

Comparing Analysis Methods for Mutation-Accumulation Data: A Simulation Study

Aurora García-Dorado¹ and Araceli Gallego

Departamento de Genética, Facultad de Biología, Universidad Complutense de Madrid, 28040 Madrid, Spain

Manuscript received December 29, 2002

Accepted for publication March 11, 2003

ABSTRACT

We simulated single-generation data for a fitness trait in mutation-accumulation (MA) experiments, and we compared three methods of analysis. Bateman-Mukai (BM) and maximum likelihood (ML) need information on both the MA lines and control lines, while minimum distance (MD) can be applied with or without the control. Both MD and ML assume gamma-distributed mutational effects. ML estimates of the rate of deleterious mutation had larger mean square error (MSE) than MD or BM had due to large outliers. MD estimates obtained by ignoring the mean decline observed from comparison to a control are often better than those obtained using that information. When effects are simulated using the gamma distribution, reducing the precision with which the trait is assayed increases the probability of obtaining no ML or MD estimates but causes no appreciable increase of the MSE. When the residual errors for the means of the simulated lines are sampled from the empirical distribution in a MA experiment, instead of from a normal one, the MSEs of BM, ML, and MD are practically unaffected. When the simulated gamma distribution accounts for a high rate of mild deleterious mutation, BM detects only ~30% of the true deleterious mutation rate, while MD or ML detects substantially larger fractions. To test the robustness of the methods, we also added a high rate of common contaminant mutations with constant mild deleterious effect to a low rate of mutations with gamma-distributed deleterious effects and moderate average. In that case, BM detects roughly the same fraction as before, regardless of the precision of the assay, while ML fails to provide estimates. However, MD estimates are obtained by ignoring the control information, detecting ~70% of the total mutation rate when the mean of the lines is assayed with good precision, but only 15% for low-precision assays. Contaminant mutations with only tiny deleterious effects could not be detected with acceptable accuracy by any of the above methods.

THE properties of mutations affecting fitness and its component traits are relevant to many evolutionary and conservation issues, and an important experimental effort is currently being devoted to their understanding. A widely used procedure is the mutation-accumulation (MA) design where, by relaxing natural selection as much as possible, deleterious mutations are allowed to accumulate in replicate lines derived from a common, genetically invariable, origin. This procedure provides information on both the rate and the effect of mutations, which do not rely on the assumption of mutation-selection balance (see reviews by GARCÍA-DORADO *et al.* 1999; KEIGHTLEY and EYRE-WALKER 1999; and LYNCH *et al.* 1999; and recent experiments by VASSILIEVA *et al.* 2000; CHAVARRÍAS *et al.* 2001 and SHAW *et al.* 2002). However, these inferences are indirect and require the use of statistical techniques whose properties depend on many factors, such as the shape of the distribution of mutational effects and the magnitude of the residual errors. Basically, the following three methods have been used to analyze MA data.

The Bateman-Mukai (BM) procedure, which has the advantage of simplicity, is more frequently used (see MUKAI *et al.* 1972). It estimates a lower bound for the rate of mutation affecting a fitness trait per gamete and generation (λ) and an upper bound for the corresponding expected deleterious effect, $E(s)$, using the trait's rates of increase in variance (ΔV) and decline in mean (ΔM) caused by mutation. The validity of these bounds is quite general, but their closeness to the true λ and $E(s)$ values (and, therefore, their usefulness) decays as the variance of the deleterious effect increases.

The second approach is a maximum-likelihood (ML) one, requiring some assumptions on the distribution of the deleterious effects (s). This method has been implemented by KEIGHTLEY (1994), assuming a gamma-distributed s or a constant deleterious effect (S). It searches for the mutation rate and the gamma parameters (or the S value), maximizing the likelihood of the observed distribution of the means of the MA and control lines. Here we concentrate on the more widely used gamma alternative. The method has been extended to a reflected gamma distribution, where $|s|$ is gamma distributed but s has negative sign with some probability (KEIGHTLEY and OHNISHI 1998). Under the more restrictive assumption of constant deleterious effects, the

¹Corresponding author: Departamento de Genética, Facultad de Biología, Universidad Complutense de Madrid, 28040 Madrid, Spain.
E-mail: augardo@bio.ucm.es

method has been further extended to analyze multigeneration data (KEIGHTLEY and BATAILLON 2000).

The third method is minimum distance (MD), and it also requires the assumption of a distribution for s . It has been implemented by García-Dorado by assuming either a reflected gamma or a mixed normal-gamma distribution for s (see GARCÍA-DORADO 1997). Here we concentrate on the reflected-gamma alternative. This method searches for the mutational parameters (*i.e.*, the mutation rate and the parameters of the reflected gamma) that minimize the Cramer-von Mises distance between the observed and theoretically predicted distribution of the line means, as is explained in the next section.

MD results have generally been obtained conditional on an estimate of ΔV but not necessarily on one of ΔM . This is of particular interest since this ΔM is estimated by comparison to a control population, and maintaining a reliable control (efficiently protected against both the accumulation of deleterious mutations and the selection of advantageous ones) is loaded with pitfalls. For FERNÁNDEZ and LÓPEZ-FANJUL (1996) *Drosophila* MA data, MD estimates characterized by low mutation rates and moderate average deleterious effects have been obtained regardless of whether or not the control information is used. These MD estimates were similar to the corresponding BM ones (GARCÍA-DORADO *et al.* 1998). Low rates and moderate effects were also obtained from MD analysis of Mukai's classical MA data (MUKAI *et al.* 1972) if the large ΔM observed was ignored. In that analysis, the MD estimate of λ was one order of magnitude lower than that of the BM, suggesting that the large ΔM observed by MUKAI *et al.* (1972) could be biased (GARCÍA-DORADO *et al.* 1998). However, it has been argued that the MD method is unable to detect mildly deleterious mutation and is strongly biased due to sampling errors departing slightly from normality (LYNCH *et al.* 1999).

Since MA experiments are extremely time- and work-consuming, it is of obvious practical interest to study the properties of the above statistical techniques to choose, in each case, the one extracting maximum information from the data. For this purpose, several simulation studies have been carried out (GARCÍA-DORADO 1997; KEIGHTLEY 1998; DENG *et al.* 1999). However, the relative merits of the different methods have not been established. Furthermore, in the ML and MD studies quoted above, data were simulated using the model assumed during estimation. Thus, the robustness of these methods to departures from the model has not been checked.

In this work we analyze MA-simulated data using BM, ML, and MD methods to compare the relative merits of the three approaches. We use reflected gamma-distributed deleterious effects (see below) and restrict the scope of this study to those MA experiments in which

lines and control are assayed only once at the end of the MA period. The properties of the MD method are analyzed both using and not using the information provided by the control. To simulate the data, we use normally or nonnormally distributed sampling errors. To check the robustness of the different estimation methods against departures from the model assumed, we introduce mutations with small constant deleterious effects, not fitting within the gamma distribution.

MATERIALS AND METHODS

The background accumulation model

For convenience, we consider a fitness trait that we denote viability and a set of isolated MA lines derived from a common isogenic strain. We assume that the number of deleterious mutations occurring per gamete and per generation is Poisson distributed with mean λ , so that $2N\lambda$ new single-copy deleterious mutations are expected to occur per generation in the N breeding individuals of each MA line. We consider nonsevere deleterious mutations that randomly drift in the MA lines due to their very low population size, so that any new mutation has a final fixation probability $1/(2N)$. Therefore, MA individuals sampled at generation t are expected to be homozygous for λF_t mutations that occurred at generation x , where F_x is the probability of identity by descent at generation t taking x as the reference noninbred generation. Thus, considering mutations that occurred at all previous generations, MA individuals are expected to be homozygous for $U = \lambda F_t$ new mutations, where $F_t = \sum_{x < t} F_x$ is the "forward cumulated inbreeding coefficient." Since each mutation is assumed to be unique, individuals from different MA lines will be homozygous for different mutations. Thus, the number of mutations accumulated per MA line at generation t is Poisson distributed with mean U . The between-line genetic variance and overall viability mean at generation t are, respectively,

$$\sigma_t^2 = F_t \lambda E(s^2) = UE(s^2),$$

$$\mu_t = \mu_0 - F_t \lambda E(s) = \mu_0 - UE(s),$$

where μ_0 is the mean viability of the original isogenic strain, before mutation accumulation. In general, F_t/t asymptotically approaches 1 for increasing t . In those experiments where a single copy of the MA genome or chromosome is transferred per generation and line, F_t equals the number t of MA generations. Thus, in the following, we use t instead of F_t for simplicity.

