Genet. Res., Camb. (2003), **82**, *pp.* 191–205. *With* 4 *figures.* © 2003 Cambridge University Press
DOI: 10.1017/S0016672303006499 Printed in the United Kingdom

Estimating numbers of EMS-induced mutations affecting life history traits in *Caenorhabditis elegans* in crosses between inbred sublines

DANIEL L. HALLIGAN[1], ANDREW D. PETERS[1,2] AND PETER D. KEIGHTLEY[1]*
[1] *Institute of Cell, Animal and Population Biology, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, UK*
[2] *Department of Zoology, University of British Columbia, 6270 University Blvd, Vancouver, BC, V6T 1Z4, Canada*

(*Received 18 June 2003 and in revised form 19 September 2003*)

Summary

Inbred lines of the nematode *Caenorhabditis elegans* containing independent EMS-induced mutations were crossed to the ancestral wild-type strain (N2). Replicated inbred sublines were generated from the F1 offspring under conditions of minimal selection and, along with the N2 and mutant progenitor lines, were assayed for several fitness correlates including relative fitness (w). A modification of the Castle–Wright estimator and a maximum-likelihood (ML) method were used to estimate the numbers and effects of detectable mutations affecting these characters. The ML method allows for variation in mutational effects by fitting either one or two classes of mutational effect, and uses a Box–Cox power transformation of residual values to account for a skewed distribution of residuals. Both the Castle–Wright and the ML analyses suggest that most of the variation among sublines was due to a few ($\sim$1·5–2·5 on average) large-effect mutations. Under ML, a model with two classes of mutational effects, including a class with small effects, fitted better than a single mutation class model, although not significantly better. Nonetheless, given that we expect there to be many mutations induced per line, our results support the hypothesis that mutations vary widely in their effects.

1. Introduction

Several important evolutionary phenomena have been hypothesized to be consequences of recurrent deleterious mutation. These include inbreeding depression (Charlesworth & Charlesworth, 1987), the evolution of sex and recombination (Kondrashov, 1988; Charlesworth, 1990), the evolution of mating systems (Charlesworth *et al*., 1990), ecological specialization (Kawecki *et al*., 1997), genetic variability for quantitative traits (Bulmer, 1989), senescence (Charlesworth, 1994) and the extinction of small populations (Lande, 1994; Lynch *et al*., 1995*b*). It has been suggested that mutation accumulation might even threaten the persistence of our own species (Muller, 1950; Kondrashov, 1995; Crow, 1997). Whether or not mutations play a role in these phenomena critically depends on parameters associated with mutations (Turelli, 1984; Caballero & Keightley, 1994), including the genomic mutation rate (U), the distribution of selection coefficients (s) and dominance coefficients (h) of new mutations.

With theory increasingly showing the potential importance of the properties of mutations, there has been a resurgence of interest in attempting to estimate U and mean s and h. Although inferring the distribution of mutation effects has received less attention (Lynch *et al*., 1999), the distribution of effects is important for several reasons. First, there is good reason to expect that mutation effects vary substantially, because genomes contain sites that vary greatly in functional significance. Second, evaluation of some evolutionary theories, such as the time to mutational meltdown, requires knowledge of the distribution of effects (Lande, 1994, 1995; Butcher, 1995; Lynch *et al*., 1995*a*). Third, estimates of U and mean s obtained from mutation accumulation experiments might be substantially biased if the distribution of mutation effects is not co-estimated.

Evidence for wide variation in effects of induced mutations comes from an analysis of the effects of ethyl methane sulphonate (EMS) mutagenesis in *C. elegans*

* Corresponding author. e-mail: peter.keightley@ed.ac.uk

(Davies *et al*., 1999; Keightley *et al*., 2000). The distribution of effects of EMS-induced mutations was evaluated by comparing an *a priori* estimate of the number of induced mutations at the molecular level with an estimate of the number of mutations detectable from fitness assays. The molecular estimate was obtained from the expected rates of EMS-induced point mutations based on experiments to measure forward mutation rates (Bejsovec & Anderson, 1988) and suppressor-induced reversion mutation rates. This yielded the prediction that approximately 45 deleterious point mutations were induced per homozygous mutant line. However, Davies and co-workers found that only 3·60 ($\pm$1·31) were detectable on the basis of fitness assays (Keightley *et al*., 2000). It is likely, therefore, that there is a large class of mutations with undetectably small, but deleterious, effects.

In the present experiment we created inbred sublines from a random selection of the EMS-induced mutant lines produced by Davies *et al*. (1999) in an attempt to refine our estimates of the number of mutations per line. By crossing the mutant lines to an inbred wild-type line and inbreeding the offspring, we produced sublines, which are expected to contain a random selection of half of the mutations present in each mutant line. By measuring the fitness of each mutant line, the wild-type control and the individual sublines, it should be possible to estimate the number of mutations present in each mutant line. The pattern of segregation of mutations among sublines should give information about the distribution of mutation effects without having to rely on information from higher order moments. We have used a modification of the Castle–Wright estimator (Castle, 1921; Wright, 1968) and a maximum likelihood (ML) method to estimate the average number of mutations per line. The ML approach can accommodate data for which the distribution of residual data points is significantly different from the expectations of a normal distribution. The method also allows two classes of mutation effect, although it was not possible to fit a continuous distribution of mutation effects owing to the computing time required. Our results are consistent with the conclusions of Davies *et al*., although we did not have the power to verify the existence of a large class of very small effect mutations.

2. Materials and methods

(i) *Generation of sublines and life history trait assays*

We arbitrarily chose ten of the 56 inbred EMS-induced mutant lines (E1–E9 and E11, collectively termed 'progenitor' lines (p-lines)) produced by Davies *et al*. plus one control line (N2), and thawed them from storage at -80 °C. Unless otherwise stated, worms were maintained at 20 °C on 3·5 cm MYOB agar plates seeded with *Escherichia coli* OP50 using standard techniques (Sulston & Hodgkin, 1988).

N2 males were generated by maintaining a few young N2 hermaphrodites on 6 cm agar plates at 25·5 °C. These were examined daily, and males were moved to agar plates containing several hermaphrodites of the same line and allowed to cross at 20 °C. This was repeated for three consecutive generations, after which time sufficient males had been generated to carry out the crosses described below. Male worms of the N2 strain were then randomly selected and crossed to hermaphrodites of the ten p-lines to produce offspring that were heterozygous for the mutations in each p-line. We checked that the ratio of male to hermaphrodite offspring did not significantly differ from the expected 1 : 1 using a χ^2 test with one degree of freedom. Two of ten p-lines (E1 and E7) produced too few offspring or insufficient males and so could not be included in the experiment.

For each of the eight remaining p-lines, ten F1 hermaphrodite offspring were chosen at random and moved to new plates. Each resulting subline was then inbred for a minimum of ten generations by transferring one larval hermaphrodite, chosen at random, to a new plate every generation. This minimizes selection by bottlenecking the population to one individual each generation and generates offspring that are homozygous for about half of the mutations in the original mutant line, with wild-type (N2) alleles at the rest of their loci. One backup plate was set up each generation in case the primary plate failed. If both of these plates failed, offspring from the previous generation's plates (kept at 16 °C in order to slow their growth) were used. This procedure yielded ten sublines per p-line, labelled E2.1–E2.10, E3.1–E3.10 etc. Only one subline (E4.10) was lost during the inbreeding process owing to the primary, backup and previous generation's plates failure to produce a viable worm, suggesting that the worms were subject to very little natural selection.

Daily productivity and longevity were measured contemporaneously for the control line (N2), the eight p-lines and their respective sublines over three assays. In each assay, each of three people (counters) assayed one worm for each p-line and subline, and eight worms for the control (N2) line per assay, giving a total of nine replicates for each p-line and subline, and 72 replicates for the control line. Within each assay, each counter's plates were randomized with respect to their position in the incubator and the order in which they were counted. Before each assay, replicates were maintained separately for three generations in an attempt to remove any possible maternal effects. If any replicates failed in one assay as a result of unnatural death owing to human error or worms crawling off the plate, extra replicates were added to the same counter's quota in the following assay.

https://doi.org/10.1017/S0016672303006499 Published online by Cambridge University Press

Daily productivity was recorded by counting the number of offspring surviving to the L3 larval stage daily for the first 5 days of productivity. Longevity was scored by recording the day on which the parental worm failed to respond to a light touch from a platinum pick and showed any loss of turgor or visible sign of decay. Four fitness correlates were obtained from the productivity data: early productivity (days 1–2), late productivity (days 3–5), total productivity (days 1–5) and relative fitness. Relative fitness (w) is a measure related to intrinsic population growth rate and is suitable for an age-structured population. To calculate w, the intrinsic growth rate of the controls (r_c) within each assay was computed by solving Eqn 1

$$\sum_x e^{-r_c x} l_c(x) m_c(x) = 1, \tag{1}$$

where $l_c(x)$ and $m_c(x)$ are the least-square means of the proportion of worms surviving to day x and fecundity at day x, respectively, for the controls within an assay. Relative fitness was then calculated separately for each individual from Eqn 2

$$w_{ijk} = \sum_x e^{-r_{ci} x} l_{ijk}(x) m_{ijk}(x), \tag{2}$$

where r_{ci} is the average intrinsic growth rate for the control lines within an assay i, and $l_{ijk}(x)$ and $m_{ijk}(x)$ are the proportions of worms surviving to day x and fecundities at day x, respectively, for assay i, worm j of line k (Charlesworth, 1994, p. 120).

