Genet. Res., Camb. (1999), **73**, *pp.* 147–164. *With 7 figures.* Printed in the United Kingdom

On the experimental design and data analysis of mutation accumulation experiments

HONG-WEN DENG*, JIAN LI AND JIN-LONG LI

Osteoporosis Research Center and Department of Biomedical Sciences, Creighton University, 601 N. 30th Street, Suite 6787, Omaha, NE 68131, USA

(*Received 14 April 1998 and in revised form 10 July 1998 and 10 September 1998*)

Summary

Characterizing deleterious genomic mutations is important. Most of the few current estimates come from the mutation–accumulation (M-A) approach, which has been extremely time- and labour-consuming. There is a resurgent interest in implementing this approach. However, its estimation properties under different experimental designs are poorly understood. By simulations we investigate these issues in detail. We found that many of the previous M-A experiments could have been more efficiently implemented with much less time and expense while still achieving the same estimation accuracy. If more than 100 lines are employed in M-A and if each line is replicated at least 10 times during each assay, an experiment of 10 M-A generations with two assays (at the beginning and at the end of M-A) may achieve at least the same estimation quality as a typical M-A experiment. The number of replicates per M-A line necessary for each assay largely depends on the magnitude of environmental variance. While 10 replicates are reasonable for assaying most fitness traits, many more are needed for viability, which has an exceptionally large environmental variance. The investigation is mainly carried out using Bateman–Mukai's method of moments for estimation. Estimation using Keightley's maximum likelihood is also investigated and discussed. These results should not only be useful for planning efficient M-A experiments, but also may help empiricists in deciding to adopt the M-A approach with manageable labour, time and resources.

1. Introduction

Some essential parameters of deleterious genomic mutations are: (1) the genomic mutation rate (U, the number of new mutation occurrences per genome per generation), (2) the mean selection coefficient ($\bar{s}$, the relative reduction in performance of mutant homozygotes relative to the wild-type homozygotes), (3) the mean dominance coefficient ($\bar{h}$, the extent to which heterozygotes express harmful effects of mutant alleles) and (4) the amount of variation in mutational effects.

Estimating these deleterious genomic mutation parameters in diverse taxa is essential for our understanding of many fundamental biological phenomena, for correctly assessing the overall risk for human health exerted by deleterious mutations, and for the continuing survival of other organisms (especially rare and endangered species). For example, U estimates are crucial in testing hypotheses on the evolution of sex and recombination (Muller, 1964; Kondrashov, 1985, 1988; Charlesworth, 1990), mate choice (Kirkpatrick & Ryan, 1991), diploidy (Kondrashov & Crow, 1991) and outbreeding mechanisms (Charlesworth & Charlesworth, 1987). They also determine the magnitude of mutation load in populations at equilibrium (Holdane, 1937; Kimura *et al.*, 1963; Burger & Hofbauer, 1994). The values of U and s are critical in determining the role of deleterious mutations in the extinction of small populations through the accumulation of mutations (Lande, 1994; Lynch *et al.*, 1995, 1996). The values of U, h and s determine the rate of input of genetic variance from mutation per generation (Deng & Lynch, 1996, 1997) and the extent to which neutral molecular variation is reduced due to background selection (Charlesworth *et al.*, 1993, 1995; Hudson & Kaplan, 1995). In finite populations variation of

* Corresponding author. Telephone (402): +280 5911. Fax: +1 (402) 280 5034. e-mail: deng@creighton.edu.

mutational effects plays an important role in the maintenance of polygenic variations (Keightley & Hill, 1990) and in determining the persistence time and extinction rate of small populations (Lande, 1994; Lynch *et al*., 1993, 1995*a*, *b*).

There are now three approaches to the estimation of these parameters:

(1) *The mutation-accumulation* (*M-A*) *approach*. This was proposed by Bateman (1959) and first employed by Mukai (1964) and Mukai *et al*. (1972). This technique estimates U and $\bar{s}$. Most estimates have come from this approach applied to *Drosophila melanogaster* (Mukai, 1979; Crow & Simmons, 1983; Keightley, 1994, 1996) and have been very hard to acquire, requiring large and long-term M-A and special chromosomal constructs or inbred/asexual lines. The data from M-A can also be analysed by the maximum likelihood method (Keightley, 1994) or the minimum distance method (Garcia-Dorado, 1997). Given the importance, there is a resurgent interest in implementing M-A experiments (Houle *et al*., 1994; Kibota & Lynch, 1996; Keightley & Caballero, 1997; Shabalina *et al*., 1998).

(2) *The inbreeding depression approach*. This was implied by Morton *et al*. (1956) in outcrossing populations, and explicitly proposed by Charlesworth *et al*. (1990) for use with highly selfing populations. Given its dependence on $\bar{h}$ that currently cannot be estimated without bias (Deng, 1998*a*), this technique *per se* estimates U only. In the highly selfing annual plants *Leavenworthia* (Charlesworth *et al*., 1994) from *Amsinckia* (Johnston & Schoen, 1995), U estimates from this approach are in line with earlier ones from the M-A in *Drosophila*.

(3) *The fitness moments approach*. This was developed by Deng & Lynch (1996, 1997). It estimates U, h and s. For two outcrossing species of cyclical parthenogenetic *Daphnia* (a freshwater microcrustacean) the preliminary data estimated by this approach generally agree with earlier ones from other species (Deng & Lynch, 1997) and also are generally compatible with earlier (Lynch, 1985) and recent (Lynch *et al*., 1998) data from the direct M-A approach in *Daphnia*.

Except for the third approach, issues of experimental design and statistical properties (bias and sampling variances) have not been formally investigated until rather recently (Deng & Fu, 1998). Such investigations, especially those under realistic biological situations, are important, since they will not only provide a basis for the correct interpretation of estimates obtained under some necessary but unrealistic assumptions, but also offer some practical guidelines for optimally designing and efficiently implementing different approaches. Under the assumption that genotypic values can be measured accurately (as is the case when each genotype is replicated many times for assay), Deng & Fu (1998) investigated and compared the three approaches under a range of parameter space and biologically plausible situations. For the M-A approach three surprising results emerged, which may potentially have significant implications for designing M-A experiments for empiricists (Deng & Fu, 1998):

(1) If sufficiently many (> 100) M-A lines are employed, estimation based on experiments of only 10 M-A generations is comparable (in terms of bias and sampling variance) to that based on experiments of 40 M-A generations.

(2) Increasing the number of assays during M-A experiments does not improve estimation quality (in terms of bias and sampling variance) very much but incurs much elevated costs, so that only two assays (one at the beginning and the other at the end of M-A) may be performed instead of multiple assays.

(3) If mutational effects are multiplicative across loci additivity is assumed, increasing the number of M-A generations results in progressively larger bias under constant effects and progressively smaller bias than predicted under variable effects (figure 2 in Deng & Fu, 1998). An important implication of this result is that, if fitness effects are variable and do act multiplicatively, the common practice (by assuming an exponential distribution of mutational effects) of inferring U to be twice the $\hat{U}$ values obtained from M-A (Mukai *et al*., 1972; Lynch *et al*., 1995) overestimates U. Throughout, the circumflex (ˆ) indicates an estimated value. The degree of overestimation increases with the number of M-A generations (figure 2 in Deng & Fu, 1998). This result concurs with the result 1 in suggesting that increasing M-A generations not only may be unnecessary but also may adversely affect the correct inference of mutation parameters.

Together, these results suggest that the designs employed by most of the earlier M-A experiments (e.g. Mukai *et al*., 1972; Ohnishi, 1977; Houle *et al*., 1994; Kibota & Lynch, 1996; Keightley & Caballero, 1997; Shabalina *et al*., 1998) may not be cost- or time-efficient. In these experiments usually fewer than 100 (most often about 50 except in Ohnishi, 1977) M-A lines were employed, M-A proceeded for over 40 generations, multiple assays (usually ~ 5) were performed and a relatively small number (~ 5 or less) of replicates were measured for each M-A line during each assay. The results of Deng & Fu (1998) also imply that, if experiments are designed properly, deleterious genomic mutations may be characterized much less expensively with M-A experiments on much smaller scales. Partially due to the fact that the previous experiments were tremendously time- and labour-consuming, the M-A approach is not widely applied. However, since the main purpose of Deng & Fu (1998) was to compare the relative efficiencies of the three currently available estimation approaches,

https://doi.org/10.1017/S0016672398003681 Published online by Cambridge University Press

genotypic values were assumed to be measured with little error. This assumption is applied to results 1 and 2 (Deng & Fu, 1998), while result 3 should be valid without this assumption. It is well known that genotypic values of quantitative traits are usually measured inaccurately due to random development instability, measurement errors and environmental effects (Falconer & Mackay, 1996). To provide useful and practical guidelines for empiricists to design efficient M-A experiments, the assumption that genotypic values are measured with little error must be examined closely.

Under the realistic situation that genotype values are usually measured with error, this study aims (1) to test the robustness of the above results 1 and 2 from Deng & Fu (1998); and (2) to investigate the statistical properties and experimental designs of the M-A approach thoroughly under a range of parameter values. The investigation is mainly carried out by using the Bateman–Mukai method of moments. Estimation using Keightley's maximum likelihood is also investigated and discussed. We conduct our investigation under both constant and biologically plausible variable mutational effects across loci. Due to the complex nature of the estimation procedure of the M-A approach, an analytical approach for fulfilling the above goals is difficult, if not entirely impossible. Computer simulations were thus adopted to accomplish our goals.

2. Simulations

(i) *Constant mutational effects*

To focus on the experimental design issues, the simulation simplifies some complex features of the actual M-A experiments (e.g. the need for raising large controls for temporal environmental changes, replacement of inadvertent or selective line losses and backups in case of line losses). These complex features may differ from one specific experiment to another, and it is almost impossible to simulate them in a meaningful and general manner. Thus we simulate the most essential processes of M-A and phenotype assays that are common to all M-A experiments. The simulation proceeds as follows:

(1) Genotypic values of the fitnesses for L genetically identical lines are set to 1·0 at the onset of the M-A.

(2) As in the *Drosophila* experiments (Mukai *et al.*, 1972), mutations are allowed to accumulate in the heterozygous state but are measured when homozygous. Therefore, mutation parameters simulated are for a special chromosome under M-A. M-A employing inbred or asexual lines is similar (though the mutation parameters, especially those for U, are different) and can be easily implemented; the results are similar and thus will not be presented in this study. Mutations, with constant effects s and h, occur in each line as a Poisson process at a rate of U per generation, and are allowed to accumulate independently for T generations. Fitness (W) equals $1-s*n$, were n is the number of mutations accumulated.

(3) All lines are assayed at time intervals of I generations, with the first assay conducted at the 0th M-A generation. At each assay generation, each of the L M-A lines is replicated (i.e. genotypes are cloned) into R sublines for fitness measurement. As is done in practice, mutations that may occur during one generation of propagating M-A lines into sublines are ignored. If the genotypic value of the ith ($i = 1, 2, \ldots, L$) M-A line at assay is W_i, the phenotypic value of fitness of its jth ($j = 1, 2, \ldots, R$) subline is determined by a random sampling from a normal distribution with mean W_i and variance $\sigma_e^2.\sigma_e^2$ accounts for environmental variance.