Simulating data

To investigate the different estimation methods, we simulated sets of diploid lines that were evaluated once for viability, after a given MA period during which deleterious mutations, completely sheltered from natural selection, fix at random. The number i of deleterious mutations fixed in each line was always assumed to be Poisson distributed with average U . For each set of parameters considered, 10–12 MA data sets were simulated (see below), each consisting of 200 MA lines. For each data set, a control population, also consisting of 200 lines, was simulated using $U = 0$. To check the robustness of the MD estimation procedure, one pure and three contaminated models were analyzed (see below).

Establishing the residual variance: To obtain results comparable to the ML estimates obtained by KEIGHTLEY (1998), we

adjusted the simulated residual variance to give a quotient $Q = \sigma_G^2/\sigma_R^2 = 20$ [where $\sigma_G^2 = UE(s^2)$ is the genetic between-line variance], although a smaller quotient ($Q = 2$) was also used.

“Pure” model: In this “pure” model, the homozygous deleterious effect of each mutation (s) was gamma distributed with shape parameter α and scale parameter β , so that the deleterious effect had expected value $E(s) = \alpha/\beta$ and the expected quadratic effect is $E(s^2) = \alpha(\alpha + 1)/\beta^2$. Thus, the average viability of each line was computed as

$$v = 1 + R - \sum s,$$

where $\sum s$ represents the sum of effects over i (the number of independent deleterious mutations accumulated in the line, sampled from a Poisson distribution with mean U), and R is the sampling error, which was sampled from a normal distribution with mean zero and variance σ_R^2 . The simpler and equivalent procedure used involves sampling $\sum s$ from a gamma distribution with shape parameter $i\alpha$ and scale parameter β . We arbitrarily used the scale parameter value (β) giving $E(s) = 0.1/U$. The residual variance was established as stated above, *i.e.*,

$$\sigma_R^2 = U\alpha(\alpha + 1)/(Q\beta^2). \quad (1)$$

For the ($U = 0.5, \alpha = 2$) case, 10 new data sets of 10 MA-control data were simulated assuming reflected-gamma-distributed effects. This means that their absolute value is gamma distributed but the sign is randomly assigned, each mutation being advantageous for viability ($s < 0$) with probability P_a .

“Residual contaminated” model: In this model, the simulated sampling error R was nonnormally distributed, but was randomly sampled from a set of residual errors estimated from the FERNÁNDEZ and LÓPEZ-FANJUL (1996) *Drosophila* viability MA data, in which viability was assayed as the percentage of adults emerging from the eggs laid by a single female. The mean viability of each MA line was obtained by averaging over 12 females (4 females per line and generation over three consecutive generations). Deviations d of each individual measure from the average of its MA line and generation were computed and joined into a single large pool. To obtain the residual error R corresponding to the mean viability of each simulated MA line, $n_e = 3$ d -values were sampled and averaged, and the resulting R was scaled to give the required σ_G^2/σ_R^2 quotient, *i.e.*, $\sigma_R^2 = U\alpha(\alpha + 1)/(Q\beta^2)$.

“Mild contaminated” model: In this model, an additional random number (i_{mild}) of mildly deleterious mutations, with constant effect $s_{\text{mild}} = 0.025$, was simulated to accumulate in the lines. In the case where $U = 0.5, \alpha = 2$, for each “pure” model MA line, a mild contaminated MA line was derived, with viability $v_M = v + 0.025 i_{\text{mild}}$, where i_{mild} was sampled from a Poisson distribution with mean $U_{\text{mild}} = 10$. Thus, the total deleterious mutation rate is $U = U_{\text{pure}} + U_{\text{mild}}$, where $U_{\text{pure}} = 0.5$ is the U value used in the pure model. This gives $U = 10.5$, which, for a 50-generation MA experiment, corresponds to $\lambda = 0.21, E(s) = 0.033$, in agreement with BM estimates from the particular MUKAI *et al.* (1972) data set that were analyzed also using MD (see GARCÍA-DORADO *et al.* 1998). The distribution of the mutational deleterious effect s is a mixture of a gamma-distributed (s_{gamma}) variable and constant s_{mild} values, mixed in the proportions $U_{\text{pure}}/U:U_{\text{mild}}/U$, respectively. We used the σ_R^2 computed for the pure model, so that the Q value, including the mild contaminant class, is 24 or 2.4, instead of 20 or 2.

“Tiny contaminated” models: Analogously to the previous case, an additional random number (i_{tiny}) of mutations with constant tiny deleterious effect $s_{\text{tiny}} = 0.0025$ were simulated to accumulate. Thus, the distribution of deleterious effect s

is a mixture of a gamma-distributed variable (s_{gamma}) and constant s_{tiny} values, and the total deleterious mutation rate is $U = U_{\text{pure}} + U_{\text{tiny}}$. For each MA line simulated for the pure model $U = 0.5, \alpha = 2$ case, a tiny contaminated line was derived, its viability being computed as $v_T = v + 0.0025 i_{\text{tiny}}$, where i_{tiny} was sampled from a Poisson distribution with mean $U_{\text{tiny}} = 10$. We used the σ_R^2 computed for the pure model, but the Q value was virtually unaffected after including the genetic variance contributed by the tiny class.

Analyzing simulated data

Moments method estimates including BM estimates: For each MA data set, we computed the mean, \bar{v} , and variance, $V(v)$, of the observed average viability v over the 200 MA lines. The increase in between-line variance ΔV and the decline in mean viability ΔM , due to mutation accumulation, were computed as

$$\Delta V = V(v) - \sigma_R^2 \quad (2)$$

$$\Delta M = \bar{v}_c - \bar{v}, \quad (3)$$

where \bar{v}_c stands for the mean viability of the corresponding control. In real MA experiments, σ_R^2 is usually estimated through ANOVA from within-line repeated measurements with very many degrees of freedom (often on the order of thousands), its sampling error being minor compared to that of the estimates of the between-line component of the variance. Here, to avoid the simulation of within-line replicate measurements, we assume that σ_R^2 is known without error.

Note that ΔM and ΔV estimate $UE(s)$ and $UE(s^2)$, respectively, but they are not per-generation rates. The lower-bound estimate for U and upper-bound estimate for $E(s)$ can be calculated as $U_{\text{BM}} = \Delta M^2/\Delta V$ and $E(s)_{\text{BM}} = \Delta V/\Delta M$ (MUKAI *et al.* 1972). These bounds are denoted BM estimates.

MD estimates: This method uses the information contained in the empirical distribution function F_n of a sample of size n , and it has been shown to be usually more robust than maximum likelihood to departures from underlying assumptions (WOODWARD *et al.* 1984; CAO *et al.* 1995).

Basically, a model is assumed, and the theoretical distribution function F_θ of the sampled variable v (the mean viability assayed in each of the n MA lines) is derived as a function of a vector of parameters θ . The empirical and theoretical distributions are compared using a distance measure. We use the Cramer-von Mises distance. In general, the distance between distributions F_1 and F_2 is defined as

$$W^2(F_1, F_2) = \int_{-\infty}^{\infty} [F_1(v) - F_2(v)]^2 dF_2(v).$$

The MD estimator of θ is defined as the value of this parameter’s vector, which minimizes $W^2(F_n, F_\theta)$. The basic idea is to estimate the “true” value of θ as the one making the assumed model closest to the sampling information given by the empirical distribution (a general introduction to MD estimation can be found in TITTERINGTON *et al.* 1985).

A simple expression to evaluate the Cramer-von Mises distance $W^2(F_n, F_\theta)$ from an empirical distribution in a size n sample to a theoretical one, given by WOODWARD *et al.* (1984), is

$$W_n^2 = \frac{1}{12n} + \sum_{i=1}^n \left[F_\theta(v_i) - \frac{i - 0.5}{n} \right]^2,$$

where n is the sample size (the number of MA lines in our case), v_i is the i th order statistic in the sample of values of v , and the term $1/(12n)$ is irrelevant to MD estimation, but is necessary to have a W_n^2 statistic with a known, tabulated, distribution, so that credible intervals can be computed (see below).