(ii) *Castle–Wright estimator of number of mutations*

The Castle–Wright estimator can be used to calculate the effective number of factors (n_e) contributing to the difference in a trait between two divergently selected inbred lines using information about the phenotypic means and variances of the two progenitor lines and their line-cross derivatives (Castle, 1921; Wright, 1968; Lande, 1981; Cockerham, 1986). We can modify this method to estimate the number of genes contributing to the fitness difference between N2 and a given p-line, assuming that all mutations are additive, unidirectional in effect and unlinked, and have equal effects. With this modification, the Castle–Wright estimator is as follows

$$\hat{n}_e = \frac{(\hat{\mu}_{\mathrm{N2}} - \hat{\mu}_i)^2 - \hat{\sigma}^2_{\hat{\mu}_{\mathrm{N2}}} - \hat{\sigma}^2_{\hat{\mu}_i}}{4\hat{\sigma}^2_{s_i}}, \tag{3}$$

where $\hat{\mu}_{\mathrm{N2}}$ and $\hat{\sigma}^2_{\hat{\mu}_{\mathrm{N2}}}$ are the observed mean and sampling variance of the trait value for N2, and $\hat{\mu}_i$ and $\hat{\sigma}^2_{\hat{\mu}_i}$ are the observed mean and sampling variance of mutant p-line i. $\hat{\sigma}^2_{s_i}$ is the segregational variance among the inbred sublines for p-line i (Lynch & Walsh, 1998). The above means and sampling variances, and the segregational variances amongst each p-line's sublines, were estimated using the MIXED procedure of SAS 6.12 (SAS Institute, 1997) for each trait. Factors included in the model were assay (1–3), counter (1–3), line (1–8), line-type (N2, p-line or subline) and subline (1–10, nested within line × line-type). Counter, assay and subline (line × line-type) were treated as random effects; all other effects were treated as fixed.

The standard error of $\hat{n}_e$ for the Castle–Wright estimator can be approximated using the delta method (Lande, 1981). Modifying this formula to use a variance estimate from sublines instead of an F1, we obtain Eqn 4.

$$\mathrm{Var}(\hat{n}_e) \simeq 4\hat{n}_e^2 \left(\frac{4\left(\hat{\sigma}^2_{\hat{\mu}_{\mathrm{N2}}} + \hat{\sigma}^2_{\hat{\mu}_i}\right)}{(\hat{\mu}_{\mathrm{N2}} - \hat{\mu}_i)^2} + \frac{\mathrm{Var}\left(\hat{\sigma}^2_{s_i}\right)}{\hat{\sigma}^4_{s_i}} \right). \tag{4}$$

This estimate ignores the correction factor proposed for the numerator of the Castle–Wright estimator ($\hat{\sigma}^2_{\hat{\mu}_{\mathrm{N2}}} + \hat{\sigma}^2_{\hat{\mu}_i}$) because it has been suggested that this would unduly complicate the variance (Cockerham, 1986).

(iii) *Likelihood approach for estimating mutational parameters*

Using a ML method to estimate the number of loci contributing to the fitness difference between N2 and a given p-line has the advantage that it uses information about the distribution of fitness values amongst sublines. Similar ML approaches have been used to estimate mutational parameters in previous experiments (Keightley, 1994; Keightley & Bataillon, 2000; Keightley *et al*., 2000; Vassilieva *et al*., 2000); the method used here is based on Keightley & Bataillon (2000). In general, these approaches assume that mutations have additive effects on fitness that follow a given distribution and that, once these effects are removed, the residual data points are normally distributed with the same environmental variance and mean. As an extension to this method, we have relaxed the assumption of normally distributed residuals by assuming instead that the residuals are distributed normally when transformed by an unknown (but estimated) power κ (Box & Cox, 1964).

Following Box and Cox (1964), we assume that for some unknown κ, observations (y) transformed by the function

$$y^{(\kappa)} = \begin{cases} \dfrac{y^\kappa - 1}{\kappa} & (\kappa \neq 0) \\ \log y & (\kappa = 0) \end{cases}$$

satisfy the full normal theory assumptions, assuming that $y > 0$. This function is continuous at $\kappa = 0$ and is therefore preferable to simply using y^κ as the transformation (Box & Cox, 1964).

https://doi.org/10.1017/S0016672303006499 Published online by Cambridge University Press

Replicates of N2 were assumed to have a mean μ and a variance V_E, and to follow a normal distribution when transformed by an unknown power (κ). The p-line and subline replicates were also assumed to have environmental variance V_E, and the number of mutations in each of the p-lines was assumed to be a Poisson random variable with mean λ. Each mutation was assumed to be unlinked from others, to have a negative effect on the trait and to fall into one of two discrete classes of effect size (s_1 and s_2), where the proportion of class 1 mutations (R) is also a parameter of the model. As a special case, we can assume that the proportion of mutations in class 1 is 1; we term this the one-class model. The model allows any number of fixed effects with any number of levels; we modelled both counter and assay as fixed effects for the experimental data.

The levels within a fixed effect were all assumed to have the same variance but different means (scaled relative to the largest level in each fixed effect). For more than one fixed effect, the total of the relevant difference between levels for each fixed effect is calculated separately for each replicate (k) of each p-line (i), and this total is labelled a_{ik} for the following equations. Because all levels are scaled relative to the largest for each fixed effect, a_{ik} can only be negative, meaning that all residuals will be positive when a_{ik} is removed, satisfying the requirement that $y>0$ for the Box–Cox transformation.

If we let X_{ik} equal the phenotypic value of p-line i replicate k then, according to the assumptions above

$$X_{ik}=\mu+x_{1i}s_1+x_{2i}s_2+a_{ik}+e_{ik}^{(\kappa)}, \tag{5}$$

where x_{1i} is the number of mutations in class 1 for p-line i and x_{2i} is the number of mutations in class 2 for p-line i, s_1 and s_2 are the effects of class 1 and class 2 mutations, respectively, (x_1+x_2) is a Poisson deviate with mean λ, and x_1 is a binomial deviate from a total of (x_1+x_2) possible mutations with a probability of success of R. $e_{ik}^{(\kappa)}$ is a transformed Gaussian deviate with mean 0 and variance V_E.

Similarly, if we let Y_{iln} equal the phenotypic value of subline l, replicate n from the p-line i, then

$$Y_{iln}=\mu+y_{1il}s_1+y_{2il}s_2+a_{iln}+e_{iln}^{(\kappa)}, \tag{6}$$

where y_1 and x_1 are binomial deviates with a total of x_1 and x_2 possible events, respectively, and probabilities of success of 0·5.

In the calculation of the likelihood for each line, the likelihood of obtaining the data for that line at every point in the probability space needs to be summed across all the possible points. In our model, there can be anywhere from 0 to an infinite number of mutations present in each p-line. Of these (j) mutations, any number m ($0\leqslant m\leqslant j$) could be in class 1; the remainder (j–m) belong to class 2. Some number p ($0\leqslant p\leqslant m$) of class 1 mutations and some number q [$0\leqslant q\leqslant (j$–$m)$] of class 2 mutations are present in each of the ten sublines of a given p-line. For each possible combination of subline class 1 and class 2 mutations, we need to calculate the likelihood of obtaining our subline data for the ten sublines belonging to each p-line. Each p-line and subline has some number of replicates (p-reps and s-reps).