(4) The mean and genetic variance of fitness at each assay are computed. The genetic variance is computed by one-way ANOVA, with M-A lines as main effects and sublines as random effects.

(5) As is the case in practice (Mukai *et al.*, 1972), the means and genetic variances of fitness among the M-A lines measured at different times are used in regression analyses to estimate the rates of change of mean fitness (M) and genetic variance (V) due to genomic mutations.

(6) With estimates of M and V, bounds on U and $\bar{s}$ can be estimated (Bateman, 1959; Mukai *et al.*, 1972). Assuming that fitness effects are additive, the mutation probability per generation at each locus is small, and mutations on all loci are independent, we have:

$$U = \frac{M^2(1+CV_s^2)}{V}, \quad \bar{s} = \frac{V}{M(1+CV_s^2)}. \tag{1a}$$

Since the squared coefficient of variation of s_i across loci ($CV_s^2 = V_s/\bar{s}^2$) is always greater than 0, we have:

$$U \geqslant \frac{M^2}{V}, \quad \bar{s} \leqslant \frac{V}{M}. \tag{1b}$$

It should be noted that in M-A experiments involving special chromosomal constructs in *Drosophila*, U is typically the haploid mutation rate for the chromosome concerned (Mukai *et al.*, 1972). In M-A experiments involving asexual lines such as in *Daphnia* (Lynch *et al.*, 1998), U is typically the mutation rate for the whole genome.

The simulation is performed for a range of parameter sets. For U and s the simulated parameter sets cover the range of the well-known values for the M-A chromosome in Mukai *et al.* (1972; $U \cong 0{\cdot}15$ and $\bar{s} \cong 0{\cdot}036$). For example, U ranges from 0·05 to 0·20 and $\bar{s}$ from 0·02 to 0·06. Determination of the

https://doi.org/10.1017/S0016672398003681 Published online by Cambridge University Press

magnitude of the environmental variance (σ_e^2) employed in simulations depends on the mutational variance (V_m, the rate of input of genetic variance from mutation per generation). It has been estimated from extensive data (Lynch, 1988; Houle *et al.*, 1994; Fernandez & Lopez-Fanjul, 1996; Houle *et al.*, 1996; Deng & Lynch, 1997) that, for most fitness traits, $V_m \cong 10^{-3}\sigma_e^2$. In simulations, unless otherwise specified, we set $\sigma_e^2 = 10^3\, V_m$. σ_e^2 reflects all error sources for genotypic value measurements, such as developmental instability and measurement error. We examined several cases in which σ_e^2 ranges from $0{\cdot}025{*}10^{-3}V_m$ to $2{\cdot}0{*}10^3 V_m$. V_m is determined by the other deleterious mutation parameters. In outcrossing populations under constant mutational effects (Deng & Lynch 1996):

$$V_m = U(hs)^2. \tag{2}$$

Throughout, unless otherwise specified, each data point represents a simulation result for one particular parameter set and is obtained by 1000 repeated random simulations.

(ii) *Variable mutational effects*

It seems clear that s_i and h_i are not constant across loci, and the few available data suggest that s_i for mutations across loci has a roughly leptokurtic distribution (Gregory, 1965; Mackay *et al.*, 1992; Santiago *et al.*, 1992; Keightley, 1994; Keightley & Caballero, 1997). As in Deng & Lynch (1996, 1997), we model the distribution ($f(s_i)$) of mutational effects s_i as:

$$f(s_i) = \frac{1}{\bar{s}} \exp(-s_i/\bar{s}), \quad (1 > s_i > 0) \tag{3}$$

where $\bar{s}^2 = V_s$ (variance of s_i). This is an exponential distribution, which is a special form of the gamma distribution modelled by Keightley (1994). We adopt it as a special alternative to the constant effects. It is used for modelling non-lethal mutations, since lethal mutations appear to represent a true discontinuity in the distribution of mutant effects (Crow & Simmons, 1983). The effect of lethals on estimation can be minimized by eliminating extreme lines as is done in practice (Mukai *et al.*, 1972). Little information exists on the distribution of h_i. Data from *Drosophila* (Crow & Simmons, 1983) and biochemical arguments suggest an inverse relationship between s_i and h_i, mutant alleles with larger effects tending to be more recessive (Kacser & Burns, 1981). Therefore, as in Deng & Lynch (1996, 1997) and Deng & Fu (1998), we approximate this relationship as:

$$h_i = \tfrac{1}{2}\exp(-13s_i). \tag{4}$$

Note that in Equation 4, $\bar{h} = 0{\cdot}36$ when $\bar{s} = 0{\cdot}03$, $h \to 0{\cdot}5$ as $s \to 0$, and $h \to 0{\cdot}0$ as $s \to 1{\cdot}0$, all in rough accordance with the data from *Drosophila* (Crow & Simmons, 1983). However, true mutational spectra may be such that the dominance of individual mutations is broadly scattered around such a function (Caballero & Keightley, 1994).

Under variable mutational effects all aspects of the simulation process are the same as in the earlier constant mutational effects, except that (1) s_i for each new mutation is not constant and is randomly drawn from an exponential distribution (eqn (3)) with mean $\bar{s}$, and (2) V_m is given by

$$V_m = \int U(sh(s))^2 f(s)ds.$$

With the exponential distribution of s_i (eqn (3)) and the above function for $h(s)$ (eqn (4)), it can be easily shown that (Deng & Lynch 1997),

$$V_m = \frac{U\bar{s}^2}{2(26\bar{s}+1)^3} \tag{5}$$

As for constant mutational effects, the simulation is performed for a range of parameter values.

Fitness traits which vary on continuous scales, such as fecundity, intrinsic rate of increased and productivity, are the focus of our simulation studies since they are employed by most of the previous M-A experiments (Lynch, 1985; Houle *et al.*, 1994; Kibota & Lynch, 1996; Keightley & Caballero, 1997; Shabalina *et al.*, 1998). For these fitness traits it has been estimated by the extensive data that σ_e^2 ranges from $10^2 V_m$ to $10^4 V_m$, with an average of $\sim 10^3 V_m$ (Lynch, 1985, 1988; Houle *et al.*, 1994; Fernandez & Lopez-Fanjul, 1996; Houle *et al.*, 1996; Deng & Lynch, 1997). An important fitness trait – viability – was also used in M-A experiments (Mukai, 1964, 1969; Mukai *et al.*, 1972). Individual viability can only take discrete values of either 0 (dead) or 1 (live). However, in M-A experiments for viability (Mukai *et al.*, 1972) it was not the individual viability that was employed in analyses. Instead, an index was constructed that was based on the ratio of the viability data from emerging homozygous and heterozygous individuals within each replicate vial for each M-A line (Mukai *et al.*, 1972). This index varies on a continuous scale from 0 to 1 as any other fitness traits simulated here. Compared with other fitness traits, σ_e^2 for viability is generally higher due to additional binomial sampling (Lynch, 1988). Therefore, in simulations of viability, $\sigma_e^2 = 2{\cdot}0 \times 10^4 V_m$. This value correctly reflects the average of V_m ($5{\cdot}0 \times 10^{-4}\,\sigma_e^2$), whose range is 10^{-4}–$10^{-5}\,\sigma_e^2$ as estimated in *Drosophila* (Lynch, 1988).

https://doi.org/10.1017/S0016672398003681 Published online by Cambridge University Press

3. Results

Data not shown reveal that there is no estimation bias under constant mutational effects. Under the exponentially variable mutational effects (equation 3), $\hat{U}$ is not significantly different from $0{\cdot}5U$, neither is $\hat{\bar{s}}$ *from* $2\bar{s}$. Throughout, '^' indicates an estimated value. These results concerning bias are not unexpected (Equation 1; Mukai *et al*., 1972; Deng & Fu, 1998). Thus, for constant mutational effects, only the

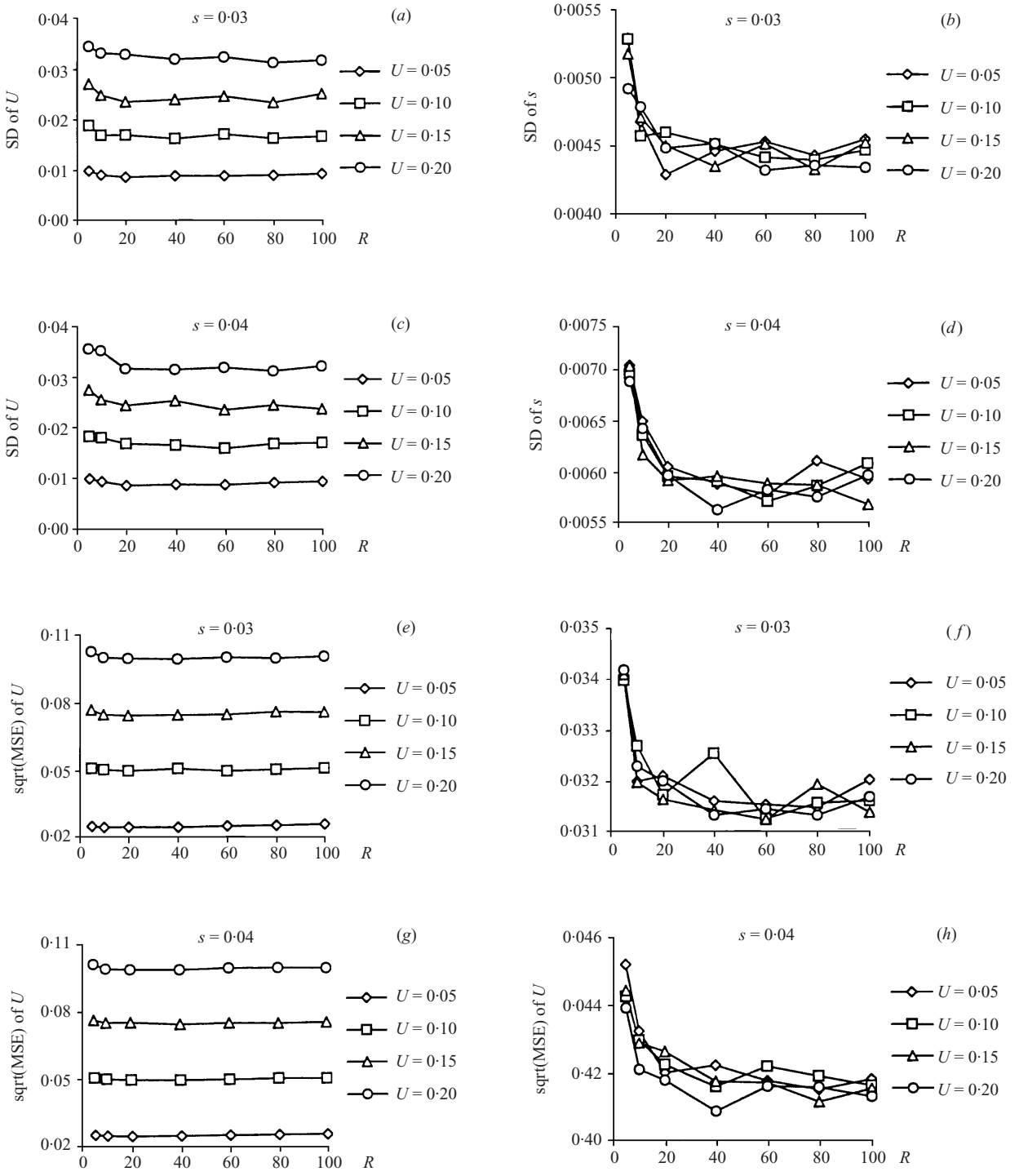

Fig. 1. The effects of different number of replicates (R) on estimation under constant (plots *a–d*) and variable mutation effects (plots *e–h*). The parameters common to all simulations for this figure are, respectively, T (the total number of M-A generations) = 40, I (the number of interval generations between each assay) = 10, L (the number of M-A lines) = 100. In all simulations, unless otherwise specified, $\sigma_e^2 = 10^3 V_m$.

