogeneous population free of DGM. In partial selfing/outcrossing populations, the number of generations needed is between \( \sim 200 \) and \( \sim 75 \) generations.

Most natural populations have experienced mutation and selection for many generations, and they were generally founded by individuals with different numbers of DGM. It can be shown easily by computer simulations (H.-W. Deng and J. Li, unpublished results) that, despite initially different distributions of DGM in founding populations, under mutation and selection, the population dynamics of fitness will quickly (in less than a few dozen generations) converge to that (under the same parameters) of populations approaching M-S balance when starting free of DGM. The specific merging point into the dynamics investigated here will largely depend on the detailed initial distributions of DGM in founding populations. Such founding populations may be formed due to migration, population admixture, and population bottlenecks. Therefore, the investigation for nonequilibrium populations approaching M-S balance when starting from a homogeneous population is of general significance in shedding light on the robustness of the Deng-Lynch method in characterizing DGM. Particularly, even starting from a homogeneous population free of DGM, the number of generations taken to reach M-S balance does not seem to be large for natural populations (Figure 3), and generally 400–2000 generations are needed. Therefore, the Deng and Lynch (1996) method should be robust and applicable for many natural nonequilibrium populations even if they are at different stages of approaching M-S balance. In addition, it is not rare for natural populations to be founded by one or a few genotypes (such as in Daphnia or in many plant populations). Therefore, our investigation of nonequilibrium populations approaching M-S balance at various stages when starting from homogeneous genotypes is not only of theoretical but also of considerable practical significance in simulating plausible situations in natural populations.

In studying inbreeding depression and mutation load in partial selfing populations, Charlesworth et al. (1990) only described some effects of mutation parameters and S on the mean fitness. However, as pointed out by Deng and Lynch (1997), mutation load is manifested not only as the mean but also the genetic variance of populations at M-S balance. Studying the dynamics of genetic variance will reveal more aspects and important information about DGM and population evolution (Deng and Lynch 1997). In this study, we demonstrate the dependency of mean and genetic variance at M-S balance on mutation parameters and S. As stated earlier, we also demonstrated the detailed dynamics of mean and genetic variances and their ratios in populations approaching M-S balance and the number of generations taken to reach M-S balance when starting from a homogeneous population free of DGM. These results should be useful not only for the characterization of DGM, but importantly, they should also be useful for a variety of other population genetics topics. For example, the dynamics of mean and genetic variance and the
Figure 4.—Mean and genetic variance of fitness for populations at M-S equilibrium under different selfing rates, S, with different parameters of constant mutation effects. (a, c, and e) The equilibrium mean of fitness under various U, h, and s, respectively. (b, d, and f) The genetic variance of fitness under various U, h, and s, respectively. In simulations for a and b, h = 0.36 and s = 0.03. In simulations for c and d, U = 1.0 and s = 0.03. In simulations for e and f, U = 1.0 and h = 0.36.

Characterization of DGM is one major challenge in the broad field of genetics (Crow 1993a,b; Bridges et al. 1994; Peck and Eyre-Walker 1997). Several ap-

number of generations taken to reach M-S balance when starting from a homogeneous population free of DGM are important for conservation genetics.
proaches have been developed and data are being accumulated. However, before the results from any approach can be treated as robust, the statistical estimation methods underlying all the approaches must be examined carefully under a variety of biologically plausible conditions. This article represents one of a series of our efforts (Deng and Lynch 1996; Deng 1998b; Li et al. 1999) to examine the robustness and statistical properties of Deng and Lynch's (1996) method for characterization of DGM in natural populations. Our effort in the near future (H.-W. Deng and J. Li, unpublished results) will examine the estimation under other conditions such as finite population sizes and linkage disequilibrium between DGM at different loci.

We are particularly grateful for the constructive comments from an anonymous reviewer who helped to improve the manuscript. This work was partially supported by a program grant and a faculty development grant from the Health Future Foundation, a grant from Hunan Normal University of People's Republic of China, and by a graduate student tuition waiver to J. Li from the Department of Biomedical Sciences of Creighton University.

LITERATURE CITED
Ritland, K., 1984 The effective proportion of self-fertilization with consanguineous matings in inbred populations. Genetics 106: 139-152.

Communicating editor: M. Slatkin