The likelihood associated with a single line's data will therefore be

$$L(\text{line}_i)=\sum_{j=0}^{\infty}\left(p(j|\lambda)\times\sum_{m=0}^{j}\left(\text{bi}(m|j)\times\prod_{k=1}^{\text{p-reps}}f(X_{ik}-ms_1-(j-m)s_2-a_{ik})\right.\right.$$
$$\left.\left.\times\prod_{l=1}^{\text{sublines}}\sum_{p=0}^{m}\left(\text{bi}(p|m)\times\sum_{q=0}^{(j-m)}\left(\text{bi}(q|(j-m))\times\prod_{n=1}^{\text{s-reps}}f(Y_{iln}-ps_1-qs_2-a_{iln})\right)\right)\right)\right), \tag{7}$$

where $p(j|\lambda)$ denotes the (Poisson) probability that the p-line i contains j mutations given the mean λ, bi($m|j$) denotes the (binomial) probability that p-line i contains m class 1 mutations given that line i contains a total of j mutations, and the probability of each mutation being class 1 is R, bi($p|m$) is the (binomial) probability that subline i, l has p class 1 mutations (given that p-line i has m), bi($q|j$–m) is the (binomial) probability that subline i, l has q class 2 mutations (given that p-line i has j–m), and f is a transformed Gaussian probability density function, shown below (adapted from Box & Cox, 1964)

$$f(y)=\frac{1}{\sqrt{2\pi\sigma_{(\kappa)}^2}}\cdot\exp\left(-\frac{1}{2}\left(\frac{(y^{(\kappa)}-\mu_{(\kappa)})^2}{\sigma_{(\kappa)}^2}\right)\right)\cdot y^{(\kappa-1)}, \tag{8}$$

where y and $y^{(\kappa)}$ are the untransformed and transformed observations as described above. There are three parameters: $\mu_{(\kappa)}$ and $\sigma_{(\kappa)}^2$ are the mean and variance of the transformed variable, respectively, and κ is the power of the transformation.

The overall log-likelihood is then obtained by adding the sum of the log-likelihoods across all p-lines

https://doi.org/10.1017/S0016672303006499 Published online by Cambridge University Press

to the log-likelihood for N2 data. The log-likelihood for the N2 data was summed over all N2 replicates, where the likelihood for each N2 replicate is:

$$L(Z_i) = f(Z_i - a_i),$$

where Z_i is the observation for N2 replicate i and a_i denotes the total effect of any fixed effects modelled.

Approximate standard errors for all parameters were calculated by fitting a quadratic function to a profile likelihood of the parameter of interest.

To verify the calculations and functionality of the ML program, simulations were carried out using the same mutational model as in the likelihood calculation.

(iv) *Likelihood maximization*

It is necessary to search the likelihood space thoroughly to be sure that any maximum found is the true global maximum. Starting values for μ, V_E and any fixed effects were estimated from the N2 data. In order to obtain starting values for the remaining parameters, a grid search was carried out, without maximization, where the likelihood was evaluated for a combination of set values for each parameter over a broad range.

Using the most likely values obtained during the grid search, a linear search was then carried out in which a series of fixed values for λ were selected about its starting value, because this is the parameter of interest. The likelihood was maximized with respect to all other parameters, using the simplex algorithm (Nelder & Mead, 1965). The simplex was then restarted using the values for μ, V_E, λ, s and κ that gave the highest likelihood during the linear search, and the likelihood was maximized with respect to all parameters. The simplex algorithm was restarted after each maximization until there was no further increase in the likelihood.

(v) *E5.2 and E5 extra line crosses*

From the primary experiment, it was clear that line E5.2 had a significantly lower relative fitness than either of its progenitor lines (E5 and N2). Under the assumptions that all mutations are deleterious, freely recombining and show no epistasis, this result is unexpected. Possible explanations are: (1) some lines might carry mixtures of mutations with both positive and negative effects on relative fitness, in which case it would be possible for sublines to have fitnesses outside the range of their progenitor lines; (2) mutations in line E5 might interact epistatically, such that they only cause the dramatic reduction in fitness visible in line E5.2 when segregated in a line cross; (3) a new spontaneous mutation occurred during the generations of selfing that produced subline E5.2.

In order to test the hypothesis that the reduction in relative fitness in subline E5.2 was due to a new spontaneous mutation, we subjected both lines E5 and E5.2 to further line crosses. If a new large-effect mutation had occurred during the generations of selfing then we would expect to see its segregation in sublines generated from a cross between E5.2 and N2, and no evidence of it in sublines produced from a cross between E5 and N2. Alternatively, if mutations present in line E5 cancelled out each others' effects on w through epistasis or by having both positive and negative effects on w, then we would expect to see further sublines (generated from lines E5 and N2) performing outside the range of the two progenitor lines.

We generated 20 sublines from both lines E5 and E5.2 using the same experimental design as for the main experiment except that two new (independently frozen) replicates of the ancestral wild-type line (labelled N2A and N2B) were thawed. E5 and E5.2 were each crossed to the males of N2A and N2B, and ten offspring from each cross were selected randomly and selfed under minimal selection conditions for seven generations. This produced 44 different lines that were then assayed for total productivity: N2A, N2B, E5.2, E5 and 20 sublines for each of E5.2 and E5. Six replicates were set up for each of these lines, giving a total of 264 data points.

The results of the productivity assay were analysed as before using the MIXED procedure of SAS 6.12 (SAS Institute, 1997). Lines E5 and E5.2 were analysed separately, and the factors included in each model were line (N2 or E5/E5.2), line-type (wild type, p-line or subline), subline (1–20, nested within line × line-type) and N2 type (A or B). N2 type and subline (line × line-type) were treated as random effects; all other effects were treated as fixed. We also attempted to estimate the number of mutations segregating in line E5.2 and E5 by applying the Castle–Wright estimator and the ML approach discussed above.

3. Results

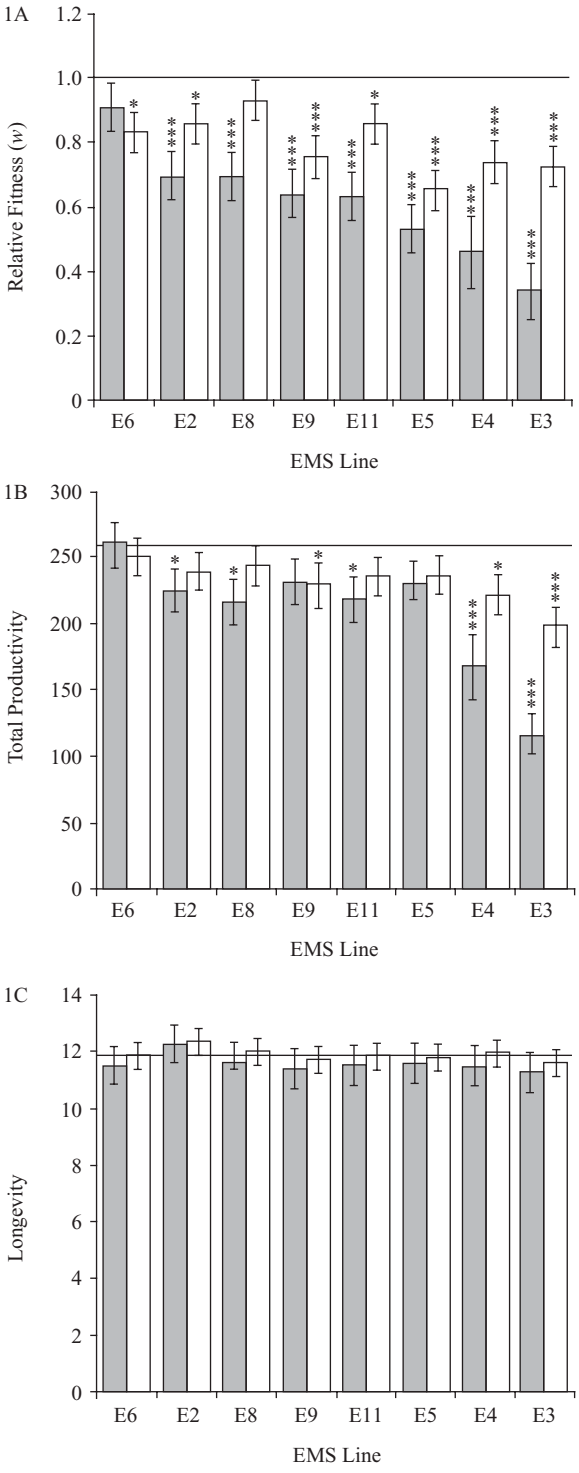
(i) *Segregation of mutant phenotypes and Castle–Wright estimates*

A total of 830 data points were obtained from the experiment for five fitness correlates, and a total of 193,157 offspring were counted to obtain the productivity data. EMS mutagenesis has the strongest effects on early productivity, and this is reflected in a large effect on relative fitness (Table 1, Fig. 1). Mutational effects on late productivity and longevity,

https://doi.org/10.1017/S0016672303006499 Published online by Cambridge University Press