This distance has been found to perform well in a number of situations (PARR and SCHUCANY 1988).

We compute the distance from the empirical distribution of our simulated experimental data to that expected under the pure model stated above with reflected-gamma-distributed effects. In computing the expected distributions, we introduce $\Delta V = UE(s^2)$, estimated from Equation 2, as a known constant. This prevents huge computation effort and minimization problems associated with the handling of too many parameters. This simplification is justified by the fact that reliable estimates of ΔV can usually be obtained through ANOVA from experimental data. For our simulated data, ΔV is estimated by subtracting the true residual variance from the observed variance of the MA lines' mean viability.

“Control-ignored” MD analysis: This is the more general analysis, where the control is assumed to be lacking or unreliable. Thus, no estimate of the viability decline $UE(s)$ based on Equation 3 is available. The three parameters, $\theta = (U, P_a, \alpha)$, are directly estimated through MD. Given the available ΔV estimate, these parameters determine $E(s^2)$ and, therefore, determine β , $E(s)$, and $\Delta M = UE(s)$.

We use a Fortran MD program (by A. García-Dorado, from a previous version by GARCÍA-DORADO and MARÍN 1998) that is available from the corresponding author. Since the distribution of the means of the lines expected under the assumed model is not analytically tractable, the program replaces it, for each $\theta = (U, P_a, \alpha)$ considered, with a distribution empirically computed from 10^4 simulated means. The program searches the minimum distance in a grid for U , P_a , and α as specified by the user. Checking a grid is a standard searching method, particularly on rough surfaces where the simplex algorithm could be trapped on local optima. The search is as follows. For each (U, P_a) pair, the distance is computed for all the α values in the grid, and the minimum distance is saved, together with the corresponding α value. This process is reiterated for each P_a value in the grid, and the minimum distance value is saved, together with the corresponding (α, P_a) pair. This was repeated for all the U values considered in the grid, giving a profile of the distance (minimized with respect to α and P_a) against U , where the overall minimum must be identified.

For U and α , we started with a grid from $\tau_i/20$ to $10\tau_i$, where τ_i is the corresponding true parameter value. For each of these parameters, the initial width of the grid steps was $\tau_i/20$. The grid covered P_a values from 0 to 0.1, with initial step width 0.0025. For all parameters, step width increased after each step by 2% of the parameter value at that cell of the grid. When, occasionally, minimum distance values corresponded to parameter values close to the grid edge, the grid was moved accordingly. When MD estimates had been obtained using this standard protocol, checking thinner grids just allowed the minimum to be more precisely located within the cell corresponding to the previous minimum, but gave no solution in different regions.

“Control-determined” MD analysis: The procedure is analogous to the previous control-ignored analysis, except that the viability decline (ΔM) is determined from the comparison between the mean viability of the MA and the control lines using Equation 3 and is introduced in the analysis as a known constant. This alternative has been used by GARCÍA-DORADO (1997). To determine MD estimates, a grid for (U, P_a) was established as before, each (U, P_a) pair determining a single (β, α) pair compatible with the ΔM and ΔV estimates from Equations 2 and 3.

“Control-supported” MD analysis: This procedure is analogous to the control-ignored analysis except that, for each (U, P_a) pair screened, α values giving $UE(s)$ outside some interval around the ΔM estimate from Equation 3 were not considered. Here we use the interval $\Delta M \pm 2 \text{ SE}(\Delta M)$, where $\text{SE}(\Delta M)$ is

the empirical standard error of ΔM in the MA data. Thus, MD estimates of U , α , and P_a [and, therefore, of $E(s)$] were constrained to give a ΔM estimate within a two-standard-error interval around the empirical estimate obtained using the control.

MD analysis assuming $P_a = 0$: This is similar to the control-ignored (CI)-MD analysis with P_a set at zero. Thus, Cramer-von Mises distances are computed through a bidimensional grid, each (U, α) pair in the grid determining a ΔM value.

ML estimates: These were obtained using a C program kindly supplied by P. D. Keightley, who has implemented the method as explained in KEIGHTLEY (1998). The basic mutational model is very similar to the one assumed in MD, although some differences must be pointed out. The main one is that the program needs both a sample of MA lines and a sample of control lines. Estimates maximize the joint likelihood of both samples and are therefore dependent upon the viability decline that would be estimated from comparison between MA and control lines (Equation 3). We have followed the procedure as described in KEIGHTLEY'S (1998) simulation study, which assumes $P_a = 0$, and we have maximized the profile likelihood against the shape parameter, as suggested by the author. This gives global ML estimates for the whole set of parameters (ML-W), μ_0 , σ_R^2 , U , α , and β , determining $E(s)$, $E(s^2)$, ΔM , and ΔV . To obtain results more directly comparable to our MD estimates (ML-C), we repeated the analysis, introducing σ_R^2 as a known constant. This also required us to introduce μ_0 as a constant equaling the observed control mean \bar{u} , which should have a negligible effect on the estimates, as ML-W estimates for μ_0 were in excellent agreement with \bar{u} . Another important difference is that maximization is not carried out by systematic screening of a grid, the simplex algorithm being used instead. As suggested by Keightley, we used different $[U, E(s)]$ starting points in the runs to prevent estimates corresponding to local maxima. In several cases, the likelihood monotonically increased as the shape parameter decreased to $\sim \alpha = 0.1$, which was associated with very large U values. When the likelihood continued to increase for U values up to $U = 50$, we classified the data set as giving no global estimates.

Mean square errors: For each parameter, the mean square error (MSE) of individual estimates was computed as the variance between replicate parameter estimates plus the squared bias of the over-replicate average estimate. This gives the MSE of estimates based on single MA data sets and, therefore, would apply to estimates from single MA experiments.

Simulated parameters and their biological meaning

The simulations are intended to test the properties of the estimation methods under plausible mutational parameter values and experimental conditions. The mutational parameters were chosen to allow comparison with published ML simulation results (see Table 5 in KEIGHTLEY 1998). Four basic cases were considered, defined by the combinations of two U (0.5, 5) and two α (0.5, 2) values.

It should be noted that U represents the final expected number of mutations accumulated per line ($U = t\lambda$ in chromosome MA lines). Thus, to interpret the simulated cases, we must keep in mind the timescale of published MA experiments, the corresponding λ and α estimates, and the observed association of high λ with low α (high kurtosis) values. For example, for *Drosophila* viability (see GARCÍA-DORADO *et al.* 1999), λ estimates range from ~ 0.01 per haploid genome (associated with a low to moderate kurtosis for the distribution of s , say $1 \leq \alpha \leq 3$) to ~ 0.5 per haploid genome (associated with $\alpha \leq 1$). The case $[U = 0.5, \alpha = 2]$ could be interpreted as a $t = 50$ MA experiment with $\lambda = 0.01$, where, for $\sim 90\%$

of the deleterious mutations, the deleterious effect is more than mild ($s > 0.05$). The case [$U = 5, \alpha = 0.5$] could be interpreted as a $t = 25$ MA experiment with $\lambda = 0.2$, where $\sim 90\%$ of the detectable deleterious mutations have only mild deleterious effects ($s < 0.05$). These two cases are approximately representative of typical published estimates of viability mutational parameters in *Drosophila* (see GARCÍA-DORADO *et al.* 1999), and they are more extensively explored.

As we use the scale parameter value (β) giving $E|s| = 0.1/U$, the average deleterious effect was 0.2 in the two cases with $U = 0.5$ and 0.02 in the two cases with $U = 5$. However, because selection was absent, the magnitude of deleterious effects is irrelevant to mutation accumulation. Since we set $Q = 20$ [where $Q = \sigma_c^2/\sigma_R^2$ and $\sigma_c^2 = \lambda E(s^2)$], the scale for s is fixed to give $E(s^2) = Q\sigma_R^2/\lambda$. In a real experiment, $\sigma_R^2 = \sigma_c^2/n_e$, where n_e is the number of individuals assayed per line, and σ_c^2 is the nongenetic variance of the fitness trait studied. Thus, using the standard definition of mutational heritability ($h_m^2 = \lambda E(s^2)/(2\sigma_c^2)$), we obtain $\sigma_R^2 = \lambda E(s^2)/(2n_e h_m^2)$, which implies $\sigma_c^2/\sigma_R^2 = 2n_e h_m^2$ and $n_e = Q/(2h_m^2)$.