https://doi.org/10.1017/S0016672398003681 Published online by Cambridge University Press

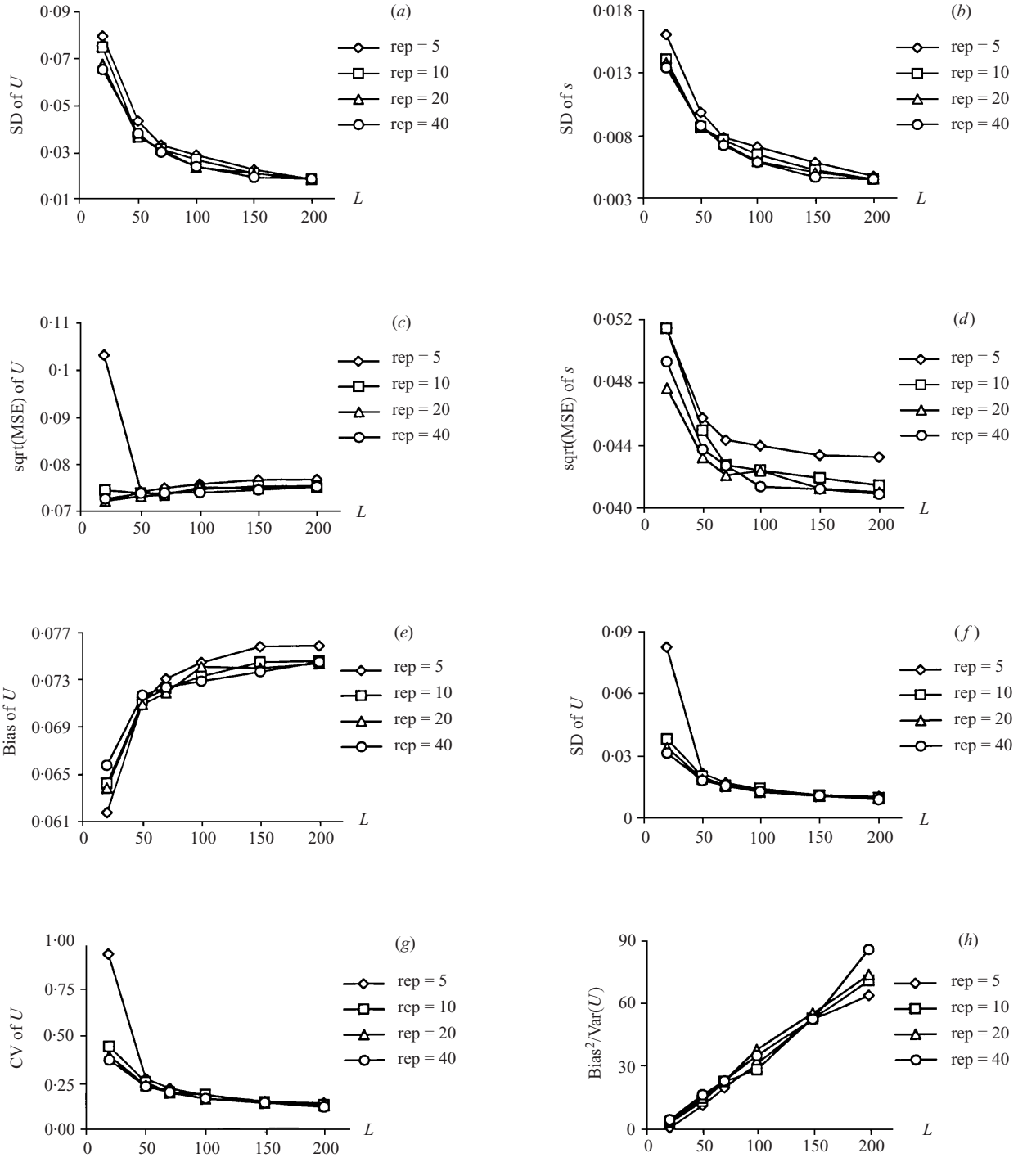

Fig. 2. The effects of different number of M-A lines (L) on estimation under constant (plots *a–b*) and variable mutation effects (plots *c–h*). $t = 40$, $I = 10$, $\bar{s} = 0{\cdot}04$, $U = 0{\cdot}15$. CV denotes the coefficient of variation. The bias of U is $(U - \hat{U})$. The expected theoretical bias of $\hat{U}$ in our simulation $U/2$ (0·075).

sampling variance as measured by one standard deviation (SD) of the estimates over the 1000 repeated simulations is presented and compared under different simulated parameters. SD is an appropriate measure for sampling properties here, since the estimates over the repeated simulations conform to normal distributions (Kolmogorov–Smirnov test, $P > 0{\cdot}50$; Sokal & Rohlf, 1995). For variable mutational effects, since all the estimates are biased, we compute their MSE (mean square error) for comparison:

$$\text{MSE} = E(\hat{x} - E(x))^2 = Var(\hat{x}) + (\bar{\hat{x}} - E(x))^2 \qquad (6)$$

where $\bar{\hat{x}}$ stands for an estimated mean. Note that if $\bar{\hat{x}}$ is unbiased, MSE is simply the variance of $\hat{x}$.

https://doi.org/10.1017/S0016672398003681 Published online by Cambridge University Press

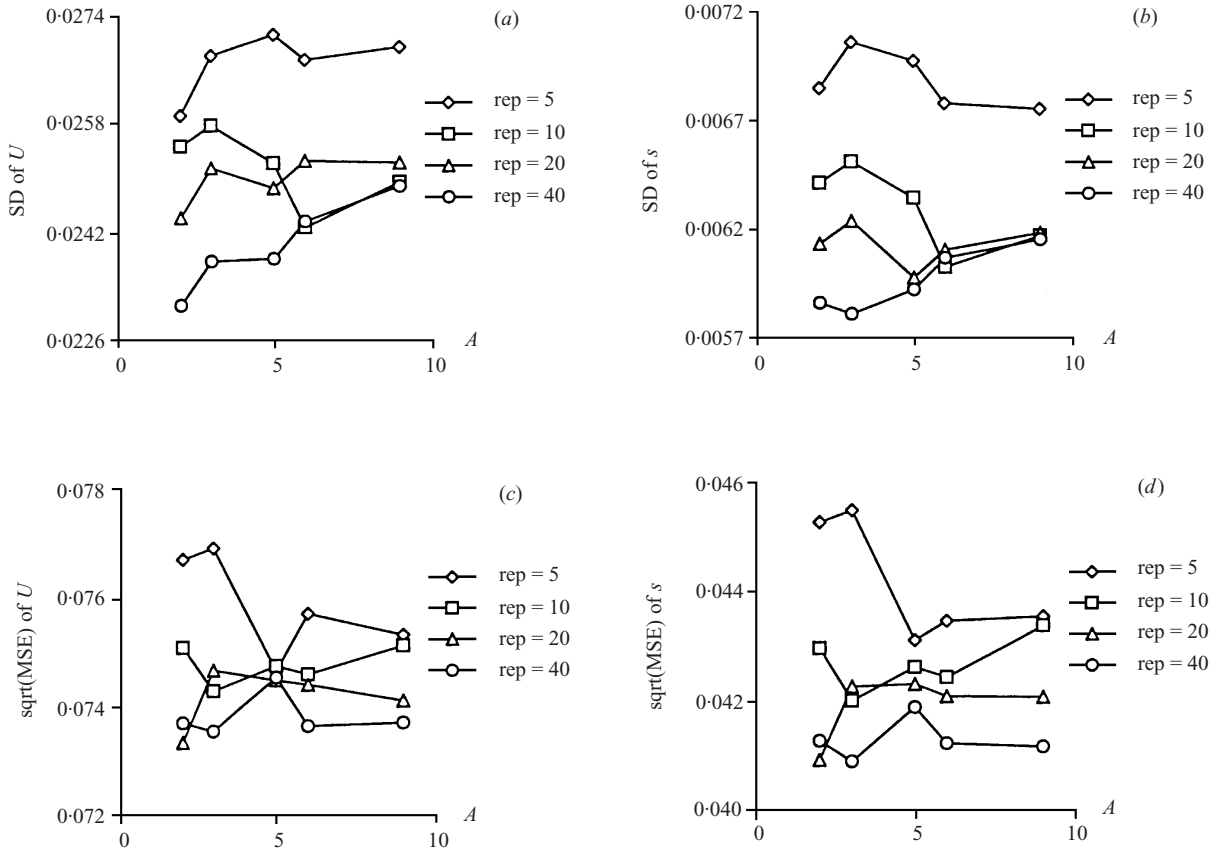

Fig. 3. The effects of different number of assays (A) on estimation under constant (plots a, b) and variable (plots c, d) mutation effects. $T = 40$, $L = 100$, $I = 10$, $R = 10$, $\bar{s} = 0{\cdot}04$, $U = 0{\cdot}15$.

Simulation results are presented in Figs. 1–6 and Tables 1–3. Simulation parameters specific to each data point are specified on each plot in Figs. 1–6, and those common parameters for all data points in each figure are specified in the legend to the respective figures. In the following, the simulation results will be summarized with respect to the effects of different design factors of M-A experiments on estimation.

(i) *The effects of different number of replicates* (R) *on estimation* (Fig. 1)

Increasing R for each M-A genotype during each assay generally increases the estimation quality. For the wide range of parameter sets simulated, SD generally decreases under constant mutational effects, as does MSE under variable mutational effects. This is true for the estimation of both U and $\bar{s}$. However, the effect of increasing R has larger effects on the estimation of $\bar{s}$ than on estimation of U. For both constant and variable mutational effects the estimation quality increases most dramatically when R increases from 5 to 10. Still better estimation can be achieved if R is increased from 10 to 20, particularly under constant mutational effects and for $\bar{s}$ estimation. After R is increased to 20, relatively small improvement in estimation can be gained. Increasing R has larger effects on estimation under constant than under variable mutational effects.