Table 1. *LS-means from Proc MIXED for lines of type N2, p-line or subline. Standard errors are shown in brackets*

Trait	N2 mean	p-line mean	Subline mean
w	1·00 (0·0293)	0·611 (0·0739)	0·792 (0·0244)
Early productivity (worms)	211 (15·9)	140 (21·0)	171 (15·7)
Total productivity (worms)	258 (8·30)	208 (16·7)	231 (7·58)
Late productivity (worms)	46·7 (10·3)	70·7 (12·1)	60·1 (9·61)
Longevity (days)	11·9 (0·367)	11·6 (0·523)	11·9 (0·367)



however, are relatively weak on average (Table 1, Fig. 1). This pattern was also noted by Keightley *et al.* (2000), who hypothesized that this was due to mutations lengthening mean development time, resulting in a decrease in early reproductive output. Deleterious mutations might therefore either increase or decrease late productivity, by delaying development or by reducing total productivity. Longevity in particular appears to be a small 'mutational target', with large amounts of environmental variation. This has also been noted in previous literature, several experiments finding little evidence for strong directional effects of mutations on longevity (Keightley & Caballero, 1997; Pletcher *et al.*, 1999; Vassilieva & Lynch, 1999; Keightley *et al.*, 2000). Neither longevity nor late productivity fit a model with only negative-acting mutations, so these traits were excluded from any of the ML analyses.

The EMS-induced mutant lines tested all had lower point estimates for *w* than N2 (seven out of eight were significantly lower; $p<0{\cdot}0001$; Fig. 1A), seven out of eight had lower point estimates for total productivity (five were significant; $p<0{\cdot}05$; Fig. 1B), and, for longevity, none were significant (for all, $p>0{\cdot}5$; Fig. 1C). For *w*, the mean values of the ten sublines fell between those of their respective p-lines and the N2 for all but one of the lines studied (Fig. 1A). This was also true for all but two lines for total productivity (Fig. 1B) and all but three lines for longevity (Fig. 1C). Most individual sublines also had point estimates for *w* between their respective p-line and N2 (Fig. 2), with one major exception: line E5.2 had a significantly lower early productivity, total productivity and *w* ($p<0{\cdot}0001$) than both line E5 and N2 from which it was derived (Fig. 2D). It is shown later that this is likely to be the result of a single large-effect spontaneous mutation that occurred during the ten generations of inbreeding needed to produce sublines.

Fig. 1. Means for relative fitness (*w*), total productivity and longevity for N2 (horizontal bar) ± standard error (grey box), compared with the means of the p-lines and sublines by line ± standard error. Asterisks above the means of the p-lines and sublines correspond to the significance of the difference between the given genotype and the wild type. * $p<0{\cdot}05$; ** $p<0{\cdot}01$; *** $p<0{\cdot}0001$.

https://doi.org/10.1017/S0016672303006499 Published online by Cambridge University Press

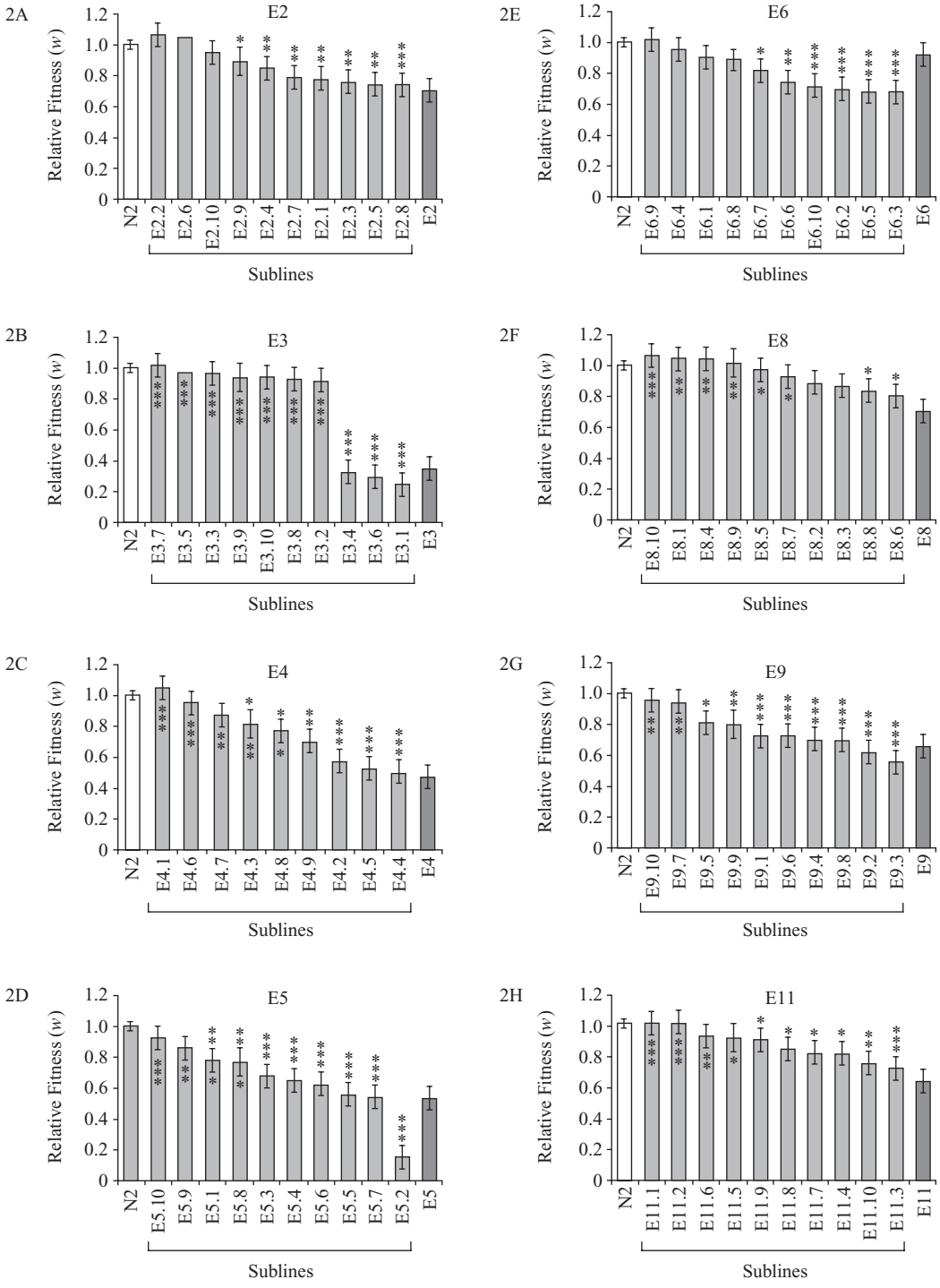


Fig. 2. Means for relative fitness by line, comparing the means for the two progenitor lines (p-line and N2) with all the sublines generated for that line (±standard errors). Asterisks above the error bars correspond to the significance of the difference between the given subline and the wild type. Asterisks below the error bars correspond to the significance of the difference between the given subline and the p-line. * $p<0{\cdot}05$; ** $p<0{\cdot}01$; *** $p<0{\cdot}0001$.

The data for this subline were therefore excluded from the following analyses.

Several of the data points for line E4 were also excluded because many of the worms died during the assay of what were considered to be unnatural causes. Most of these deaths were a result of the worms desiccating after crawling onto the plastic edge of the agar plate. Significantly ($p<0{\cdot}0001$) more worms from line E4 and its sublines (17 worms) died in this manner than from the rest of the experiment put together (only two worms). It is conceivable that line E4 contains a behavioural mutation that

https://doi.org/10.1017/S0016672303006499 Published online by Cambridge University Press

Table 2. *ANOVA table for mixed-model general linear models (GLMs) of relative fitness* (w)*, early productivity, total productivity, late productivity and longevity. Random effects were estimated by restricted maximum likelihood (REML) and significance was tested with* Z *scores rather than* F *statistics*

Trait	Effect	Variance	df_{num}	df_{den}	F	Z
w	Line		7	76·8	1·82	
	Line type		1	78·7	6·72*	
	Subline (line × line-type)	0·0311				5·07**
	Assay	$6{\cdot}52 \times 10^{-5}$				0·28
	Counter	$9{\cdot}32 \times 10^{-5}$				0·3
	Residual	0·0587				19·1**
Early productivity	Line		7	77·1	1·48	
	Line-type		1	77	5·36*	
	Subline (line × line-type)	1240				5·21**
	Assay	683				0·323
	Counter	0·928				0·912
	Residual	2110				19·4**
Total productivity	Line		7	76·3	1·65	
	Line-type		1	78	2·37	
	Subline (line × line-type)	1570				5·18**
	Assay	86·7				0·90
	Counter	0·00				–
	Residual	2570				19·1**
Late productivity	Line		7	77·5	5·46**	
	Line-type		1	80·4	2·21	
	Subline (line × line-type)	252				3·73**
	Assay	262				0·98
	Counter	11·8				0·69
	Residual	1430				19·1**
Longevity	Line		7	75·7	0·55	
	Line-type		1	80·6	0·62	
	Subline (line × line-type)	0·132				0·67
	Assay	0·360				0·92
	Counter	0·00				–
	Residual	8·76				18·8**

* $p < 0{\cdot}05$.
** $p < 0{\cdot}001$.

causes the worms to be more likely to die in this manner.