FERNÁNDEZ and LÓPEZ-FANJUL (1996) obtained $Q = 2.12$ for egg-to-adult viability after 105 MA generations, and $Q \approx 3.6$ was obtained for competitive viability in the MUKAI *et al.* (1972) data set ($t = 40$) that has been analyzed by MD. We used $Q = 20$ (as assumed in most KEIGHTLEY 1998 simulations) and $Q = 2$ as reasonable bounds on the likely value of Q .

Numbers of replicates

For each case analyzed, 10 data sets were simulated, each consisting of 200 MA lines and 200 control lines. An exception is the case $U = 0.5, \alpha = 0.5, Q = 20$ (the first case we analyzed) for which two additional data sets were simulated after two analyses failed to produce global MD estimates. For the remaining cases no additional data were simulated to replace MD failures. Overall, 152 data sets were simulated, each subject to several analyses as shown in the tables. Some additional ML-W estimates were obtained from KEIGHTLEY (1998, see Table 5) and also correspond to 10 replicates (each with 200 MA and 200 control lines) per parameter set.

RESULTS

Pure model data: Results of different estimation methods (BM, MD, and ML) from pure model simulated data are given in Tables 1–5. ML-W estimates obtained by KEIGHTLEY (1998) are included for comparison. To allow direct comparison with Keightley’s ML-W results, we refer to the estimated expected absolute effect on the trait ($\hat{E}|s| = \hat{\alpha}/\hat{\beta}$, where the hat stands for estimation), relative to the true expected value ($E(s^*) = \hat{E}|s|/E|s|$). For cases simulated using $P_a = 0$, very similar estimates were obtained for $E(s)$ and $E|s|$.

BM underestimated U and overestimated $E(s^*)$. The biases were smaller for $\alpha = 2$ than for $\alpha = 0.5$, as expected since the BM bounds became closer to the true parameter with decreasing coefficient of variation of s , which is equal to $1/\sqrt{\alpha}$ in the gamma distribution.

MD estimates were generally less biased and had smaller MSE than BM estimates, the difference being larger for small α , as expected. The control-determined (CD)-MD analysis was not usually the best MD alternative, leading to larger MSE on the average. This short-

TABLE 1

Pure model results for case 1

Method	U	MSE ^{1/2} (U)	α	MSE ^{1/2} (α)	$E(s^*)^a$	MSE ^{1/2} ($E(s^*)$)	F^b
BM	0.17	0.33			3.32	2.54	0.00
CI-MD	0.53	0.15	0.72	0.60	1.04	0.31	0.17
CD-MD	0.58	0.25	1.07	1.35	1.11	0.49	0.17
CS-MD	0.50	0.13	0.69	0.39	1.12	0.30	0.08
ML-W ^c	0.95	1.32	0.52	0.52	0.90	0.45	0.00

Estimates (averaged over replicates), and the square root of their MSE, were obtained using different estimation methods on MA data simulated using the pure model with $U = 0.5, \alpha = 0.5$, and $Q = 20$.

^a $|s|$ relative to the true expected simulated value.

^b Fraction of samples that failed to provide a global minimum for the distance or a global maximum for the likelihood.

^c Estimates were obtained by KEIGHTLEY (1998) from 10 simulated MA lines and controls and are included here for comparison.

coming can be due to the limited precision of the ΔM estimate but is partially overcome if the MD estimate of ΔM is allowed to vary two standard deviations around the empirically estimated ΔM . Thus, the control-supported (CS)-MD analysis is a good choice to incorporate reliable empirical information on ΔM .

On the whole, MD showed no bias for U or $E(s^*)$ estimates. One exception is the case with $U = 5, \alpha = 0.5$, where an underestimate for U and an overestimate for $E(s^*)$ were obtained, the bias being always smaller than those observed for BM. Furthermore, the case $U = 5, \alpha = 0.5$ with $Q = 2$ is the only one where CS-MD estimates had MSE smaller than that of the CI-MD, for both U and $E(s)$. This is not surprising, as large U and small Q render the distribution of the means of the MA lines more similar to a normal curve, masking the high kurtosis of the distribution of effects and increasing the relevance of the information about ΔM . In any case, the bias was always relatively small.

For all the pure model cases where no advantageous mutations were simulated (those in Tables 1–4), MD estimates for P_a averaged 0.02. For cases with $Q = 20$, 90% of analyses gave $P_a < 0.05$, but with $Q = 2$ this dropped to 78% (results not shown). Empirical standard deviations were of the order of the corresponding estimates. Table 5 allows us to compare estimates for the case $U = 0.5, \alpha = 2, Q = 20$ for MA data simulated with $P_a = 0$ or $P_a = 0.1$. The quality of BM estimates decreases with $P_a > 0$, as expected. MD estimates for P_a were reasonable but had large standard deviations, indicating that P_a values of ~ 0.05 can pass undetected in an MD analysis. The MSEs of MD estimates for U, α , and $E(s^*)$ obtained by assuming no advantageous mutations [$P_a = 0$ (PA0)-MD in Table 5] were not smaller than those obtained in CI- or CS-MD analyses

TABLE 2
Pure model results for case 2

Estimation method	Q	U	$MSE^{1/2}(U)$	α	$MSE^{1/2}(\alpha)$	$E(s^*)^a$	$MSE^{1/2}(E(s^*))$	F^b
BM	20	0.34	0.20			1.54	0.69	0.00
CI-MD	20	0.46	0.10	2.34	1.01	1.05	0.19	0.00
CD-MD	20	0.55	0.14	2.47	2.16	0.93	0.25	0.10
CS-MD	20	0.50	0.13	1.88	0.76	0.97	0.19	0.00
ML- W^c	20	0.65	0.30	1.66	0.91	0.87	0.25	0.00
BM	2	0.36	0.19			1.37	0.61	
CI-MD	2	0.41	0.17	3.80	2.58	1.12	0.24	0.10
CD-MD	2	0.70	0.86	2.56	1.70	0.81	0.92	0.10
CS-MD	2	0.46	0.24	3.78	2.46	1.10	0.32	0.10
ML- W^c	2	0.67	0.44	3.89	5.20	0.90	0.36	0.10
ML- C^d	2	1.93	4.32	3.04	3.52	0.83	0.46	0.00

Estimates (averaged over replicates), and the square root of their MSE, were obtained using different estimation methods on MA data simulated using the pure model with $U = 0.5$ and $\alpha = 2$.

^a $|s|$ relative to the true expected simulated value.

^b Fraction of samples that failed to provide a global minimum for the distance or a global maximum for the likelihood.

^c Estimates for the case $\sigma_c^2/\sigma_R^2 = 20$ are from KEIGHTLEY (1998) and also correspond to 10 simulated MA lines and controls.

^d Estimates were obtained from our 10 simulated MA lines and controls using the true original mean and σ_R^2 as a known parameter.

for samples simulated with $P_a = 0$ and were larger when the samples had been simulated using $P_a = 0.1$. Furthermore, when data were simulated using $P_a = 0.1$, the analysis assuming $P_a = 0$ often produced no global estimates. Thus, the general analysis allowing the user to

estimate P_a should be preferred whenever there is uncertainty of whether P_a is zero.

ML analysis showed a tendency to overestimate U . This seems to be due to a few very large estimates, despite the fact that the maximum was not searched for

TABLE 3
Pure model results for case 3

Method	Q	U	$MSE^{1/2}(U)$	α	$MSE^{1/2}(\alpha)$	$E(s^*)^a$	$MSE^{1/2}(E(s^*))$	F^b
BM	20	1.61	3.40			3.08	2.12	0.00
CI-MD	20	3.14	2.41	1.76	1.96	1.85	1.24	0.00
CD-MD	20	3.35	1.98	1.13	0.78	1.60	0.78	0.40
CS-MD	20	4.28	2.40	0.95	0.74	1.40	0.76	0.10
ML- W^c	20	5.93	4.15	1.05	1.55	1.11	0.47	0.00
BM	2	1.68	3.37			3.21	4.36	
CI-MD	2	2.97	2.74	1.83	2.02	1.95	1.38	0.10
CD-MD	2	2.50	2.60	3.36	3.21	2.20	1.23	0.60
CS-MD	2	3.81	1.66	1.86	1.99	1.55	0.63	0.60
ML- W^c	2	10.30	18.44	1.39	1.55	1.66	1.09	0.50
ML- C^d	2	3.88	1.97	0.98	0.97	1.41	0.60	0.50

Estimates (averaged over replicates), and the square root of their MSE, were obtained using different estimation methods on MA data simulated using the pure model with $U = 5$ and $\alpha = 0.5$.