(ii) *The effects of different number of M-A lines* (L) *on estimation* (Fig. 2)

Increasing L improves estimation of $\bar{s}$ under both constant and variable mutational effects (plots b and d), and estimation of U under constant mutational effects (plot a), as reflected by the decreasing SD or MSE. These are not unexpected. However, counter-intuitively, with an increasing L the MSE for U under variable mutational effects generally increases (except when $R = 5$ and L increases from 20 to 50), though only very slightly (plot c), as reflected by the scale of the y-axis. For the data in plot c, we analysed in detail (plots e–h) the SD, bias ($U - \hat{U}$), coefficient of variation (CV) of $\hat{U}$, and the ratio of the squared bias to the variance of $\hat{U}$. It can be seen that, as L increases, the SD and CV of $\hat{U}$ decrease (as expected) (plots f and g). Although all the estimates are not significantly different from the theoretical expectation of $U/2$, the bias increases (though the magnitude is very small)

https://doi.org/10.1017/S0016672398003681 Published online by Cambridge University Press

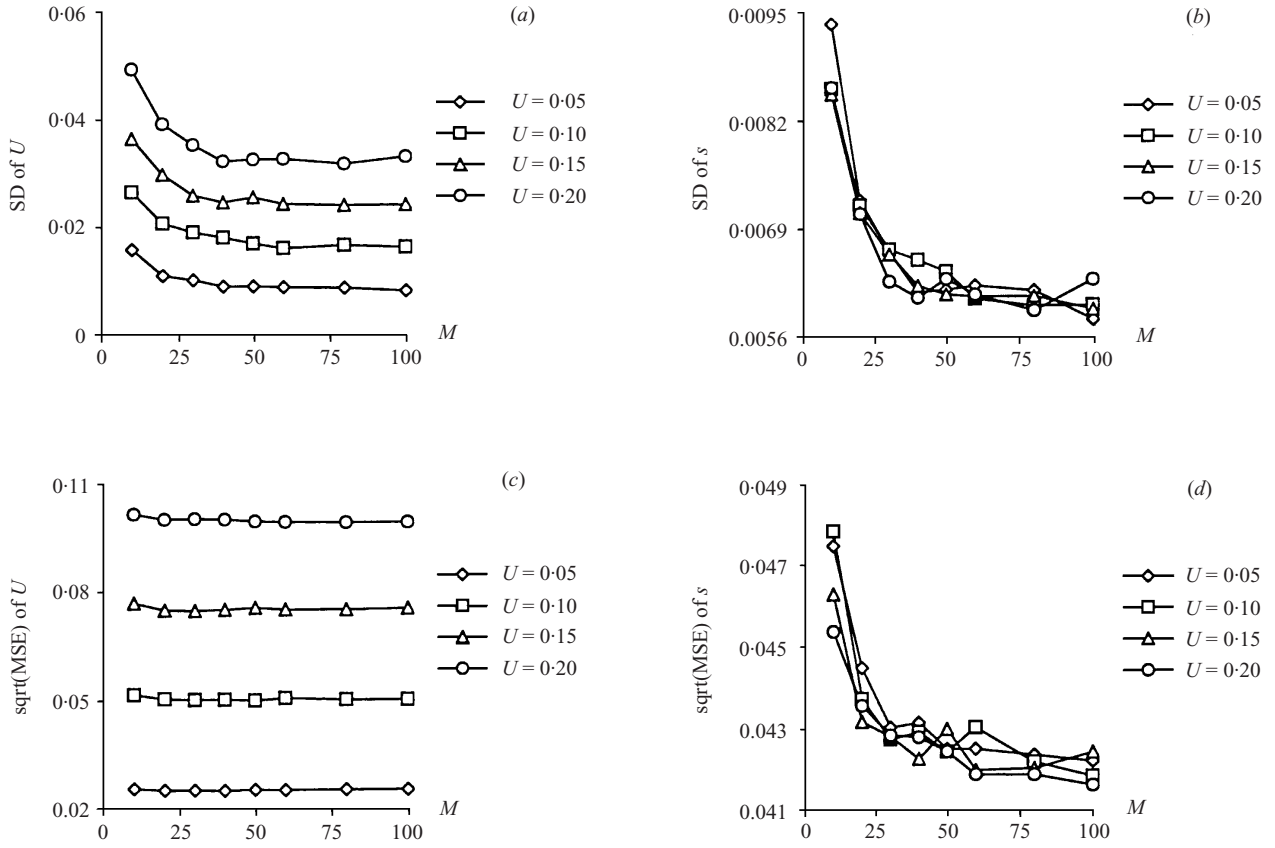

Fig. 4. The effects of different number of M-A generations (M) on estimation under constant (plots *a*, *b*) and variable mutation effects (plots *c*, *d*). $L = 100$, $R = 10$, $\bar{s} = 0{\cdot}04$, $U = 0{\cdot}15$; the number of assays (A) is 2.

and converges closer to the theoretical expectation of $U/2$ (which equals 0·75; plot *e*). Since the ratio of the bias² to the variance of $\hat{U}$ (plot *h*) increases, the change in MSE with increasing L due to the increase in bias² outweighs the decrease in sampling variance of $\hat{U}$. Therefore, MSE increases (though very slightly) with increasing L. However, the magnitude is miniscule, as can be noted from the scale of the *y*-axis (plot *c*). If R is very small (only 5), increasing L from 20 to 50 can greatly decrease MSE (plot *c*) under variable effects. In all cases, adding more M-A lines generally improves estimation more efficient when $L < 100$ than when $L > 100$.

(iii) *The effects of different number of assays* (A) *on estimation* (Fig. 3)

Increasing A generally does not change the estimates very much. Due to the random nature of the simulation process, the SD and MSE fluctuate with A, but the magnitude of fluctuation is small as reflected by the scales of the *y*-axis. Therefore, two assays (one at the beginning and the other at the end of the M-A experiment) can essentially achieve about the same estimation quality as multiple assays. This may be partially due to the autocorrelation structure of the M-A data over generations.

(iv) *The effects of different number of M-A generations* (M) *on estimation* (Fig. 4)

Increasing M improves estimation of $\hat{U}$ under constant effects (plot *a*) and estimation of $\bar{s}$ under both constant and variable effects (plots *b* and *d*). The improvement of estimation for U under variable mutational effects (plot *c*) is only slight when the true mutation rate is relatively high ($U = 0{\cdot}1$–$0{\cdot}2$) and almost negligible when U is relatively small ($U = 0{\cdot}05$). Analyses not shown indicate that, under variable mutational effects, the SD of $\hat{U}$ decreases with an increasing M, but the magnitude of bias increases and converges closer to the theoretical expectation of $U/2$ (though $\hat{U}$ is not significantly different from the theoretical expectation of $U/2$). These two trends cancel each other out to a certain extent and cause the relative inertia of estimation to the increasing M under variable effects. This situation is similar to that found for the effects of different L on estimation as presented earlier in detail (Fig. 2).

https://doi.org/10.1017/S0016672398003681 Published online by Cambridge University Press

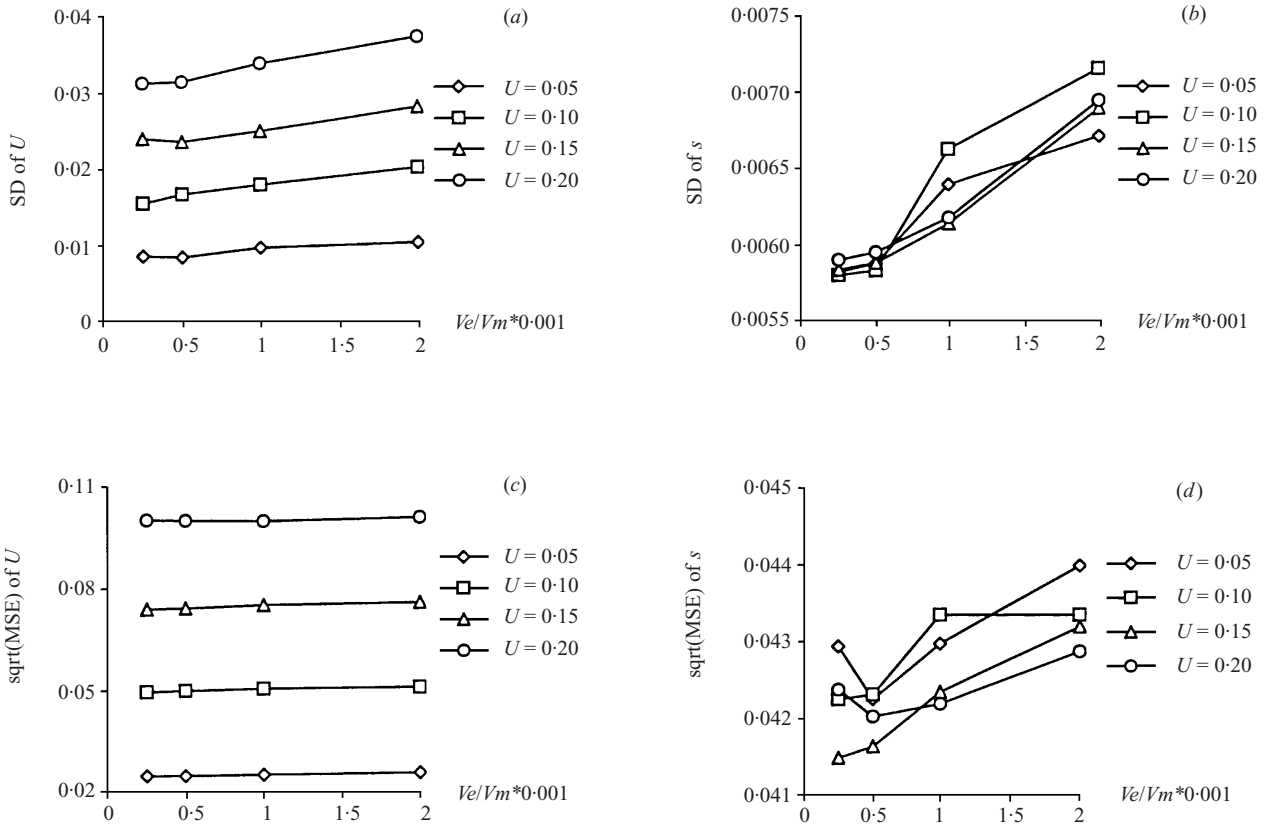

Fig. 5. The effects of σ_e^2 on estimation under constant (plots *a*, *b*) and variable mutation effects (plots *c*, *d*). $A = 2$, $L = 100$, $M = 40$, $R = 10$, $\bar{s} = 0{\cdot}04$.