To estimate the variability among sublines, we performed a mixed-model analysis (SAS Institute, 1997). The effects of counter and assay on all three traits are non-significant but there is significant variation among sublines within lines for most traits (Table 2, Fig. 2). This suggests that much of the variation among sublines within a line is due to a few mutations of large effect, or that there is substantial variation in mutational effects or epistasis among mutations. The large variation in relative fitness among sublines for several p-lines can be seen in Fig. 2. For example, the sublines of line E3 (Fig. 2B) appear to have a bimodal distribution of relative fitness values, implying that there is one large-effect mutation segregating amongst them. Contrasts between p-line E3 and E3 sublines show that three of the sublines (E3.1, E3.4 and E3.6) are significantly different from N2 ($p \leqslant 0{\cdot}0005$) but not E3, whereas the other seven sublines are significantly different from E3 ($p \leqslant 0{\cdot}0001$) but not N2 (Fig. 2B). This pattern is most striking in line E3, although most sublines for the other p-lines show significant differences from one progenitor but not the other. Very few sublines were nonsignificantly different from either progenitor (seven out of 78, excluding subline E5.2). Similarly, very few were significantly different from both (seven out of 78, excluding subline E5.2) (although, in three of these cases, the subline performed worse than either parent). This limited evidence is suggestive of one or two major effect mutations (rather than many similarly sized small effect mutations) for most of the lines tested.

The Castle–Wright estimator was used to estimate the effective number of segregating factors within each mutant p-line. These estimates were then averaged over all eight p-lines to give estimates for each trait (Table 3). Estimates of the effective number of factors using the Castle–Wright estimator are quite low and, despite the large standard errors, are not substantially different from the numbers estimated by

https://doi.org/10.1017/S0016672303006499 Published online by Cambridge University Press

Table 3. *Results for Castle–Wright approach for estimating gene number, for the five traits studied on a line-by-line basis. Approximate standard errors for the number of effective factors are shown in brackets after the estimate along with the average effect* (s). *Overall averages and the standard error of the average are also shown at the base of the Table. Many of the estimates, especially for those traits that were either a small mutational target or might have had their values both increased and decreased as a result of mutagenesis, were incalculable or negative. If the sampling error of the progenitor lines was large, the estimate could be negative, making the calculation of* s *impossible. If the variance amongst the sublines was estimated to be 0 then the estimate for the number of effective factors would be infinite*

	w		Early productivity		Total productivity		Late productivity		Longevity	
Line	n_e	*s*	n_e	*s*	n_e	*s*	n_e	*s*	n_e	*s*
E2	1·53 (2·78)	0·248	1·41 (3·57)	0·214	0·313 (0·832)	0·234	0·116 (0·461)	−1·28	−0·228 (2·20)	–
E3	0·769 (0·848)	0·749	0·653 (0·828)	0·753	0·536 (0·593)	0·754	−0·247 (2·97)	–	−16·5 (2390)	–
E4	1·50 (2·09)	0·440	1·23 (2·02)	0·387	1·21 (2·06)	0·318	∞ (—)	–	∞ (—)	–
E5	2·61 (3·59)	0·289	2·60 (4·39)	0·257	0·281 (0·861)	0·198	1·61 (2·90)	−0·970	−0·305 (3·54)	–
E6	0·0552 (0·196)	0·422	−0·0662 (0·337)	–	∞ (—)	–	0·215 (0·621)	−1·26	∞ (—)	–
E8	5·85 (20·0)	0·127	9·27 (46·8)	0·0905	1·18 (2·91)	0·152	−0·151 (0·761)	–	∞ (—)	–
E9	1·98 (3·10)	0·261	2·44 (5·01)	0·211	0·451 (1·57)	0·161	1·05 (2·41)	−0·801	−0·469 (4·25)	–
E11	3·56 (5·91)	0·196	2·12 (4·92)	0·191	0·661 (1·48)	0·193	−0·0161 (0·0701)	–	∞ (—)	–
Mean	2·23 (2·71)	0·341	2·46 (5·96)	–	— (—)	–	— (—)	–	— (—)	–

Davies *et al.* (1999) or Keightley *et al.* (2000). The Castle–Wright estimator assumes equal effects but, if this assumption is violated, the estimator will underestimate the number of mutations present. Any single large-effect mutation segregating amongst the sublines produced from a cross will lead to a large amount of among-subline variance, reducing the number of factors estimated. It is possible to correct for this bias if the variation of effects is known (Zeng, 1992); alternatively, a ML approach can be used that allows more than one class of mutation effect.

(ii) *Likelihood analysis*

We verified the utility of our ML approach using simulations, the results of which are shown in Table 4. Each set of parameter values in Table 4 was used to simulate 50 data sets. We then used the ML approach to estimate the parameter values from the data. Mean estimates for all parameters do not differ significantly from the simulated values. However, the estimates of some parameters appear to be noisier than others; estimates of κ have the largest standard deviations. Because the accuracy of the estimate of κ depends on the number of data points modelled, the two-class model simulations were designed to have a comparable number of data points per simulation to the experimental data. For each simulation, parameter values were estimated from 600 data points (in comparison to 830 data points for the actual experiment). Over the five sets of simulations, there is a high correlation between the simulated and average estimated values for κ ($r = 0{\cdot}927$ for one class of mutational effects; $r = 0{\cdot}997$ for two classes of mutational effects).

The one-class model allows one class of mutational effects and assumes additivity; in this respect, it is comparable to the Castle–Wright estimator. The number of mutations estimated for the least noisy traits are all similar, low and not substantially different from the Castle–Wright estimates but have smaller standard errors (Table 5). The two-class model allows for two classes of mutations with different effects. It was expected that including variable effects in this way would lead to higher estimates for the number of mutations with correspondingly lower average effects (Keightley, 1998). However, for the three least noisy traits, the most likely mutational model found was a few ($\sim 0{\cdot}13$) very-large-effect mutations ($\sim 70\%$) and many ($\sim 1{\cdot}3$) medium-effect mutations ($\sim 20\%$) (Table 5). The large-effect class seems to emerge as a result of the large-effect mutation segregating in line E3 (Table 3, Fig. 2B). With the one-class model, the fitness reduction associated with line E3 can only be explained away with multiple medium-effect mutations; therefore, the number of mutations estimated with the two-class model is lower (albeit not significantly) than that for the one-class model. For all three

https://doi.org/10.1017/S0016672303006499 Published online by Cambridge University Press

Table 4. *Simulation results for maximum likelihood one- and two-class models. Relative fitness data was simulated according to the models described for the ML analyses. For the one-class model, two sublines per p-line were modelled for a total of 20 simulated p-lines with three replicate data points per p-line and subline. For the two-class model, more (30) p-lines with more (five) replicate data points were modelled owing to the extra number of parameters to be estimated. There were 50 replicates per parameter combination and standard deviations over the 50 replicates are shown in brackets*

One-class model

Simulated values				Estimated values			
λ	s	V_E	κ	λ	s	V_E	κ
1	0·05	0·001	1	1·10 (0·545)	0·0496 (0·00840)	0·000984 ($9{\cdot}28 \times 10^{-5}$)	0·658 (2·39)
1	0·1	0·001	1	0·979 (0·196)	0·0999 (0·00307)	0·000979 ($8{\cdot}13 \times 10^{-5}$)	1·10 (1·87)
2	0·05	0·001	1	2·02 (0·498)	0·0487 (0·00437)	0·000990 ($9{\cdot}83 \times 10^{-5}$)	1·04 (2·44)
2	0·1	0·001	2	1·98 (0·315)	0·0995 (0·00192)	0·000977 ($8{\cdot}86 \times 10^{-5}$)	1·66 (1·72)
2	0·1	0·001	2	2·01 (0·305)	0·100 (0·00188)	0·000995 ($9{\cdot}51 \times 10^{-5}$)	1·85 (1·59)