^a $|s|$ relative to the true expected simulated value.

^b Fraction of samples that failed to provide a global minimum for the distance or a global maximum for the likelihood.

^c Estimates for the case $\sigma_c^2/\sigma_R^2 = 20$ are from KEIGHTLEY (1998) and also correspond to 10 simulated MA lines and controls.

^d Estimates were obtained from our 10 simulated MA lines and controls using the true original mean and σ_R^2 as known parameter.

TABLE 4
Pure model results for case 4

Method	U	$MSE^{1/2}(U)$	α	$MSE^{1/2}(\alpha)$	$E(s^*)^a$	$MSE^{1/2}(E(s^*))$	F^b
BM	3.33	1.74			1.52	0.57	0.00
CI-MD	4.97	1.6	2.15	1.80	1.00	0.30	0.00
CD-MD	6.56	2.36	3.20	3.61	0.85	0.29	0.10
CS-MD	7.06	2.92	2.30	2.16	0.80	0.36	0.00
ML-W ^c	7.70	6.73	2.98	3.65	0.87	0.33	0.10

Estimates (averaged over replicates), and the square root of their MSE, were obtained using different estimation methods on MA data simulated using the pure model with $U = 5$, $\alpha = 2$, and $Q = 20$.

^a $|s|$ relative to the true expected simulated value.

^b Fraction of samples that failed to provide a global minimum for the distance or a global maximum for the likelihood.

^c Estimates were obtained by KEIGHTLEY (1998) from 10 simulated MA lines and controls and are included here for comparison.

$U > 50$. Thus, one of the ML-C analyses in Table 2 gave $U = 13$, and one ML-W analysis in Table 3 gave $U = 42$, these cases being insensitive to the starting position in the search algorithm. This could be ascribed to the finding of maxima in regions where the likelihood profile is too flat, which might be due to the use of the simplex algorithm to maximize the likelihood. Again, the exception is the case $U = 5$, $\alpha = 0.5$ (Table 4), where ML-W overestimates $E(s^*)$ and, after the outliers are removed, underestimates U , as occurred in the MD analysis.

ML estimates for U had MSEs larger than those obtained using CI- or CS-MD estimation (Tables 1–4). For example, the average over all ML estimates in Tables 1–4 of the relative $MSE^{1/2}$ (*i.e.*, the average of the “ $MSE^{1/2}/$

true parameter value”) is 5.95 times larger than that for CS-MD. However, Keightley’s ML program and our MD program do not allow the comparison of both methods under identical conditions. The ML program maximizes the joint likelihood of the MA and control samples, allowing the user to specify μ_0 and σ_R^2 as known values (ML-C analyses). Then, the estimate for ΔM is conditioned on, but not determined by, the assumed μ_0 value. The MD program uses only the MA sample, although we can incorporate information from the control mean (\bar{v}_c) to bound the ΔM estimate or to set this at a given constant (CS or CD alternatives). Since in the MD analysis reported here we have introduced the true σ_R^2 as known, the fair comparison is that between ML-C (where σ_R^2 is set at its true value and μ_0 is set at \bar{v}_c) and a MD alternative that incorporates information about \bar{v}_c without determining a fixed ΔM . One such alternative (not necessarily the best one) is the CS-MD analysis using a 2-SE interval around ΔM . Relative $MSE^{1/2}$ averaged over cases in Tables 2 and 3 gives 4.51 for ML-C estimates of U and 0.39 for CS-MD ones, so that average relative $MSE^{1/2}$ for U is 11 times larger for ML-C than for CS-MD estimates. Removing an outlier ML-C estimate giving $U = 13$, the average relative $MSE^{1/2}$ for U is just 4.6 times larger for ML-C than for CS-MD estimates. $MSE^{1/2}$ ’s for $E|s|$ using ML or CS-MD were roughly similar on average.

Although for a fair comparison with the MD analysis of our simulated data the appropriate choice is ML-C, the difference between ML-W and ML-C was negligible when they were computed for the same data sets.

All this suggests that, when the analyzed data match the model assumed in the MD or ML analysis, MD allows good estimation in the absence of a control assay and that differences between ML and MD estimates could be due to differences in the number of estimated parameters, the nature of the data used (MA and control *vs.* only MA data), and/or the searching algorithms.

TABLE 5
Effect of the P_a value on the estimates

Method	True P_a	MD P_a	$MSE^{1/2}(P_a)$	U	$MSE^{1/2}(U)$	α	$MSE^{1/2}(\alpha)$	$E(s^*)^a$	$MSE^{1/2}(E(s^*))$	F^b
BM	0			0.34	0.20			1.51	0.69	
CI-MD	0	0.000	0.001	0.46	0.10	2.34	1.01	1.05	0.19	0.00
CD-MD	0	0.012	0.034	0.55	0.14	2.47	2.16	0.93	0.25	0.10
CS-MD	0	0.004	0.012	0.50	0.13	1.88	0.76	0.97	0.19	0.00
PA0-MD	0	—	—	0.47	0.12	2.58	1.27	1.05	0.21	0.00
BM	0.1			0.22	0.28			1.90	0.94	
CI-MD	0.1	0.073	0.060	0.51	0.18	2.50	1.56	1.04	0.25	0.10
CD-MD	0.1	0.080	0.052	0.48	0.11	2.22	0.75	1.05	0.18	0.00
CS-MD	0.1	0.075	0.058	0.48	0.16	2.53	1.25	1.08	0.25	0.10
PA0-MD	0.1	—	—	0.54	0.26	1.36	0.87	0.97	0.34	0.40

Estimates (averaged over replicates), and the square root of their MSE, were obtained using different MD estimation methods on MA data simulated using different P_a values (pure model with $U = 0.5$, $\alpha = 2$, and $Q = 20$).

^a $|s|$ relative to the true expected simulated value.

^b Fraction of samples that failed to provide a global minimum for the distance.

TABLE 6
Estimates for residual contaminated data

Case	Q	Method	U	$MSE^{1/2}(U)$	α	$MSE^{1/2}(\alpha)$	$E(s^*)^a$	$MSE^{1/2}(E(s^*))$	F^b
$U = 0.5, \alpha = 2$	20	BM	0.34	0.17			1.43	0.46	
		CI-MD	0.51	0.09	2.63	1.37	1.00	0.14	0.00
		CS-MD	0.49	0.09	2.81	1.59	1.02	0.12	0.00
		ML-C ^c	0.51	0.10	2.17	0.84	0.98	0.16	0.00
$U = 0.5, \alpha = 2$	2	BM	0.34	0.20			1.50	0.58	
		CI-MD	0.43	0.10	4.58	2.73	1.16	0.20	0.30
		CS-MD	0.56	0.25	4.27	2.69	1.07	0.21	0.00
		ML-C ^c	0.50	0.37	3.26	2.26	1.22	0.50	0.00

Estimates (averaged over replicates), and the square root of their MSE, were obtained using different estimation methods on MA data simulated using the residual contaminated model for the two more representative parameter sets.

^a $|s|$ relative to the true expected simulated value.

^b Fraction of samples that failed to provide a global minimum for the distance or a global maximum for the likelihood.

^c Estimates were obtained from our 10 simulated MA lines and controls using the true original mean and σ_R^2 as known parameter.

Both ML and MD showed a trend to moderately overestimate α , although, in all cases, the true parameter value was included within two standard deviations around the corresponding estimate. Therefore, individual estimates are not expected to significantly differ from the true parameter values (note that the standard deviation between replicate estimates would correspond to the standard error of individual estimates obtained in a single data set or experiment). For both MD and ML analyses, the U and $E(s^*)$ estimates were largely insensitive to α above some threshold value around 10 or 20. It should be remembered that both the kurtosis coefficient and the coefficient of variation of a gamma distribution asymptotically decrease with increasing α , so that large α differences (or large MSE for α estimates) can be scarcely relevant for large α values. In a few data sets, the ML profile plotted against α increased up to the maximum value we used to build the ML profile ($\alpha = 20$). For these cases, a constant-effect model (equivalent to $\alpha = \infty$) showed a slightly larger maximum likelihood than the $\alpha = 20$ gamma model showed, but the estimates for U and $E(s)$ remained the same under both models (gamma or constant effects). These data sets were used to compute the average ML estimates for U and $E(s^*)$ but are excluded in the computation of the average α estimate.