Table 1. *Direct comparison of two M-A experimental designs*

Experimental designs	Constant		Variable		Total M-A	Total lines assayed	M-A generations
	$\hat{U}$	$\hat{s}$	$\hat{U}$	$\hat{\bar{s}}$			
A	0·161 (0·049)	0·040 (0·010)	0·080 (0·022) [0·073]	0·080 (0·019) [0·044]	2000	1250	40
B	0·159 (0·041)	0·0397 (0·009)	0·0800 (0·020) [0·073]	0·078 (0·017) [0·042]	1000	2000	10

Numbers within parentheses are SDs and those within square brackets are the square roots of MSE (for biased estimates only). Total M-A is equal to the product of the total numbers of M-A lines and M-A generations. In design A, $M = 40$, $L = 50$, $R = 5$, $I = 10$ generations (i.e. assays are performed every 10 generations, so there is a total of five assays including the one at the beginning of the M-A). In design B, $M = 10$, $L = 100$, $R = 10$, $I = 10$ generations (i.e. assays are performed every 10 generations, so there is a total of two assays). In simulations, $U = 0{\cdot}15$, $\bar{s} = 0{\cdot}04$, $\bar{h} = 0{\cdot}21$ and $\sigma_e^2 = 10^3 V_m$, which are close to the estimates by Mukai *et al.* (tables 5 and 7 of Mukai *et al.*, 1972) for an M-A chromosome.

(v) *The effects of normal environmental variance* (σ_e^2) *on estimation* (Fig. 5)

The magnitude of σ_e^2 is measured relative to the mutational variance (V_m). Increasing the magnitude of σ_e^2 decreases estimation quality for U under constant mutational effects (plot *a*) and $\bar{s}$ (plots *b* and *d*). However, the rate of change of estimation quality is relatively small (especially for U) and the trend is roughly linear (except for that due to the random simulation process). Under variable effects the estimation of U changes little with different σ_e^2/V_m. Again, detailed analyses not shown indicate that, under variable effects, with increasing σ_e^2/V_m the SD

https://doi.org/10.1017/S0016672398003681 Published online by Cambridge University Press

Table 2. *The effects of line loss due to high mutation pressure in M-A experiments for two experimental designs*

Experimental designs	Constant				Variable			
	$\hat{U}$		$\hat{s}$		$\hat{U}$		$\hat{s}$	
	NLL	LL	NLL	LL	NLL	LL	NLL	LL
A	0·31 (0·05)	0·35 (0·06)	0·040 (0·007)	0·034 (0·005)	0·153 [0·149]	0·207 [0·098]	0·080 [0·043]	0·055 [0·017]
B	0·32 (0·08)	0·32 (0·08)	0·040 (0·009)	0·040 (0·009)	0·158 [0·146]	0·159 [0·145]	0·079 [0·043]	0·079 [0·042]

NLL and LL are denoted in the legend in Fig. 6. In simulations, $\bar{s} = 0{\cdot}04$, $U = 0{\cdot}3$, $\bar{h} = 0{\cdot}21$ and $\sigma_e^2 = 10^3 V_m$. In design A, $M = 40$, $L = 100$, $R = 10$, $I = 10$ (i.e. assays are performed every 10 generations, so there is a total of five assays including the one at the beginning of the M-A). In design B, $M = 10$, $L = 100$, $R = 10$, $I = 10$ (i.e. assays are performed every 10 generations, so there is a total of two assays). The total lines assayed are 5000 for design A and 2000 for design B. The total M-A (defined in the legend to Table 2) is 4000 for design A and 1000 for design B.

for $\hat{U}$ decreases while the bias of $\hat{U}$ increases (though not significantly different from the expected $U/2$); the overall effect is that the MSE of $\hat{U}$ remains relatively stable with different σ_e^2/V_m.

(vi) *Direct comparison of two experimental designs* (Table 1)

Design A is similar to those implemented by previous researchers (Mukai *et al*., 1972; Houle *et al*., 1994; Kibota & Lynch, 1996; Keightley & Caballero, 1997; Shabalina *et al*., 1998). Design B is an alternative that may be adopted based on the previous simulation results. Compared with design A, design B takes much less time and probably much less labour due to the reduced total M-A (the product of M-A generation and the total number of M-A lines). However, the estimation as reflected by SD or MSE is at least indistinguishable for the two designs, and design B may be slightly better.

(viii) *The effect of high mutation pressure on estimation* (Table 2, Fig. 6)

In most of the previous simulations, we use the parameters that are close to those reported in Mukai *et al*. (1972). In these simulations the mutation pressure on fitness is relatively small. Hence line losses due to severe fitness decreases from potentially-many non-lethal mutations (stochastically accumulated) are very rare and are thus ignored. However, mutation parameters to be estimated are unknown beforehand. If the mutation pressure on fitness is in fact high (due either to an elevated U or a larger $\bar{s}$), line losses will occur relatively often (even without lethals) and will be a problem that cannot be ignored for estimation. Simulations are conducted in which sublines with fitness less than 0·1 are excluded from analysis, a protocol that was used in Mukai *et al*. (1972). The results are shown in Fig. 6 and Table 2. It can be seen that the progressing M-A generations, line losses (plots *e* and *f* in Fig. 6) will increase exponentially and the mean and genetic variance measured will deviate to a larger extent from the true values expected under no line losses. The effects of line losses are more significant under variable than under constant effects, due to the larger genetic variance among M-A lines under variable effects (note the scale difference in plots *e* and *f* in Fig. 6). Statistically speaking, the estimation employing data from the first 10 M-A generations is about the same as that using all 40 M-A generations under the unrealistic situations of no line loss (with both constant and variable effects; Table 2). Under the realistic situation of line losses, estimation employing data from the first 10 M-A generations is better than that using all 40 M-A generations under constant effects. This is reflected by the more accurate point estimates and the smaller MSE, which can be inferred from the reported means and SDs by Equation (6). Although under variable effects with line losses estimation is less biased with the data from 40 M-A generations, they deviate to a larger extent from the expected values than the estimation using the data of the first 10 M-A generations. The expected values are $\hat{U} = U/2$, and $\hat{\bar{s}} = 2\bar{s}$ under the exponential distribution for s_i (Equation 3). Then the usual practice (Mukai *et al*., 1972; Crow & Simmons, 1983; Lynch *et al*., 1995) or inferring U to be $2\hat{U}$ and $\bar{s}$ to be $\hat{\bar{s}}/2$ will overestimate U and underestimate $\bar{s}$ under variable effects with line losses. Therefore, appropriately reducing M-A generations can not only greatly reduce the cost of the M-A experiments while preserving comparable statistical properties of the estimation (Fig. 4, Table 1), but also may actually achieve better inferences due to the less serious problem of line losses.

https://doi.org/10.1017/S0016672398003681 Published online by Cambridge University Press

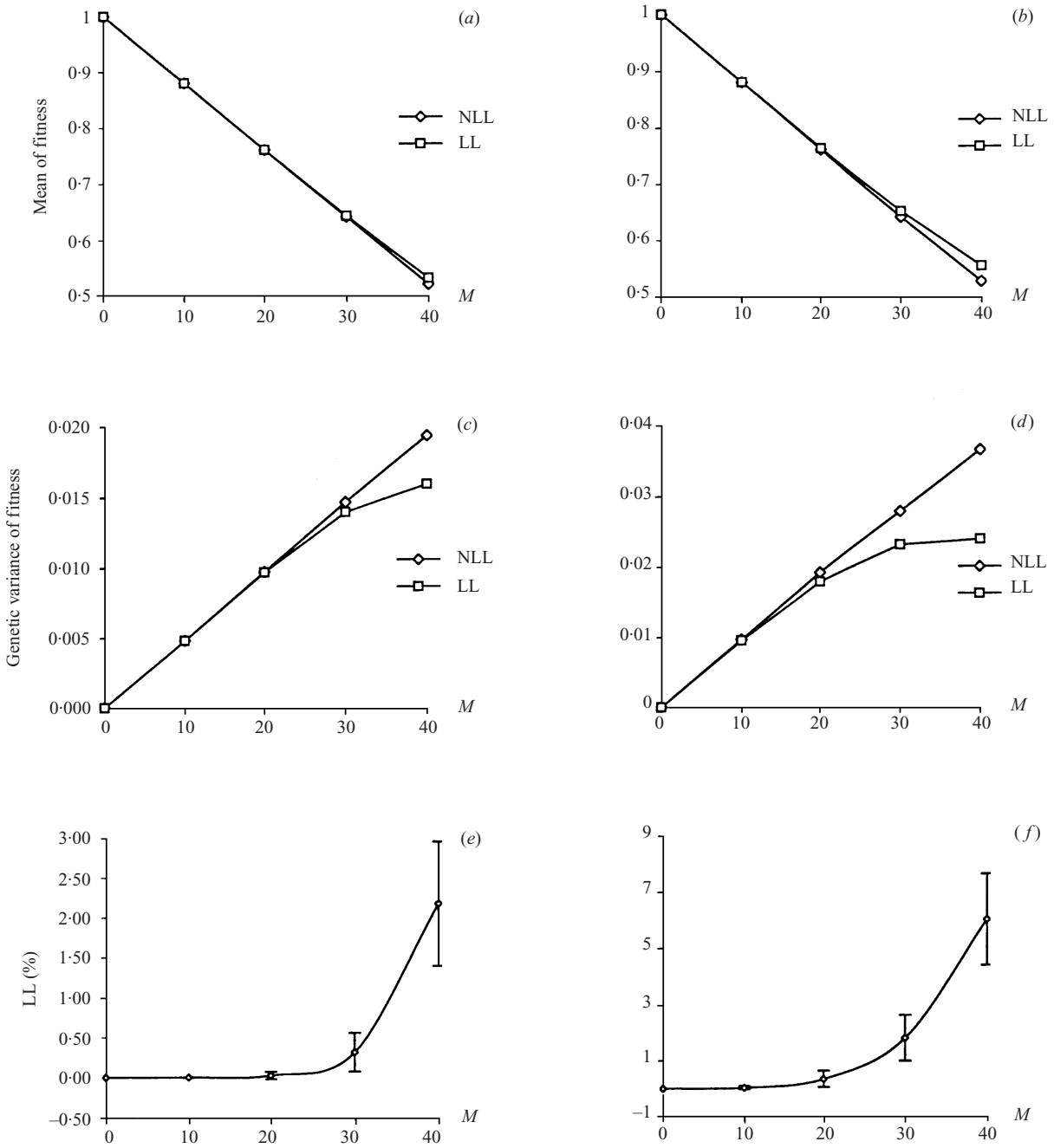

Fig. 6. The effects of line losses due to a high mutation pressure in M-A experiments. $A = 5$, $L = 100$, $R = 10$, $\bar{s} = 0{\cdot}04$, $U = 0{\cdot}3$, $M = 40$, $I = 10$. NLL indicates data points where all sublines are available for assay in simulations no matter what fitnesses they have (an unrealistic situation). LL indicates data points where sublines with fitness less than 0·1 (classified as lethals in Mukai *et al.*, 1972) are excluded from analyses (a more realistic approach). The three plots in the left-hand panel (plots *a*, *c* and *e*) are for constant effects, and the three plots in the right-hand panel (plots *b*, *d* and *f*) are for variable effects.