Two-class model

Simulated values						Estimated values					
λ	s_1	s_2	R	V_E	κ	λ	s_1	s_2	R	V_E	κ
1	0·05	0·02	0·4	0·0001	1	0·957 (0·169)	0·0504 (0·00192)	0·0201 (0·000956)	0·394 (0·108)	$9{\cdot}87 \times 10^{-5}$ ($5{\cdot}79 \times 10^{-6}$)	1·11 (2·57)
4	0·05	0·03	0·6	0·0001	−2	4·08 (0·390)	0·0501 (0·000532)	0·0300 (0·000661)	0·594 (0·0661)	$9{\cdot}86 \times 10^{-5}$ ($5{\cdot}40 \times 10^{-6}$)	−2·14 (2·71)
2	0·1	0·03	0·6	0·001	1	2·07 (0·379)	0·0993 (0·00388)	0·0316 (0·00735)	0·588 (0·0943)	0·000980 ($6{\cdot}38 \times 10^{-5}$)	0·876 (1·13)
1	0·05	0·03	0·4	0·001	2	1·28 (0·495)	0·0475 (0·0143)	0·0299 (0·0102)	0·354 (0·194)	0·000990 ($6{\cdot}74 \times 10^{-5}$)	2·27 (1·21)
3	0·05	0·03	0·4	0·001	−1	3·16 (1·17)	0·0534 (0·0192)	0·0317 (0·0138)	0·422 (0·263)	0·000990 ($6{\cdot}31 \times 10^{-5}$)	−0·935 (1·28)

Table 5. *ML parameter estimates for the one- and two-class models of mutation effects. Approximate standard errors are shown in brackets after the parameter estimate. The log-likelihood (loglik) associated with each parameter combination is shown in the final column. This analysis excludes data for subline E5.2 but includes all data from line E3*

Model	Trait	λ	s			μ	V_E	κ	loglik
One-class	w	1·64 (0·731)	0·229 (0·0261)			1·04 (0·0330)	0·0562 (0·00447)	1·31 (0·150)	−26·6
	Early	1·76 (0·706)	0·162 (0·0215)			241 (5·39)	1680 (129)	2·25 (0·198)	−1928·9
	Total	1·49 (0·767)	0·142 (0·0128)			278 (4·26)	1930 (158)	2·31 (0·187)	−1900·6
			s_1	s_2	R				
Two-class	w	1·41 (0·680)	0·743 (0·0589)	0·213 (0·0228)	0·0884 (0·117)	1·03 (0·0329)	1·32 (0·150)	1·32 (0·150)	−21·2
	Early	1·38 (0·663)	0·635 (0·0408)	0·163 (0·0155)	0·0905 (0·128)	241 (4·05)	2·25 (0·198)	2·25 (0·198)	−1920·9
	Total	1·81 (0·768)	0·612 (0·0496)	0·0814 (0·0240)	0·0691 (0·0955)	277 (5·75)	2·25 (0·183)	2·25 (0·183)	−1891·5

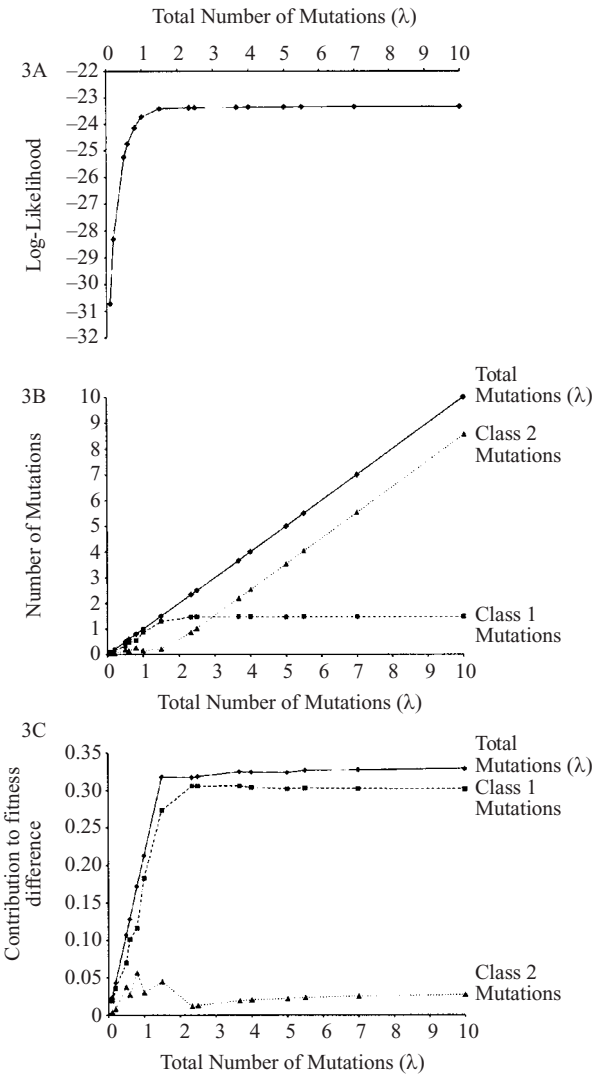


Fig. 3. Plots of total numbers of mutations against log-likelihood (A), number of mutations (B) and contribution to fitness difference of mutations (C), for class 1 mutations (squares), class 2 mutations (triangles) and class 1 + class 2 mutations (diamonds). The number of class 1 or class 2 mutations was calculated by multiplying the proportion of class 1 or class 2 mutations (R or $1-R$) by the total number of mutations. The contribution to fitness difference from class 1 or class 2 mutations is calculated by multiplying the number of class 1 or class 2 mutations by their estimated effect size.

traits studied, the two-class model fitted significantly better than the one-class model ($p<0{\cdot}0001$ in all cases).

The above analysis appears to be dominated by the single large-effect mutation in line E3. Because this might obscure patterns caused by smaller-effect mutations in the other lines, we applied the two-class ML model to our data, excluding line E3. The results of this analysis are shown in Fig. 3A–C for the trait w; similar results were found for early and total productivity. For w, there is virtually no change in log-likelihood above approximately 1·5 mutations

https://doi.org/10.1017/S0016672303006499 Published online by Cambridge University Press

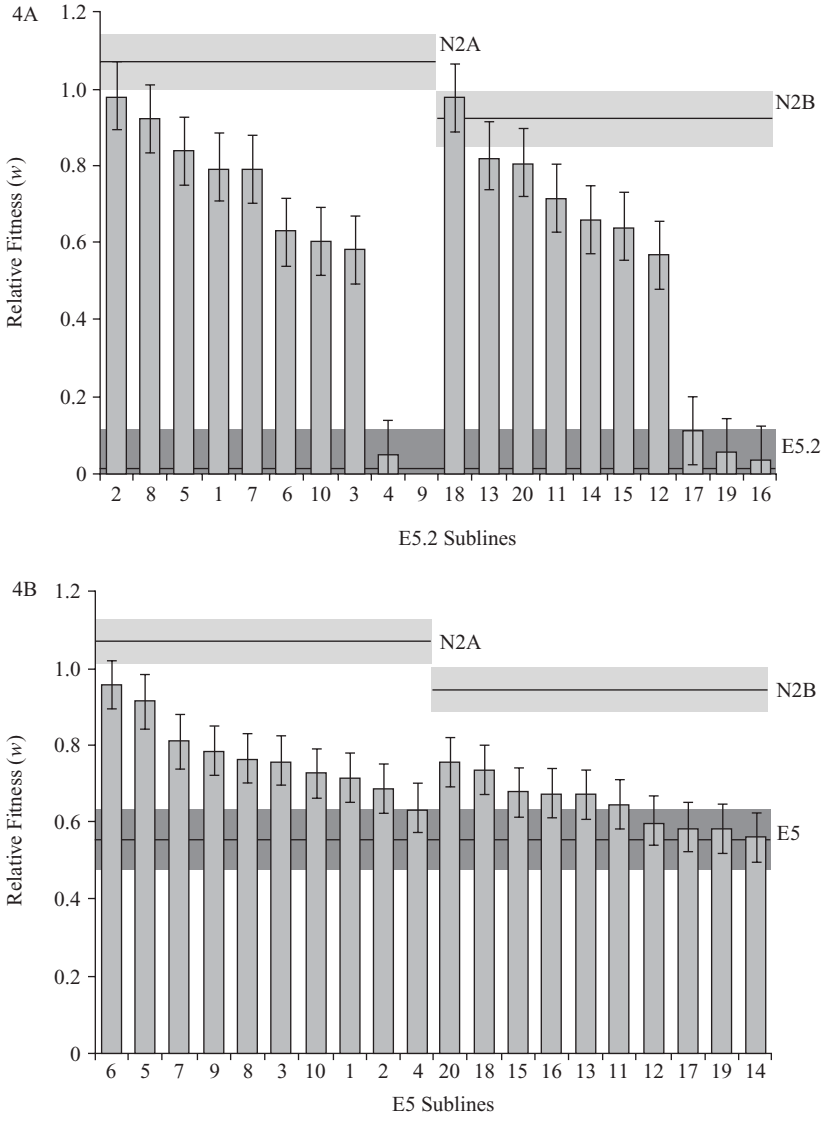


Fig. 4. Means of the sublines and progenitor lines from the analysis of lines E5 and E5.2. Progenitor lines (E5, E5.2, N2A and N2B) are shown as horizontal bars (±standard error as a grey box) above and below the sublines that they correspond to.