A problem common to both ML and MD analyses is that, on occasion, they do not provide a global maximum for the likelihood (or a global minimum for the distance), corresponding to a set of parameter estimates. This occurred with similar frequencies under CI-MD, CS-MD, ML-W, or ML-C, but in different data sets. It was more common in CD-MD analysis. In those occasions, the distance profile decreased (or the likelihood profile increased) with U and then became flat. This

was more likely to occur under large U and small Q , when the distribution of the trait mean in the lines is expected to be closer to normality and, therefore, less informative.

Effect of nonnormal residual errors: The empirical distribution of the per-vial random deviations (d) had skewness coefficient -0.47 . This distribution and that for R (for $n_e = 3$) significantly departed from normality ($p < 0.1\%$).

Table 6 gives results of BM, MD (CI and CS), and ML-C analyses on data simulated following the residual contaminated model, for the two more representative cases. All but one of the estimates (BM, MD, or ML) showed smaller MSE than those obtained under the analogous pure model, irrespective of the U , α , or Q values used to simulate the data. Although we do not have an explanation for this unexpected result, it may be due to sampling error, as there happened to be no outliers in the case of ML-C, which partly explains the lower MSE.

Detection of mutations with small deleterious effect: With gamma-simulated deleterious effects, the case $U = 5$, $\alpha = 0.5$ implies a high rate of mutation with mild deleterious effects ($U_{s < 0.05} = 4.34$). CI-MD as well as both ML analyses detects more than half of these (Table 3). On average, MD detected $\sim 70\%$ of the simulated deleterious mutation rate, ML gave overestimates due to outlier U values, and BM estimated a mutation rate $\sim 30\%$ of the simulated one. Both ML and MD estimates of the average effect were more accurate than BM estimates.

When constant-effect mildly deleterious mutations were added to the $U = 0.5$, $\alpha = 2$ pure model case (mild contaminated model), both the BM and the CI-MD analyses underestimate U and overestimate $E|s|$ (Table 7). BM detects $\sim 35\%$ of the total mutation rate,

TABLE 7
Estimates for mild contaminated data

Method	Q	U	$MSE^{1/2} (U)$	α	$SD(\alpha)$	$E s $	$MSE^{1/2} (E s)$	F^a
True value		10.5				0.0333		
BM	24.2	3.36	7.15	—	—	0.10	0.07	0.00
CI-MD	24.2	7.50	5.50	0.28	0.23	0.04	0.02	0.10
CD-MD	24.2	—	—	—	—	—	—	1.00
CS-MD	24.2	—	—	—	—	—	—	1.00
ML-W ^b	24.2	—	—	—	—	—	—	1.00
ML-C ^c	24.2	—	—	—	—	—	—	1.00
BM	2.4	3.89	6.63	—	—	0.09	0.06	—
CI-MD	2.4	1.55	9.10	1.89	1.79	0.13	0.11	0.10
CD-MD	2.4	—	—	—	—	—	—	1.00
CS-MD	2.4	—	—	—	—	—	—	1.00
ML-C	2.4	—	—	—	—	—	—	1.00

Estimates (averaged over replicates) for the different parameters and the square root of their MSE from MA data simulated using the mild contaminated models for $U = 0.5$ and $\alpha = 2$ are shown (SDs are given instead of MSE in the case of α , as the expected α value is not defined for the mild contaminated model).

^a Fraction of samples that failed to provide a global minimum for the distance or a global maximum for the likelihood.

^b Estimates, including those of the original average and σ_R^2 , obtained from our 10 simulated MA lines and controls.

^c Estimates obtained from our 10 simulated MA lines and controls using the true original mean and σ_R^2 as known parameter.

irrespective of the data precision (*i.e.*, for both $Q = 24$ and $Q = 2.4$), although it estimates an average effect ($E(s) = 0.10$) that is very different from the mild contaminant effect ($E(s) = 0.025$), from the average effect of the uncontaminated gamma distribution ($E(s) = 0.2$), and from the true average of the contaminant-gamma mixed distribution ($E(s) = 0.0333$). In contrast, the quality of MD estimates depends upon the data assay precision. For large Q , CI-MD estimates were clearly better than BM estimates. Besides showing considerably smaller MSE, CI-MD detected $\sim 71\%$ of the total mutation rate and gave good estimates of the average deleterious effect. Although the constant-effect mild contaminants are not modeled by the gamma distribution, CI-MD estimated a gamma distribution where the frequency of the mild-effect class accounts for $\sim 50\%$ of the contaminant “mild” mutations, and this was attained making no use of the control-based ΔM estimate. Due to the assumption of gamma-distributed effects, this MD detection of a fraction of mild contaminant mutations was obtained at the expense of overestimating the rate of moderately deleterious mutation (see Figure 1). With low Q , CI-MD becomes even less efficient than BM, detecting only $\sim 15\%$ of the total mutation rate, BM and MD estimates for U being of the same order of magnitude. Surprisingly, when the control-based ΔM estimate was incorporated (*i.e.*, in CD or CS-MD analysis) none of the samples provided global point estimates for the parameters (*i.e.*, no global minimum for the distance was found). The same occurred with ML esti-

mates, which depend upon the control average and, therefore, upon the control-based ΔM estimate. The fact that no estimates are found when $UE(s)$ is forced to its true value suggests that no gamma distribution both reasonably fits the gamma contaminated distribution and produces the true overall $UE(s)$ value.

For the tiny contaminated model (Table 8), BM, CI-MD, and CS-MS did not allow us to detect the additional rate of tiny mutations, irrespective of the assay precision. Estimates constrained to the observed ΔM (CD-MD or ML-C) account for larger fractions of the tiny-contaminant class when assay precision is large, suggesting that gamma distributions with low α may account better for tiny contaminant effects than for mild effects. However, although ML-C estimates of U and $E|s|$ are reasonable when $Q = 20$, the difference between methods is scarcely relevant due to the very large $MSE^{1/2}$ values. These values were about as large as or larger than the true parameter. Therefore, estimates do not significantly differ from zero, from the true value, or from the estimate obtained using an alternative method. Furthermore, about one-half the cases provided no global estimates.

DISCUSSION

For pure model simulated data, both ML and MD estimates for the total mutation rate (U) and the average deleterious effect ($E|s|$) were less biased than BM estimates, as expected. CI-MD was among the better choices, despite being computed without making use of control

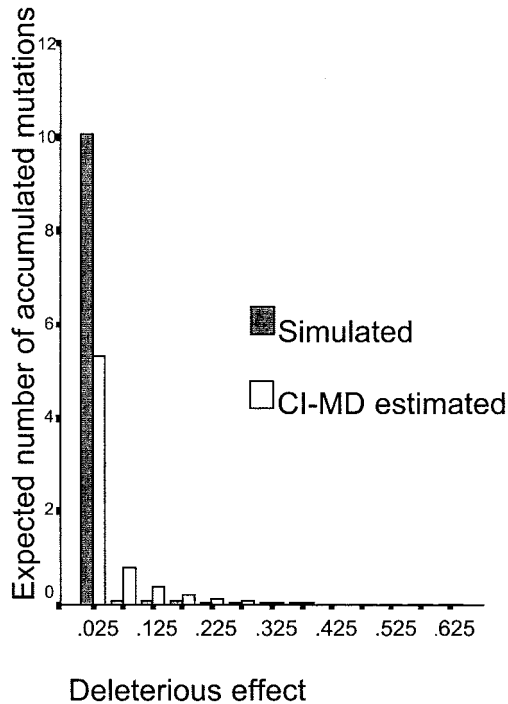


FIGURE 1.—Expected number of mutations accumulated per line plotted against deleterious effect (grouped in 0.05 width classes) for the “mild contaminated” model.

information. To incorporate the control-based estimate of the mutational rate of viability decline (ΔM), CS-MD estimates should be preferred to CD-MD ones, since the latter are associated with larger MSE and with a larger proportion of distance profiles showing no minimum. Although the MD power to estimate the rate of favorable mutation (P_a) is poor, it is convenient to simultaneously estimate this parameter to prevent any bias in the estimates of α , U , and $E(s)$. This result is model dependent, and we have not checked the behavior of our MD estimates in cases where favorable effects are not obtained from a reflected-gamma distribution. SHAW *et al.* (2002) have recently published a ML alternative for a model where favorable mutations are simulated by displacing a gamma distribution of deleterious effects. Simulation showed that the method is efficient when the displacement is known. Therefore, information on the distribution of favorable mutations for fitness traits (which are not necessarily favorable for overall fitness) is required to choose the more appropriate estimation model.