(viii) *Different experimental designs for fitness traits with exceptionally high* σ_e^2 (Table 3)

All the previous studies are not for fitness traits with normal magnitudes of σ_e^2 ($= \sim 10^3 V_m$). For fitness traits with exceptionally high σ_e^2, such as viability, additional simulations were performed in which $\sigma_e^2 = 2{\cdot}0 \times 10^4 V_m$. It can easily be seen that if σ_e^2 is indeed exceptionally high, it is almost impossible to estimate U and $\bar{s}$ without tremendous efforts. The experimental design *a* in Table 3 represents one that is similar to a typical M-A experimental design employed by most previous researchers (Mukai *et al.*, 1972; Houle *et al.*, 1994; Kibota & Lynch, 1996; Keightley & Caballero,

https://doi.org/10.1017/S0016672398003681 Published online by Cambridge University Press

Table 3. *Different experimental designs for fitness traits with exceptional high* σ_e^2

	Experimental designs				Constant effects		Variable effects		Total	Total lines
ED	*M*	*A*	*L*	*R*	$\hat{U}$	$\hat{s}$	$\hat{U}$	$\hat{s}$	M-A	assayed
a	40	5	50	10	0·230 (4·211)	0·040 (0·026)	0·090 [0·733]	0·080 [0·057]	2000	2500
b			50	40	0·168 (0·066)	0·040 (0·012)	0·082 [0·073]	0·080 [0·046]	2000	10000*
c			50	80	0·161 (0·048)	0·040 (0·010)	0·080 [0·073]	0·080 [0·044]	2000	20000
d	40	2	100	50	0·157 (0·036)	0·040 (0·008)	0·079 [0·073]	0·079 [0·042]	4000	10000*
e			50	100	0·160 (0·045)	0·040 (0·010)	0·079 [0·074]	0·080 [0·045]	2000	10000*
f	10	2	100	13	0·259 (6·536)	0·042 (0·052)	0·073 [1·787]	0·086 [0·098]	1000	2600
g			100	50	0·191 (0·199)	0·040 (0·017)	0·088 [0·109]	0·081 [0·050]	1000	10000*
h			100	100	0·163 (0·061)	0·040 (0·011)	0·081 [0·074]	0·080 [0·045]	1000	20000

ED denotes different experimental designs, *M* the total number of M-A generations in the experiment, *A* the number of assays performed, *L* the number of M-A lines, and *R* the replicates per line in each assay.
* Several experimental designs with the same total lines assayed in the experiments. The experiments were simulated under the ideal situation of no line losses. In simulations, $U = 0{\cdot}15$, $\bar{s} = 0{\cdot}04$, which are close to the estimates by Mukai *et al.* (tables 5 and 7 of Mukai *et al.*, 1972) for an M-A chromosome. $V_m = 5 \times 10^{-5}\ \sigma_e^2$, which is about the average for viability in *Drosophila* (Lynch, 1988).

1997; Shabalina *et al.*, 1998), but with more replicates (10) assayed for each M-A line. It can easily be seen that this design cannot reliably estimate U and $\bar{s}$. Even with 40 M-A generations, and assays performed every 10 generations (a total of five assays), only when genotypic values can be estimated more accurately with many replicates ($R > 40$) for each M-A line at each assay can U and $\bar{s}$ be relatively reliably estimated. With only 10 M-A generations and a total of two assays, many more replicates ($R > \sim 100$) would be needed in assays for each M-A line to achieve reasonably good estimates.

The design of M-A experiments is very important, especially those that will be tremendously labour-intensive. This can be demonstrated by comparing experimental designs *b*, *d*, *e* and *g*, in all of which 10000 sublines will have to be measured. However, the design *d* apparently will yield better (sometimes much better) estimates (as reflected by SD or MSE) than designs *b*, *e* and *g*. Note, in design *d*, that although M-A proceeds for 40 generations, only two assays need to be performed for the 100 M-A lines each with 50 replicate sublines. However, for design *b* in which five assays are performed but with a reduced number of M-A lines and replicates, the estimation (especially that under constant mutational effects) is inferior to design *d*.

4. Discussion

By our extensive simulations including σ_e^2, we conclude that *with at least 10 sublines replicated for each genotype at each assay*, the results 1 and 2 (see Section 1) from Deng & Fu (1998), ignoring σ_e^2, are robust for fitness traits *with normal magnitude of* σ_e^2. Namely, **if** sufficient (> 100) M-A lines are employed and genotypic values can be estimated relatively accurately, estimation based on experiments of only 10 M-A generations is comparable in estimation to that based on experiments of 40 M-A generations. However, *an efficient experimental design would critically depend on the magnitude of* σ_e^2 *of the fitness traits under study. If* σ_e^2 *is exceptionally high such as for viability, experiments with 40 M-A generations and two assays may be more efficient*. If the mutation pressure to be characterized is high, using data from more M-A generations may potentially have a more serious problem of line losses and introduce larger estimation biases under constant effects and larger inference biases under variable effects. Increasing the number of assays during M-A experiments does not improve estimation quality very much (in terms of bias and sampling variance), so that only two assays (one at the beginning and the other at the end of M-A) may be performed instead of multiple assays.

We re-analyzed the data of Kibota & Lynch (1996) in the light of our simulation results. In their experiments, 50 initially isogenic *E. coli* lines were allowed to accumulate spontaneous deleterious muta-

https://doi.org/10.1017/S0016672398003681 Published online by Cambridge University Press

tions for 300 generations and assays were performed for the lines from the M-A generations 0, 100, 120, 200, 250 and 300. The mutation pressure is relatively small and line losses during M-A are rare (Kibota & Lynch, 1996). Mean fitness and genetic variance among M-A lines at different M-A assay generations are extrapolated from figure 3 in Kibota & Lynch (1996). Applying Equation (1 *b*) to the inferred values revealed that $\hat{U} = 0{\cdot}0033$ and $\hat{\bar{s}} = 0{\cdot}017$ (from the data of the assays at M-A generations of 0 and 100 only), $\hat{U} = 0{\cdot}0036$ and $\hat{\bar{s}} = 0{\cdot}016$ (from the data of the assays at M-A generations 0 to 300 only), $\hat{U} = 0{\cdot}0041$ and $\hat{\bar{s}} = 0{\cdot}015$ (from the data of all six assays). These estimates are so close so that additional M-A beyond the 100th M-A generation may not have been necessary. The standard errors of these estimates cannot be computed without the original data on individual M-A lines. However, as shown in plots *c* and *d* of Figs. 3 and 4, under plausible variable mutational effects the statistical properties (sampling variance and bias as reflected by the composite index MSE) change little with an increasing number of M-A generations or the number of assays. This is especially true for the U estimates. This concrete example demonstrates that the experimental design issues are very important and that multiple assays may not be necessary if each assay is implemented accurately. It should be noted that $\hat{U}$ and $\hat{\bar{s}}$ here are different from those reported in Kibota & Lynch (1996). This is because their estimates were corrected for the sampling variation of the rate of changes of mean and genetic variance during M-A (see the legend to figure 3 of Kibota & Lynch, 1996), which can only be done based on their original data for different M-A lines.

In investigating the design and statistical properties for the M-A experiments in our simulations we employed the most commonly used estimation method: Bateman–Mukai's method of moments; this is computationally efficient and easy to implement in practice. Keightley (1994) developed an alternative maximum likelihood (ML) estimation method, for which the data from multiple assay generations cannot be analysed with the current available program (Dr Keightley is still working on this problem; P. D. Keightley, personal communication). A gamma distribution is assumed to fit the distribution of s_i with shape parameter β and scale parameter α:

$$\bar{s} = \beta/\alpha \quad \text{and} \quad \sigma^2_{s_i} = \beta/\alpha^2. \tag{7}$$

$\beta \leftarrow \infty$ is the limiting case for all the mutant effects being equal. As $\beta \rightarrow 0$, the distribution of mutant effects becomes increasingly leptokurtic. As with Bateman–Mukai's method of moments, Keightley's method *cannot* estimate U, and all the distribution parameters α and β *simultaneously* and *individually* from M-A data. Estimates are strongly confounded with other parameters to be estimated in the model. The reason for this may be that the M-A data do not contain enough information. One parameter must be assumed in order to estimate the other parameters. Keightley (1994) estimated β by assuming a U value. In Bateman–Mukai's method of moments, it can easily be seen that if U is assumed known, the mean and variance of s can be estimated (eqn (1 *a*). Further, if a gamma distribution is assumed as in Keightley (1994), β can easily be derived from the mean and variation of s easily from Equation (7). Therefore, contrary to the general belief, Keightley's method (1994) *does not* yield estimates on more parameters from M-A data about deleterious genomic mutations than Bateman–Mukai's method of moments. Keightley's method (1994) may provide a way of testing for variability in mutational effects and a way of discriminating alternative distributions of mutation effects by quantifying the likelihood. However, the power and the accuracy are generally unknown and may also critically depend on other unknown parameter such as U.

To apply Keightley's (1994) estimation method, we simulated M-A for 40 generations with $L = 100$, $R = 10$, $\bar{s} = 0{\cdot}04$, $U = 0{\cdot}15$, and $\sigma^2_e = 10^3 V_m$. The mutation pressure is not very high, line losses are very rare and are thus ignored. Means of M-A lines are obtained and genetic variance and σ^2_e computed by one-way ANOVA at assays conducted at M-A generations 10 and 40 respectively. Keightley's ML method is then applied to the data obtained at M-A generations 10 (panel *A* of Fig. 7) and 40 (panel *B* of Fig. 7), respectively. Due to the relatively large computational demand of M-A estimation, such simulations are repeated 'only' five times for which the resultant profile likelihood curves are presented. It can be seen (plots *a* and *b* of panels *A* and *B* in Fig. 7) that under both constant and variable effects, the profile of the likelihood peaks near the simulated values (Ut, where t is the M-A generation). This is also approximately true for the likelihood profiles for β estimation (plots *c* and *d* of panels *A* and *B* in Fig. 7). The log likelihood profile curves for β estimation change very slowly along different β, which suggests that Keightley's M-A approach may have very little power in distinguishing different distributions of mutational effects. It is also apparent that estimation of β can be sensitive to an assumed U, especially under constant mutation effects (plots *e* and *f* of panels *A* and *B* in Fig. 7). Likewise, estimation of U is also sensitive to an assumed β (plots *g* and *h* of panels *A* and *B* in Fig. 7). Data not shown revealed that Keightley's M-L program may fail to find global maxima, even with the starting value of the other parameter(s) set close to the true (but generally unknown) values in implementing th M-L program. It is also seen, by visual inspection of the corresponding plots of panels *A* and *B* in Fig.

https://doi.org/10.1017/S0016672398003681 Published online by Cambridge University Press