(Fig. 3A) (lower confidence limit of 0·487 mutations), suggesting that any number of mutations above ∼1·5 is equally supported by the data. As this estimate of total mutation number increases, the number of class 1 (medium-effect) mutations in the best fitting model remains constant (at ∼1·5); only the number of class 2 (small-effect) mutations increases (Fig. 3B), and these have correspondingly lower effects on fitness, such that their total contribution to the average fitness difference remains more or less constant (Fig. 3C). The only way to distinguish between a model with a few small-effect mutations (e.g. four total mutations, ∼2·5 of which have very small effects of ∼0·8%) and a model with many very-small-effect mutations (e.g. 20 total mutations, ∼18·5 of which have very small effects of ∼0·1%) is to use information about the distribution of these mutations amongst the sublines. It is unlikely, given the number of sublines used in this experiment and the level of environmental variation, that it would be possible to distinguish between these distribution patterns. For all traits, when line E3 was removed, a model with two classes of mutations is more likely than a model with one class, but not significantly so ($p<0·1$).

Estimates of κ, from the two-class ML model including line E3, were tested to see whether they increased the fit to normality of the N2 data after it was transformed, using a Ryan–Joiner normality test (Ryan & Joiner, 1976). Because N2 replicates were assumed to have no mutations, the residual data points could be calculated simply by removing the fixed effects estimated from the ML model. N2 data for both w and early productivity departed significantly ($p<0·025$) from the expectation of a normal

https://doi.org/10.1017/S0016672303006499 Published online by Cambridge University Press

distribution without the transformation, but not once transformed ($p > 0{\cdot}1$). When the same tests were carried out for total productivity, the data did not significantly depart from a normal distribution, with or without the Box–Cox transformation ($p > 0{\cdot}1$). For relative fitness, a significant increase in the likelihood ($p = 0{\cdot}0285$) is obtained when κ is estimated instead of being fixed at 1. The same is true of early productivity ($p < 0{\cdot}0001$) and total productivity ($p < 0{\cdot}0001$).

(iii) *E5.2 and E5 extra line crosses*

Unexpectedly, line E5.2 had a significantly lower relative fitness than either of its progenitors, E5 and N2. To investigate this further, 20 sublines were generated from both lines E5.2 and E5 by crossing them to two freshly thawed replicates of the ancestral N2 (N2A and N2B). Subline 9 generated from line E5.2 was lost during the generations of selfing, owing to the extremely low fitness of the line. Even if this lost line is ignored, it is clear that there is one large-effect mutation present in E5.2, which is segregating amongst the sublines (Fig. 4A). Sublines 4, 16, 17 and 19, which appear to contain this mutation, are not significantly different from their progenitor line E5.2 but are all significantly different from N2 ($p \leqslant 0{\cdot}0001$). Of the remaining 15 sublines, ten are significantly different from both E5.2 and the N2 replicate from which they were generated; only five are not significantly different from line N2. This indicates that there are likely to be some other smaller-effect mutations segregating amongst the sublines of this cross.

Using the Castle–Wright estimator, we estimated that there were 2·64 mutations segregating in line E5.2 (SE 2·39) with an average effect of 0·645. Using our one-class ML model, we estimated that E5.2 contained 1·00 (SE 1·42) mutations with an average effect of 0·957. For the two-class model, we estimated that E5.2 contained 2·00 (SE 2·00) mutations, and that 0·500 of these had an effect of 0·689, whereas the remainder had a lower effect size of 0·278.

All of the extra 20 sublines produced from line E5 had fitness values that were intermediate between the two progenitor lines and there appears to be no evidence of a single large-effect mutation of the size that was observed in the original line E5.2 (Fig. 4B). Using the Castle–Wright estimator, we estimated that there were 5·88 mutations (SE 8·28) with an average effect of 0·212. Applying our one-class ML model, we found the most likely model contained 8·87 mutations, although this model was not a significantly better fit than any models with more than $\sim 0{\cdot}5$ mutations. The most likely two-class model tended towards the results from the one-class model.

We have established that E5.2 contains a single large-effect mutation but we were unable to detect this mutation in the progenitor line E5, suggesting that the mutation occurred spontaneously during the generations of inbreeding that produced line E5.2. Alternatively, it is possible, although unlikely, that the mutation causing the reduction in fitness is present in line E5 but that another tightly linked mutation masked its effects. These mutations might then have been separated after a recombination event during the period of inbreeding that led to line E5.2 but none of the other 29 sublines.

4. Discussion

Davies *et al.* (1999) compared the number of EMS-induced mutations detectable from fitness assays to the number estimated to have been induced in the DNA. They estimated that they had induced an average of at least 45 amino-acid-changing mutations that would be deleterious under natural conditions per homozygous line they studied. However, Davies *et al.* (1999) were able to detect only 1·6 (SE 0·21) (assuming equal effects) or 2·5 (assuming a γ distribution of effects) mutations affecting productivity. Subsequently, Keightley *et al.* (2000) found that only 3·60 ($\pm$1·31) mutations could be detected per line on average with effects on relative fitness. The aim of the present experiment was to estimate more accurately the number of induced mutations per EMS-induced mutant line, by producing sublines for a random selection of the mutant lines. The use of sublines allows large-effect mutations to segregate and it should therefore be possible to determine whether the fitness difference between a wild-type line and a single EMS-induced mutant line is primarily due to few or many mutations with correspondingly large or small effects on fitness. This information in turn can then be used to draw inferences about the distribution of mutation effects.

We used a modification of the Castle–Wright estimator (Castle, 1921) to estimate the number of mutations segregating per line and their average effect. With this approach, we estimated that there were 2·23 mutations on average affecting relative fitness (SE 2·71) and 2·46 on average affecting early productivity (SE 5·96). We also developed a maximum-likelihood approach to estimate the number of mutations, which can allow for variable mutation effects, modelled as two classes of effects. Under the assumption of two mutation classes, ML estimates of mutation numbers are lower than either the Castle–Wright or ML estimates under a one mutational class model. This surprising result seems to be a consequence of the segregation of a single large-effect mutation in one line (E3), which is modelled as several medium-effect (15–20 %) mutations under the one-class model but as a single large-effect mutation under the two-class model. When line E3 was removed from the analysis,

https://doi.org/10.1017/S0016672303006499 Published online by Cambridge University Press

it was found that the most likely two-class model consisted of approximately 1·5 medium-effect (~20%) mutations plus several smaller-effect mutations affecting w. However, it proved impossible to determine the number and corresponding effect size of these smaller-effect mutations, despite the extra power afforded by producing sublines. Our data are therefore consistent with both a model with several small effect mutations (~3 mutations with an effect size of ~1%) and a model with many very small effect mutations (>20 mutations with an effect size <0·2%). Distinguishing between these models would clearly require very much more data.

Our estimates of mutation number are dependent on how we treat variability in effects of mutations. If it is assumed that all mutations have the same effect then it is possible to obtain a concrete estimate of their number, but this is not possible if we assume that there are two classes of effects. Unfortunately, we were unable to test the fit of a γ or other continuous distribution because of the limits of computing power, but such an analysis might provide greater support for a leptokurtic distribution of mutation effects than the two-class model.

There are at least three possible explanations for the difference between the numbers of mutations estimated to have been induced and the number of mutations detected at the phenotypic level. If the estimate of at least 45 deleterious mutations induced per p-line is correct then our results suggest that the distribution of mutation effects is highly leptokurtic and that a large class of mutations have undetectable effects in laboratory assays. This is consistent with several other direct and indirect estimates of the shape of the distribution of mutation effects. For example, transposable elements provide an opportunity to control the number of mutational events at the DNA level, and experiments using these have provided estimates of the distribution of mutation effects. Analysis of the effects of P-element insertions in *Drosophila melanogaster* on metabolic parameters (Clark *et al*., 1995) and bristle numbers (Lyman *et al*., 1996) suggest that mutations with the smallest effects are the most frequent. Similarly, there is direct evidence from Tn10 insertions in *E. coli* for an L-shaped distribution of mutational effects (Elena *et al*., 1998; Elena & Lenski, 1997). A second possibility is that each p-line carries many fewer than 45 deleterious mutations on average, because our estimates of the number carried are indirect. A possible way to resolve this would be to estimate the number of mutational events at the DNA sequence level directly (Denver *et al*., 2000). Finally, it is possible that assaying fitness under standard lab conditions would not reveal every large effect deleterious mutation and that assaying fitness under a range of environments could reveal many more potentially large-effect deleterious mutations.