Pure model ML estimates occasionally gave large outlier U values, causing $\text{MSE}^{1/2}$ and SD for U up to 10 times those for MD estimates and considerably exceeding the BM ones. As ML is asymptotically efficient under some regularity conditions (VAN DER VAART 1998, pp. 65 and 120), larger SD values were unexpected. However, it should be noted that Keightley and Bataillon, using a constant-effect ML version to analyze data that had been simulated with a gamma model,

obtained MSE values for ML estimates of U that were slightly smaller than those for BM estimates (see cases evaluated at generations 0 and 80 in Table 2 from KEIGHTLEY and BATAILLON 2000). Thus, a ML analysis based on a simpler, although less appropriate, model may work better. This suggests that the common occurrence of outliers in estimates of U obtained by ML under the gamma model may be due to difficulties in maximizing over too many parameters and/or in the search algorithm used in the programs.

Both ML and MD analyses gave reasonable estimates for the shape parameter α , although these estimates were highly variable with large α values, exhibiting an upward bias and high MSE. However, U and $E(s)$ estimates were very insensitive to variation in α once α is large. This could be ascribed to the fact that the shape of the gamma distribution becomes quite stable to variation in α when α is large.

For pure model data, the precision of the estimates is not greatly affected by the precision with which the trait mean is assayed ($Q = \sigma_G^2/\sigma_R^2$). However, the probability of finding MD or ML global estimates with $Q = 20$ is, on average, about four times larger than that with $Q = 2$.

For ML estimates, asymptotic 95% support limits can be computed on the basis of a drop in natural log likelihood of two from the maxima in the likelihood profiles (see KEIGHTLEY 1994). The distribution of the Cramer-von Mises distance (W^2) between the empirical (F_n) and the true distribution (F_θ) does not depend on F_θ whenever F_θ is continuous (*i.e.*, it is distribution free), but only on the sample size n . Since, for $n > 20$, the probability of a distance $W^2 > 0.46$ is ~ 0.05 (SCOTT 2000), the set of parameter values associated with $W^2 < 0.46$ define a 95% credible interval, analogous to that defined by the ML support limits. This can be straightforwardly computed from the distance profile. In both cases (ML and MD), credible intervals are very wide, often being of infinite length. However, it should be noted that these credible intervals are not standard confidence intervals, but intervals of parametric values that could not be rejected in a goodness-of-fit test. Standard confidence intervals do not depend on the distribution of the W^2 distance itself, but on the distribution of the MD estimates (the $\hat{\theta}$ value that minimizes W^2) as a function of the true parameter value (θ). Unfortunately, the distribution of MD estimates is unknown, although it can be approached using bootstrap. Analogously, a confidence interval for ML estimates would not depend on the distribution of the likelihood itself, but on the distribution of the ML estimates. If, for a given true θ , MD (or ML) $\hat{\theta}$ estimates have a high probability of being close to θ , confidence intervals should be small. However, W^2 could still be relatively small (or the likelihood relatively high) for parameter values well outside the confidence interval, thus giving a credible interval much larger than the confidence interval. Furthermore, it should be noted

TABLE 8
Estimates for tiny contaminated data

Method	Q	U	$MSE^{1/2} (U)$	α	$SD(\alpha)$	$E s $	$MSE^{1/2} (E s)$	F^a
True value		10.5				0.0119		
BM	20	0.52	9.98	—	—	0.24	0.23	0.00
CI-MD	20	0.51	9.99	1.95	0.92	0.19	0.18	0.00
CD-MD	20	2.13	8.45	0.39	0.12	0.06	0.05	5.00
CS-MD	20	0.58	9.92	1.63	0.75	0.17	0.17	0.10
ML-C	20	7.66	9.48	0.18	0.19	0.04	0.04	0.50
BM	2	0.60	9.90			0.21	0.20	
CI-MD	2	0.39	10.11	4.21	0.52	0.23	0.22	0.30
CD-MD	2	1.77	8.76	1.49	2.15	0.09	0.11	0.60
CS-MD	2	0.86	9.65	1.45	0.41	0.13	0.13	0.50
ML-C	2	3.32	8.19	1.30	1.75	0.09	0.11	0.30

Estimates (averaged over replicates) for the different parameters and the square root of their MSE from MA data simulated using the tiny contaminated models for $U = 0.5$ and $\alpha = 2$ (SDs are given instead of MSE in the case of α , as the expected α value is not defined for the tiny contaminated model).

^a Fraction of samples that failed to provide a global minimum for the distance or a global maximum for the likelihood.

that the asymptotic validity of ML errors (as well as the ML asymptotic efficiency) depends on specific assumptions that are not trivial (VAN DER VAART 1998, pp. 65 and 120) and have not been checked even under the pure model. Even if those assumptions were met, the validity of MD or asymptotic ML credible intervals depends on the variable being drawn from the distribution assumed by the model. All this considered, errors or confidence intervals based on bootstrapping, which are more directly interpretable, should be preferred if the corresponding computational effort can be undertaken. It would be worth exploring the properties of bootstrapping in a future study.

Using sampling residuals estimated from *Drosophila* viability assays (instead of normally distributed ones) did not have an appreciable effect on the estimates and their MSEs (BM, MD, or ML), even when the residual variance was large ($Q = 2$) and only $n_e = 3$ single-female sampling errors d were averaged per MA line (d , estimated from FERNÁNDEZ and LÓPEZ-FANJUL 1996 data). This result suggests that MD estimates based on the Fernández and López-Fanjul data (GARCÍA-DORADO 1997) are unlikely to be biased due to nonnormal residuals, as $n_e = 12$ d values were averaged in those data to assay the viability of each line. Thus, due to the central limit theorem, they should have a distribution closer to a normal curve. However, this conclusion cannot be extrapolated to residuals with other unknown nonnormal distributions, and strongly asymmetric distributions may induce estimation bias.

It has been proposed that detecting common deleterious mutations with small effect requires an estimate of the rate of mean decline ΔM obtained from the

evaluation of a reliable control population (see Equation 3) and that this might explain why CI-MD mutation rates estimated from the MUKAI *et al.* (1972) data by ignoring the control information were only 8% of the BM estimates obtained using such ΔM (GARCÍA-DORADO *et al.* 1999; LYNCH *et al.* 1999). However, for the pure model with $U = 5$, $\alpha = 0.5$, and $E(s) = 0.02$ (which gives a mild mutation rate of $U_{s<0.05} = 4.35$), CI-MD average estimates for $U_{s<0.05}$ (2.35 or 2.18 for $Q = 20$ or $Q = 2$, respectively) were larger than the BM bounds for total U (1.61 or 1.68 for $Q = 20$ or $Q = 2$, respectively). Furthermore, the rate of mean decline (ΔM_{CI-MD}) predicted from CI-MD is fairly accurate (94 and 90% of the simulated decline for $Q = 20$ or $Q = 2$, respectively).

On the other hand, it has been noted (GARCÍA-DORADO 1997) that a high rate of mutations with “small” deleterious effects, not fitting the gamma distribution assumed for s , could pass undetected in a CI-MD analysis. To check this possibility, MA data were simulated using mild contaminated or tiny contaminated models, where the distribution of deleterious effects departed from a gamma distribution due to a very common class of deleterious mutations with constant small effect. It must be noted that even the continuity assumption is violated in this contaminated model.