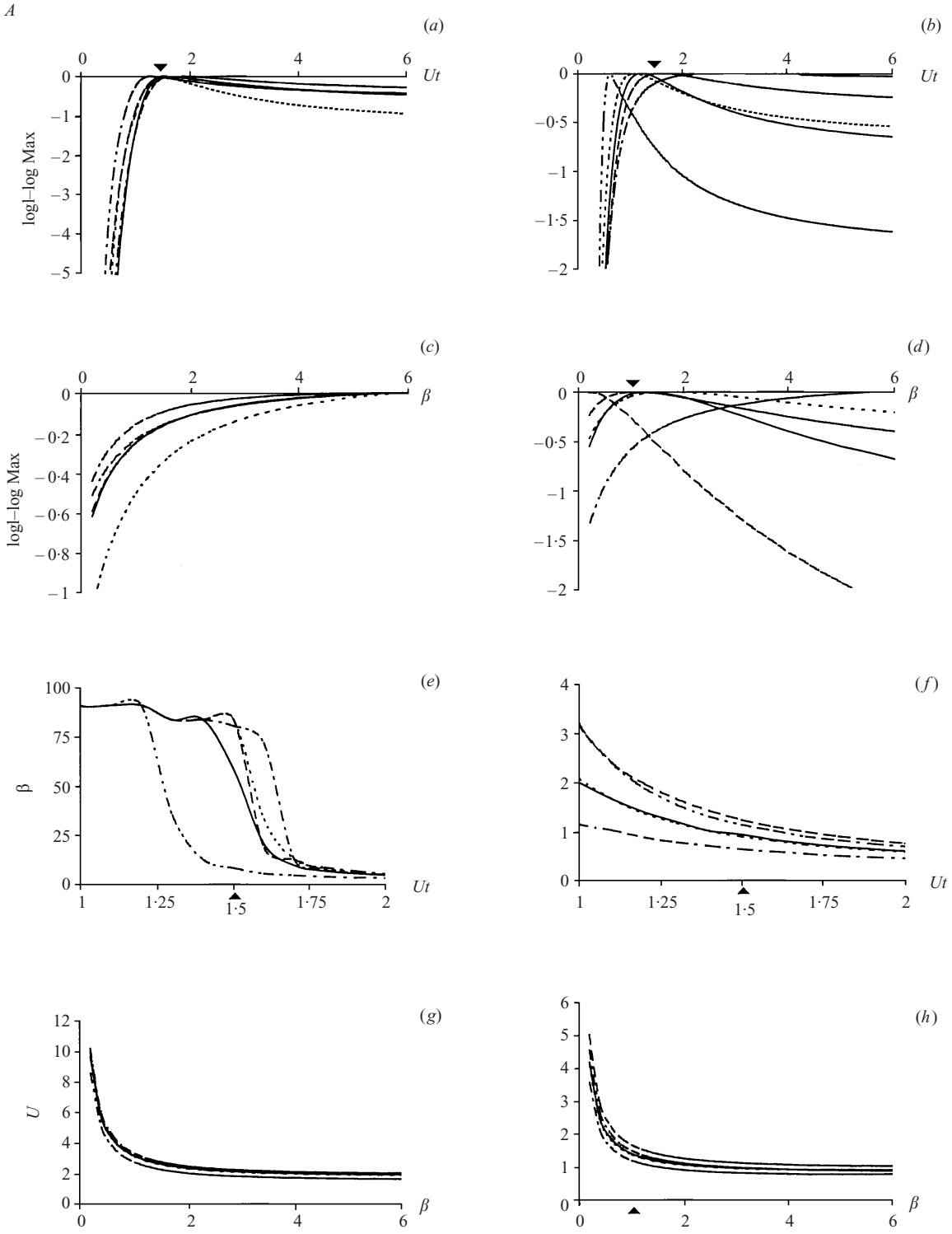

Fig. 7A. For description see opposite.

7 that, apparently, estimation using data from the fortieth M-A generation is not significantly better than that from the tenth M-A generation. This is consistent with earlier extensive simulation results that the additional M-A generations beyond the tenth M-A generation may not be very fruitful.

It should be noted that our conclusions and recommendations here are mainly based on statistical criteria, such as sampling variance and bias. There may be practical constraints (such as manpower and available facilities) preventing large numbers of replicates from being measured at each assay. In this case the M-A experiments may have to be spread over more M-A generations and assays performed at

https://doi.org/10.1017/S0016672398003681 Published online by Cambridge University Press

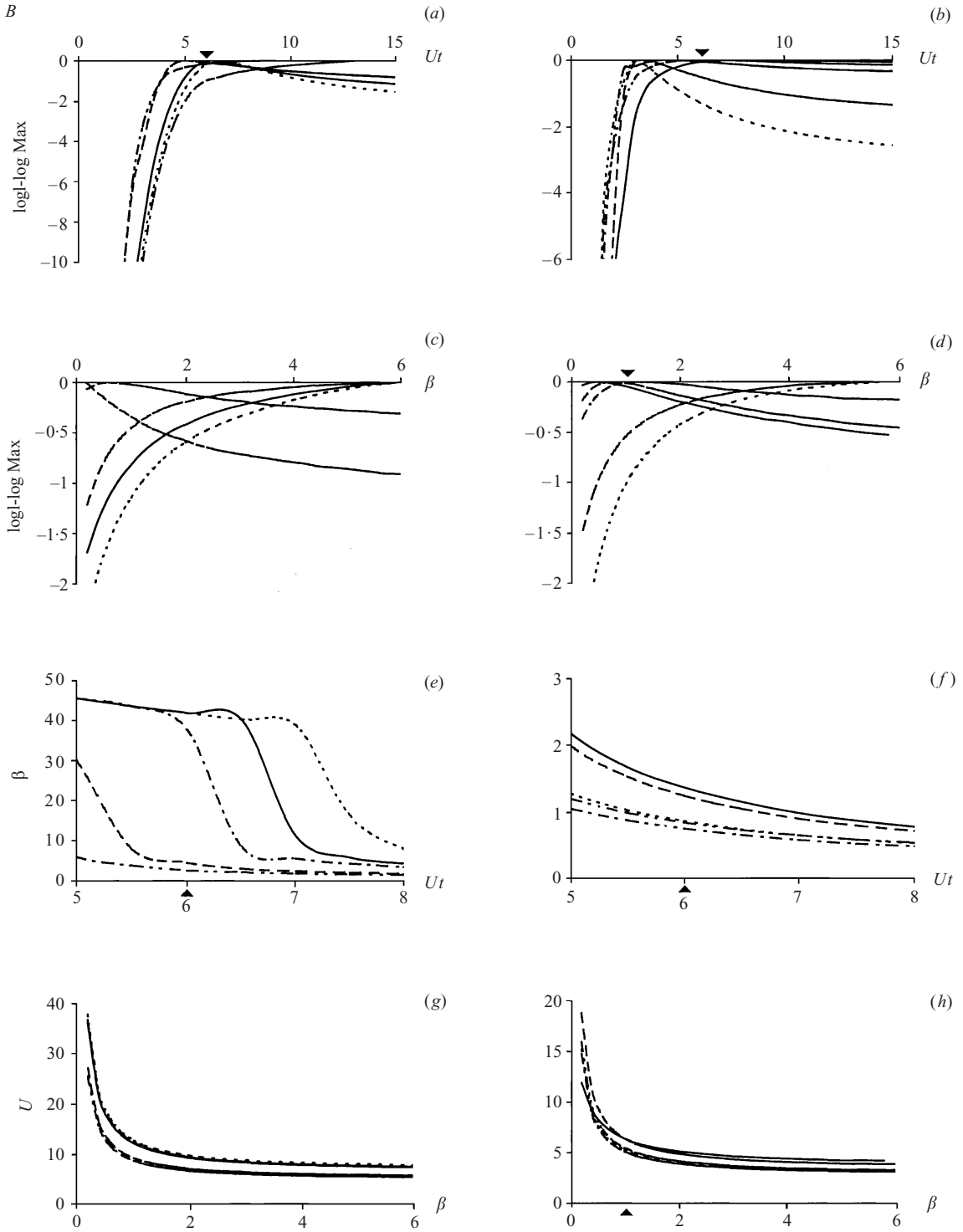

Fig. 7. Analyses of M-A data using Keightley's ML estimation. Panel (*A*): Estimation using M-A lines at the tenth M-A generation: The profiles of log likelihood as a function of Ut (where t is the M-A generation at assay) are shown under constant (plot *a*) and variable (plot *b*) effects and those of β are shown in plots *c* (constant effect) and *d* (variable effect). The estimation of β as a function of an assumed U is shown in plots *e* (under constant effects) and *f* (under variable effects). The estimation of U as a function of an assumed β is shown in plots *g* (under constant effects) and *h* (under variable effects). Panel (*B*): Estimation using M-A lines at the fortieth M-A generation. Plots correspond to those in panel (*A*). The black triangle pinpoints the true parameter values for the simulations. Under constant effects, $\beta \to \infty$.

multiple times, so that at each assay relatively fewer replicates per M-A line need to be assayed. In these situations the exact experimental design depends not only on statistical criteria but also on the specific practical constraints at hand. The C program for simulation studies here is available upon request for

https://doi.org/10.1017/S0016672398003681 Published online by Cambridge University Press

empiricists planning M-A experiments. By accounting for the practical constraints, power and different experimental designs can be experimented with our program, so that an efficient M-A experiment can be designed.

In this study we simulate only the most essential processes common to all M-A experiments. Some other procedures that differ from one experiment to another are not simulated. One of these is adequate control for potential environmental changes in mean fitness. Controls for such temporal environmental changes can be implemented by raising large case populations, by freezing a large sample of individuals for later revival and measurement, or by storing a large number of seeds/eggs for later planting/hatching and measurements. There will be measurement error for the mean of these controls, since the number of control individuals to be measured is finite. More importantly, the intrinsic mean of these controls may not be stable temporally due to finite numbers of controls raised and measured, potential effects of freezing and revival on individuals, and/or different periods of dormancy on the quality of seeds or eggs, etc. The larger the number of M-A generations, the more serious the potential problem of the change in the intrinsic mean values of the controls. This is also consistent with the idea that M-A experiments should not proceed for an excessive number of M-A generations and, if at all possible, great efforts should be made in each assay to measure genotypic values as accurately as possible.

This study focuses on the goals for efficiently estimating U and $\bar{s}$ via M-A experiments. Characterizing deleterious genomic mutations also involves characterizing mutational effects (additive, multiplicative or epistatic), which may also be accomplished in M-A experiments (Mukai, 1969) by the trends of the mean of M-A lines with M-A generations. Although the trend of M-A line means over M-A generations is valuable, M-A experiments may be a poor strategy for characterizing mutational effects. Even with over 60 generations of M-A, the evidence for epistatic mutational effects (Mukai, 1969) is not conclusive.