If the distribution of mutation effects is L-shaped and the vast majority of deleterious spontaneous mutations have nearly neutral (but still deleterious) effects on fitness then this could have implications for several areas of evolutionary theory. For example, mildly detrimental mutations on the border of neutrality are the most damaging to population viability if the effective population size is larger than a few individuals (Lande, 1994). Second, mutations of very small effect are undetectable in the vast majority of fitness assays, leading to underestimates of the mutation rate, which has implications for our understanding of the evolution of sex. It is thought that that the diploid mutation rate must be above one per generation for sexual reproduction to be maintained by deleterious mutations alone (Kondrashov, 1988, 1995). Many estimates of the mutation rate from mutation accumulation experiments that do not account for variability in the effects of mutations fall well below this value (see reviews by Drake *et al*., 1998; Keightley & Eyre-Walker, 1999; Lynch *et al*., 1999). However, these might be substantial underestimates if the degree of variation in mutation effects is high.

We thank B. Charlesworth, J. Christians and two anonymous referees for helpful comments and suggestions, and J. Elrick for technical assistance.

References

Bejsovec, A. & Anderson, P. (1988). Myosin heavy-chain mutations that disrupt *Caenorhabditis elegans* thick filament assembly. *Genes and Development* **2**, 1307–1317.

Box, G. E. P. & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society Series B* **26**, 211–252.

Bulmer, M. G. (1989). Maintenance of genetic variability by mutation-selection balance: a child's guide through the jungle. *Genome* **31**, 761–767.

Butcher, D. (1995). Muller's ratchet, epistasis and mutation effects. *Genetics* **141**, 431–437.

Caballero, A. & Keightley, P. D. (1994). A pleiotropic nonadditive model of variation in quantitative traits. *Genetics* **138**, 883–900.

Castle, W. E. (1921). An improved method of estimating the number of genetic factors concerned in cases of blending inheritance. *Proceedings of the National Academy of Sciences of the USA* **81**, 6904–6907.

Charlesworth, B. (1990). Mutation–selection balance and the evolutionary advantage of sex and recombination. *Genetical Research* **55**, 199–221.

Charlesworth, B. (1994). *Evolution in Age-structured Populations*. Cambridge, UK: Cambridge University Press.

Charlesworth, D. & Charlesworth, B. (1987). Inbreeding depression and its evolutionary consequences. *Annual Review of Ecology and Systematics* **18**, 237–268.

Charlesworth, D., Morgan, M. T. & Charlesworth, B. (1990). Inbreeding depression, genetic load, and the evolution of outcrossing rates in a multilocus system with no linkage. *Evolution* **44**, 1469–1489.

Clark, A. G., Wang, L. & Hulleberg, T. (1995). Spontaneous mutation rate of modifiers of metabolism in *Drosophila*. *Genetics* **139**, 767–779.

Cockerham, C. C. (1986). Modifications in estimating the number of genes for a quantitative character. *Genetics* **114**, 659–664.

Crow, J. F. (1997). The high spontaneous mutation rate: is it a health risk? *Proceedings of the National Academy of Sciences of the USA* **94**, 8380–8386.

Davies, E. K., Peters, A. D. & Keightley, P. D. (1999). High frequency of cryptic deleterious mutations in *Caenorhabditis elegans*. *Science* **285**, 1748–1751.

Denver, D. R., Morris, K., Lynch, M., Vassilieva, L. L. & Thomas, W. K. (2000). High direct estimate of the mutation rate in the mitochondrial genome of *Caenorhabditis elegans*. *Science* **289**, 2342–2344.

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

Elena, S. F. & Lenski, R. E. (1997). Test of synergistic interactions among deleterious mutations in bacteria. *Nature* **390**, 395–397.

Elena, S. F., Ekunwe, L., Hajela, N., Oden, S. A. & Lenski, R. E. (1998). Distribution of fitness effects caused by random insertion mutations in *Escherichia coli*. *Contemporary Issues in Genetics and Evolution* **7**, 349–358.

Kawecki, T. J., Barton, N. H. & Fry, J. D. (1997). Mutational collapse of fitness in marginal habitats and the evolution of ecological specialisation. *Journal of Evolutionary Biology* **10**, 407–430.

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. & 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. & Eyre-Walker, A. (1999). Terumi Mukai and the riddle of deleterious mutation rates. *Genetics* **153**, 515–523.

Keightley, P. D. & Bataillon, T. M. (2000). Multigeneration maximum-likelihood analysis applied to mutation–accumulation experiments in *Caenorhabditis elegans*. *Genetics* **154**, 1193–1201.

Keightley, P. D., Davies, E. K., Peters, A. D. & Shaw, R. G. (2000). Properties of ethylmethane sulfonate-induced mutations affecting life-history traits in *Caenorhabditis elegans* and inferences about bivariate distributions of mutation effects. *Genetics* **156**, 143–154.

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

Kondrashov, A. S. (1995). Contamination of the genome by very slightly deleterious mutations: why have we not died 100 times over? *Journal of Theoretical Biology* **175**, 583–594.

Lande, R. (1981). The minimum number of genes contributing to quantitative variation between and within populations. *Genetics* **99**, 541–553.

Lande, R. (1994). Risk of population extinction from fixation of new deleterious mutations. *Evolution* **48**, 1460–1469.

Lande, R. (1995). Mutation and conservation. *Conservation Biology* **9**, 782–791.

Lyman, R. F., Lawrence, F., Nuzhdin, S. V. & Mackay, T. F. C. (1996). Effects of single P-element insertions on bristle number and viability in *Drosophila melanogaster*. *Genetics* **143**, 277–292.

Lynch, M. & Walsh, B. (1998). *Genetics and analysis of quantitative traits*. Sinauer Associates.

Lynch, M., Conery, J. & Borger, R. (1995*a*). Mutational meltdowns in sexual populations. *Evolution* **49**, 1067–1080.

Lynch, M., Conery, J. & Buerger, R. (1995*b*). Mutation accumulation and the extinction of small populations. *American Naturalist* **146**, 489–518.

Lynch, M., Blanchard, J., Houle, D., Kibota, T., Schultz, S., Vassilieva, L. & Willis, J. (1999). Perspective: spontaneous deleterious mutation. *Evolution* **53**, 645–663.

Muller, H. J. (1950). Our load of mutations. *American Journal of Human Genetics* **2**, 111–176.

Nelder, J. A. & Mead, R. (1965). A simplex method for function minimization. *Computer Journal* **7**, 308–313.

Pletcher, S. D., Houle, D. & Curtsinger, J. W. (1999). The evolution of age-specific mortality rates in *Drosophila melanogaster*: genetic divergence among unselected lines. *Genetics* **153**, 813–823.

Ryan, T. A. & Joiner, B. L. (1976). *Normal probability plots and tests for normality*. Pennsylvania: The Pennsylvania State University.

SAS Institute (1997). The MIXED procedure. In *SAS/STAT® Software: Changes and Enhancements Through Release 6.12*, pp. 571–701. Cary, NC: SAS Institute.

Sulston, J. & Hodgkin, J. (1988). *The Nematode* Caenorhabditis elegans, pp. 587–606. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Turelli, M. (1984). Heritable genetic variation via mutation–selection balance: Lerch's zeta meets the abdominal bristle. *Theoretical Population Biology* **25**, 138–193.

Vassilieva, L. L. & Lynch, M. (1999). The rate of spontaneous mutation for life-history traits in *Caenorhabditis elegans*. *Genetics* **151**, 119–129.

Vassilieva, L. L., Hook, A. M. & Lynch, M. (2000). The fitness effects of spontaneous mutations in *Caenorhabditis elegans*. *Evolution* **54**, 1234–1246.

Wright, S. (1968). *Evolution and the genetics of populations. I. Genetic and biometric foundations*. Chicago, IL: University of Chicago Press.

Zeng, Z. B. (1992). Correcting the bias of Wright's estimates of the number of genes affecting a quantitative character: a further improved method. *Genetics* **131**, 987–1001.