For the mild contaminated model, BM detected only ~35% of the deleterious mutation rate, and the CD-MD or ML-C methods, constrained to produce exactly the observed change in the mean relative to the control, usually failed to produce global estimates for all the parameters. In contrast, a substantial fraction of mild contaminants is detected by CI-MD (*i.e.*, in the absence of control information) when the means of the MA lines

were assayed with high precision ($Q = 20$). In this case, CI-MD analysis detects 71% of the simulated mutation rate, estimates a gamma distribution of s close to the true mild contaminant-gamma mixed distribution (Figure 1), and accounts for more than half of the mutational decline in mean. With low precision data ($Q = 2$), CI-MD becomes even less efficient than BM, detecting only $\sim 15\%$ of the deleterious mutation rate. Even then, BM and CI-MD estimates are of the same order ($U_{\text{BM}}/U_{\text{CI-MD}} = 2.5$). This suggests that the large differences between these estimates found for MUKAI *et al.* (1972) data ($U_{\text{BM}}/U_{\text{CI-MD}} = 13$) should not be ascribed to CI-MD missing mild deleterious effects not fitting into the gamma distribution (GARCÍA-DORADO *et al.* 1998).

The above results show the convenience of checking the agreement between the CI-MD-estimated $UE(s)$ value and a ΔM estimate on the basis of comparison to a reliable control, which would support the validity of the MD estimates. This is particularly useful for low-precision data. Even when the MA means have been assayed with low precision, the findings of CD-MD estimates and their agreement with the CI-MD ones, as was the case in the analysis of the Fernández and López-Fanjul MA data (GARCÍA-DORADO 1997), suggest that the gamma model provides an adequate description of the distribution of deleterious effects, including mild ones. Finally, if CS-MD estimates can be obtained, they appear to be the best way to incorporate control information into the analysis.

No method, however, allowed reliable detection of a contaminant class of mutations with tiny effects. Although ML appeared to obtain reasonable average estimates when the trait is assayed with high precision ($Q = 20$), individual estimates were highly variable, giving $\text{MSE}^{1/2}$ about the true parameter value being estimated. This confirms the view that MA experiments should not be intended to investigate mutations with extremely small effects (GARCÍA-DORADO 1997; KEIGHTLEY and EYRE-WALKER 1999).

Our conclusions refer to the analysis of data obtained from a single assay of a fitness trait in a MA experiment. DENG *et al.* (1999) found that BM estimates from a single assay in 10-generation MA experiments achieved about the same quality as those from longer MA experiments. However, this conclusion depends on the particular mutational parameter values used in those simulations giving relatively large rates of mean decline, which can be efficiently estimated after a short MA period. On the other hand, ML and MD make use of the information contained in the observed shape of the distribution of the means of the lines $f(v)$, which approaches normality as the expected number of mutations accumulated per line ($U = t\lambda$) increases. After sufficiently long MA periods, $f(v)$ becomes virtually normal and uninformative. At that time, neither ML nor MD global estimates are found, and only BM estimates can be obtained. There-

fore, the optimal MA period depends on the mutation rate.

In the absence of *a priori* information on the true deleterious mutation rate, it could be convenient to assay the trait more than once during the MA experiment, until estimates can be obtained with reasonable precision. Incorporating data from multiple generations in an ML framework has proven to be more efficient than incorporating data from single-generation ML for a constant s model (KEIGHTLEY and BATAILLON 2000). However, the gain in precision should be weighed against the increased experimental effort required for repeated evaluation. A multigeneration ML method for a gamma- s model has been developed by SHAW *et al.* (2002). Although the procedure is cumbersome, it might be worth studying the relative efficiency of a multigeneration analysis.

We have found that single-assay data, analyzed by CI-MD and CS-MD, can provide accurate information if the trait is assayed when most lines have accumulated a small number of deleterious mutations. This requires MA lines and control (if available) being assayed as precisely as possible at the end of the MA period. These CI-MD and CS-MD estimates have relatively small MSEs and are relatively robust against contamination with mutations of mild deleterious effects. ML can also be used if a reliable control assay is available, but some criterion should be developed to indicate whether the ML surface is too flat, in which case the parameter estimates should be ignored. Bootstrap errors can be obtained for MD or ML estimates. If both MD and ML fail to provide global estimates, BM bounds should be computed.

We are indebted to P. D. Keightley for kindly providing his ML program and to J. F. Crow, C. López-Fanjul, and S. Otto for useful discussions of the manuscript. This work was supported by grant PB98-0814-C03-01 from the Ministerio de Educación y Cultura, Spain.

LITERATURE CITED

- CAO, R., A. CUEVAS and R. FRAIMAN, 1995 Minimum distance density-based estimation. *Comput. Stat. Data Anal.* **20**: 611–631.
- CHAVARRÍAS, D., C. LÓPEZ-FANJUL and A. GARCÍA-DORADO, 2001 The rate of mutation and the homozygous and heterozygous mutational effects for competitive viability: a long-term experiment with *Drosophila melanogaster*. *Genetics* **158**: 681–693.
- DENG, H.-W., J. LI and J.-L. LI, 1999 On the experimental design and data analysis of mutation accumulation experiments. *Genet. Res.* **73**: 147–164.
- FERNÁNDEZ, J., and C. LÓPEZ-FANJUL, 1996 Spontaneous mutational variances and covariances for fitness-related traits in *Drosophila melanogaster*. *Genetics* **143**: 829–837.
- GARCÍA-DORADO, A., 1997 The rate and effects distribution of viability mutation in *Drosophila*: minimum distance estimation. *Evolution* **51**: 1130–1139.
- GARCÍA-DORADO, A., and J. M. MARÍN, 1998 Minimum distance estimation of mutational parameters for quantitative traits. *Biometrics* **54**: 1097–1114.
- GARCÍA-DORADO, A., J. L. MONEDERO and C. LÓPEZ-FANJUL, 1998 The mutation rate and the distribution of mutational effects of viability and fitness in *Drosophila melanogaster*. *Genetica* **102/103**: 255–265.
- GARCÍA-DORADO, A., C. LÓPEZ-FANJUL and A. CABALLERO, 1999

- Properties of spontaneous mutations affecting quantitative traits. *Genet. Res.* **74**: 341–350.
- KEIGHTLEY, P. D., 1994 The distribution of mutation effects on viability in *Drosophila melanogaster*. *Genetics* **138**: 1315–1322.
- KEIGHTLEY, P. D., 1998 Inference of genome-wide mutation rates and distributions of mutation effects for fitness traits: a simulation study. *Genetics* **150**: 1283–1293.
- KEIGHTLEY, P. D., and T. M. BATAILLON, 2000 Multigeneration maximum-likelihood analysis applied to mutation-accumulation experiments with *Caenorhabditis elegans*. *Genetics* **154**: 1193–1201.
- KEIGHTLEY, P. D., and A. EYRE-WALKER, 1999 Terumi Mukai and the riddle of deleterious mutation rates. *Genetics* **153**: 515–523.
- KEIGHTLEY, P. D., and O. OHNISHI, 1998 EMS-induced polygenic mutation rates for nine quantitative characters in *Drosophila melanogaster*. *Genetics* **148**: 753–766.
- LYNCH, M., J. BLANCHARD, D. HOULE, T. KIBOTA, S. SCHULTZ, 1999 Spontaneous deleterious mutation. *Evolution* **53**: 645–663.
- MUKAI, T., S. I. CHIGUSA, L. E. METTLER and J. F. CROW, 1972 Mutation rate and dominance of genes affecting viability in *Drosophila melanogaster*. *Genetics* **72**: 333–355.
- PARR, W. C., and W. R. SCHUCANY, 1988 Minimum distance and robust estimation. *J. Am. Stat. Assoc.* **75**: 616–624.
- SCOTT, W. F., 2000 Tables of the Cramer-von Mises distributions. *Commun. Stat. Theory Methods* **29**: 227–235.
- SHAW, F. H., C. J. GEYER and R. G. SHAW, 2002 A comprehensive model of mutations affecting fitness and inferences for *Arabidopsis thaliana*. *Evolution* **56**: 453–463.
- TITTERINGTON, D. M., A. F. M. SMITH and U. E. MARKOV, 1985 *Statistical Analysis of Finite Mixture Distributions*. John Wiley & Sons, New York.
- VAN DER VAART, A. W., 1998 *Asymptotic Statistics*. Cambridge University Press, Cambridge, UK.
- VASSILIEVA, L. L., A. M. HOOK and M. LYNCH, 2000 The fitness effects of spontaneous mutations in *Caenorhabditis elegans*. *Evolution* **54**: 1234–1246.
- WOODWARD, W. A., W. C. PARR, W. R. SCHUCANY and H. LINDSLEY, 1984 A comparison of minimum distance and maximum likelihood estimation of a mixture proportion. *J. Am. Stat. Assoc.* **79**: 590–598.

Communicating editor: S. P. OTTO