The M-A approach was proposed nearly four decades ago (Bateman, 1959). However, few scientists have the determination to carry out M-A experiments. This is at least partially because (1) the statistical properties and the experimental design issues have never been formally and thoroughly investigated and therefore (2) the few M-A experiments implemented are extremely time- and labour-consuming and their scales are beyond the resources and capacity of most laboratories. However, our results here indicate that almost all previous M-A experiments (e.g. Mukai *et al*., 1972; Houle *et al*., 1994; Kibota & Lynch, 1996; Keightley & Caballero, 1997; Shabalina *et al*., 1998) may not have been efficiently or appropriately implemented. Our results not only provide some practical guidelines for designing efficient and powerful M-A experiments, but also clearly demonstrate that, if efficiently implemented, M-A experiments for most fitness traits are probably well within the available resources for many laboratories. Therefore, more scientists will be able to employ the M-A approach to characterize deleterious genomic mutations in a wider range of taxa.

Indisputably, characterizing deleterious genomic mutations is important. However, even if the importance is realized by an increasing number of scientists and revealed in more and more biological aspects, the estimates are few and thus are needed (Peck & Eyre-Walker, 1997). Of the three available estimation approaches, under their respective assumptions. Deng & Lynch's fitness moments approach (Deng & Lynch, 1996, 1997) is generally more efficient and statistically better than the other two (Deng & Fu, 1998), especially with a modified design in outcrossing populations (Deng, 1998*b*). However, different approaches make different assumptions whose validity may be difficult to consolidate in a specific experimental setting (Keightley, 1994; Peck & Eyre-Walker, 1997; Deng & Fu, 1998; Lynch *et al*., 1998). Examples of these assumptions are mutation–selection balance for the fitness moments approach and the mean fitness approach, no line losses due to selection during M-A, etc. Applying multiple approaches to the same organism and/or characterizing deleterious mutations in diverse organisms may provide a cross-check of the results (together with the underlying assumptions for deriving these results) and eventually crystallize the deleterious mutation parameters.

H.-W. D. would like to thank Professor M. Lynch for years of advice, continuous encouragement and support. We are also grateful for Professor P. Keightley for providing us with his C program to implement the maximum likelihood estimation, and for his discussions on a number of issues concerning his maximum likelihood estimation. Although Dr Keightley and ourselves disagree on some points raised concerning his ML estimation, we appreciate his time for the extensive discussions. Comments from Professor T. Mackay and anonymous reviewers greatly helped to improve the manuscript. The work was supported by a grant from National Institutes of Health R01 AR45349 and a Health Future Foundation grant from Creighton University, Nebraska, and by graduate student tuition waiver to J. L. and J.-L. L. from the Department of Biomedical Sciences of Creighton University.

References

Bateman, A. J. (1959). The viability of near-normal irradiated chromosomes. *International Journal of Radiation Biology* **1**, 170–180.

Buger, R. & Hofbauer, J. (1994). Mutation load and quantitative genetic traits. *Journal of Mathematical Biology* **32**, 193–218.

Caballero, A. & Keightley, P. D. (1994). A pleiotropic nonadditive model of variation in quantitative traits. *Genetics* **138**, 883–900.
Charlesworth, B. (1990). Mutation–selection balance and the evolutionary advantage of sex and recombination. *Genetical Research* **55**, 199–221.
Charlesworth, B., Charlesworth, D. & Morgan, M. T. (1990). Genetic loads and estimates of mutation rates in highly inbred plant populations. *Nature* **347**, 380–382.
Charlesworth, B., Morgan, M. T. & Charlesworth, D. (1993). The effects of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289–1303.
Charlesworth, D. & Charlesworth, B. (1987). Inbreeding depression and its evolutionary consequences. *Annual Review of Ecology and Systematics* **18**, 237–268.
Charlesworth, D., Lyons, E. E. & Litchfield, L. B. (1994). Inbreeding depression in two highly inbreeding populations of *Leavenworthia*. *Proceedings of the Royal Society of London*, *Series B* **258**, 209–214.
Charlesworth, D., Charlesworth, B. & Morgan, M. T. (1995). The pattern of neutral molecular variation under the background selection model. *Genetics* **41**, 1619–1632.
Crow, J. F. & Simmons, M. J. (1983). The mutation load in *Drosophila*. In *The Genetics and Biology of* Drosophila (ed. M. Ashburner, H. L. Carson & J. N. Thompson), vol. 3c, pp. 1–35. London: Academic Press.
Deng, H.-W. (1998*a*). Estimating (over)dominance coefficient and discriminating dominance vs overdominance as the genetic cause of heterosis. *Genetics* **148**, 2003–2014.
Deng, H.-W. (1998*b*). Characterizing deleterious genomic mutations in outcrossing populations. *Genetics* (in press).
Deng, H.-W. & Fu, Y.-X. (1998). On the three methods for estimating deleterious genomic mutation parameters. *Genetical Research* **71**, 223–236.
Deng, H.-W. & Lynch, M. (1996). Estimation of the genomic mutation parameters in natural populations. *Genetics* **144**, 349–360.
Deng, H.-W. & Lynch, M. (1997). Inbreeding depression and inferred deleterious mutation parameters in *Daphnia*. *Genetics* **147**, 147–155.
Falconer, D. S. & Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics*. Harlow, UK: Longman.
Fernandez, J. & Lopez-Fanjul, C. (1996). Spontaneous mutational variances and covariances for fitness-related traits in *Drosophila melanogaster*. *Genetics* **143**, 829–837.
Garcia-Dorado, A. (1997). The rate and effects distribution of viability mutation in *Drosophila*: minimum distance estimation. *Evolution* **51**, 1130–1139.
Gregory, W. C. (1965). Mutation frequency, magnitude of change and the probability of improvement in adaptation. *Radiation Botany* **5** (Suppl.), 429–441.
Haldane, J. B. S. (1937). The effects of variation on fitness. *American Naturalist* **71**, 337–349.
Houle, D., Hughes, K. A., Hoffmaster, D. K., Ihara, J., Assimacopoulos, S., Canada, D. & Charlesworth, B. (1994). The effects of spontaneous mutation on quantitive traits. I. Variances and covariances of life history traits. *Genetics* **138**, 773–785.
Houle, D., Morikawa, B. & Lynch, M. (1996). Comparing mutational variabilities. *Genetics* **143**, 1467–1483.
Hudson, R. & Kaplan, N. L. (1995). Deleterious background selection with recombination. *Genetics* **141**, 1605–1617.
Johnston, M. & Schoen, D. J. (1995). Mutation rates and dominance levels of genes affecting total fitness in two angiosperm species. *Science* **267**, 226–229.
Kacser, H. & Burns, J. A. (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. & Caballero, A. (1997). Genomic mutation rates for lifetime reproductive output and lifespan in *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences of the USA* **94**, 3823–3827.
Keightley, P. D. & Hill, W. G. (1990). Variation maintained in quantitative traits with a mutation–selection balance: pleiotropic side effects on fitness. *Proceedings of the Royal Society of London*, *Series*, *B* **253**, 291–296.
Kibota, T. T. & Lynch, M. (1996). Estimates of the genomic mutation rate deleterious to overall fitness in *E. coli*. *Nature* **381**, 694–696.
Kimura, M., Maruyama, T. & Crow, J. F. (1963). The mutation load in small populations. *Genetics* **48**, 1303–1312.
Kirkpatrick, M. & Ryan, M. J. (1991). The evolution of mating preferences and the paradox of the lek. *Nature* **350**, 33–38.
Kondrashov, A. S. (1985). Deleterious mutations as an evolutionary factor. II. Facultative apomixis and selfing. *Genetics* **111**, 635–653.
Kondrashov, A. S. (1988). Deleterious mutations and the evolution of sexual reproduction. *Nature* **336**, 435–440.
Kondrashov, A. S. & Crow, J. F. (1991). Haploidy or diploidy: which is better? *Nature* **351**, 314–315.
Lande, R. (1994). Risk of population extinction from new deleterious mutations. *Evolution* **48**, 1460–1469.
Lynch, M. (1985). Spontaneous mutations for life-history characters in an obligate parthenogen. *Evolution* **39**, 804–818.
Lynch, M. (1988). The rate of polygenic mutation. *Genetical Research* **51**, 137–148.
Lynch, M., Burger, R., Butcher, D. & Gabriel, W. (1993). The mutation meltdown in small asexual population. *Journal of Heredity* **84**, 339–344.
Lynch, M., Conery, J. & Burger, R. (1995). Mutation accumulation and the extinction of small populations. *American Naturalist* **146**, 489–518.
Lynch, M., Conery, J. & Burger, R. (1996). Mutation meltdowns in sexual populations. *Evolution* **49**, 1067–1080.
Lynch, M., Latta, L., Hicks, J. & Giorgianni, M. (1998). Mutation, selection, and maintenance of life-history variation in a natural population. *Evolution* (in press).
Mackay, T. F. C., Lyman, R. F. & Jackson, M. S. (1992). Effects of *P* element insertions on quantitative traits in *Drosophila melanogaster*. *Genetics* **130**, 315–322.
Morton, N. E., Crow, J. F. & Muller, H. J. (1956). An estimate of the mutational damage in man from data on consanguineous marriages. *Proceedings of the National Academy of Sciences of the USA* **42**, 855–863.
Mukai, T. (1964). Spontaneous mutation rate of polygenes controlling viability. *Genetics* **50**, 1–19.
Mukai, T. (1969). The genetic structure of natural populations of *Drosophila melanogaster*. VII. Synergistic interaction of spontaneous mutant polygenes controlling viability. *Genetics* **61**, 749–761.
Mukai, T. (1979). Polygenic mutations. In *Quantitative Genetic Variation* (ed. J. N. Thompson Jr & J. M. Thoday), pp. 177–196. New York: Academic Press.
Mukai, T., Chigusa, S. I., Mettler, L. E. & Crow, J. F. (1972). Mutation rate and dominance of genes affecting viability in *Drosophila melanogaster*. *Genetics* **72**, 335–355.

Muller H. J. (1964). The relation of recombination to mutational advance. *Mutation Research* **1**, 1–9.
Ohnishi, O. (1977). Spontaneous and ethyl methanesulfonate-induced mutations controlling viability in *Drosophila melangoaster*. II. Homozygous effect of polygentic mutations. *Genetics* **87**, 520–545.
Peck, J. R. & Eyre-Walker, A. (1997). The muddle about mutations. *Nature* **87**, 135–136.
Santiago, E., Albornoz, J., Dominguez, A., Toro, M. A. & Lopez-Fanjul, C. (1992). The distribution of spontaneous mutations on quantitative traits and fitness in *Drosophila melanogaster*. *Genetics* **132**, 771–781.
Shabalina, A. S., Yampolsky, L. Y. & Kondroshov, A. S. (1998). Rapid decline of fitness in panmictic *Drosophila* populations under relaxed selection *Proceedings of the National Academy of Sciences of the USA* **94**, 13034–13039
Sokal, R. R. & Rohlf, F. J. (1995). *Biometry*, 3rd edn. New York: W. H. Freeman.