

TOWARD A REALISTIC MODEL OF MUTATIONS AFFECTING FITNESS

PETER D. KEIGHTLEY¹ AND MICHAEL LYNCH²¹*University of Edinburgh, Institute of Cell, Animal and Population Biology, West Mains Road, Edinburgh EH9 3JT, United Kingdom**E-mail: p.keightley@ed.ac.uk*²*Indiana University, Department of Biology, Bloomington, Indiana 47405**E-mail: mlynch@bio.indiana.edu*

Abstract.—Analysis of a recent mutation accumulation (MA) experiment has led to the suggestion that as many as one-half of spontaneous mutations in *Arabidopsis* are advantageous for fitness. We evaluate this in the light of data from other MA experiments, along with molecular evidence, that suggest the vast majority of new mutations are deleterious.

Key words.—*Arabidopsis*, mutation, mutation accumulation, mutational effects distribution, statistical inference.

Received July 16, 2002. Accepted October 3, 2002.

A recent analysis of a mutation-accumulation (MA) experiment in *Arabidopsis thaliana* led to the conclusion that roughly half of new spontaneous mutations increase reproductive fitness (Shaw et al. 2002). This inference was based on the divergence for two fitness-related traits, number of fruits per plant and number of seeds per fruit, among inbred sublines maintained by selfing with minimal selection for 17 generations. The method used to infer the shape of the distribution of mutational effects, Markov chain Monte Carlo maximum likelihood, represents a welcome advance over previous methods based on maximum likelihood (ML) (Keightley 1994; Keightley and Ohnishi 1998), principally because data recorded over several generations can be analyzed simultaneously. The only previous attempt at multigeneration ML analysis assumed a model with equal mutational effects (Keightley and Bataillon 2000). To accommodate positive or negative mutational effects, Shaw et al. (2002) assumed a one-sided gamma distribution with a displacement parameter that represents a significant departure from the reflected gamma distribution employed in previous studies (Keightley 1994; García-Dorado 1997; Keightley and Ohnishi 1998).

Although uncertainties remain with respect to the form of the mutational-effect distribution (García-Dorado et al. 1999; Keightley and Eyre-Walker 1999; Lynch et al. 1999; Bataillon 2000), a great deal of evidence from several sources strongly suggests that the overall effects of mutations are to reduce fitness. Indirect evidence comes from asymmetrical responses to artificial selection on life history traits, suggesting that variance for these traits is maintained by downwardly skewed distributions of mutational effects (Frankham 1990). More direct evidence comes from spontaneous MA experiments in *Drosophila* (Fry et al. 1999; Chavarrias et al. 2001), *Caenorhabditis elegans* (Keightley and Caballero 1997; Vassilieva and Lynch 1999; Vassilieva et al. 2000), wheat (T. Bataillon pers. comm.), yeast (Wloch et al. 2001; Zeyl and DeVisser 2001), *Escherichia coli* (Kibota and Lynch 1996), and a different MA experiment in *Arabidopsis* (Schultz et al. 1999). All of these experiments detected downward trends in MA line population mean fitness relative to control populations as generations accrued. As far as we know, there is no case of even a single MA line maintained by bottle-

necking that showed significantly higher fitness than its contemporary control populations. Ethyl methanesulfonate (EMS) mutagenesis experiments, in which controls are given identical treatment to mutagenized lines, other than a dose of mutagen, have also shown consistently strongly negative effects on fitness traits in *Drosophila* (Mukai 1970; Ohnishi 1977; Mitchell and Simmons 1977; Temin 1978; Keightley and Ohnishi 1998; Yang et al. 2001) and *C. elegans* (Keightley et al. 2000). Similarly, transposable element insertional mutagenesis leads to reduced fitness in *Drosophila* (Mackay et al. 1992) and *E. coli* (Elena and Lenski 1997).

Although the above studies have focused on the fitness effects of mutations in the context of laboratory environments, substantial indirect evidence derived from molecular studies supports the contention that most mutations in natural populations are deleterious (e.g., Ohta 1995; Li 1997; Fay et al. 2001). The fraction of amino-acid altering mutations that is deleterious enough to be removed by selection is approximated by $C = 1 - K_n/K_s$, where K_n and K_s are the substitution rates at nonsynonymous and synonymous sites, respectively. If mutations are neutral on average, C , the proportion of “missing” amino-acid substitutions, would have an expected value of 0.0. However, in all taxa examined so far, average values of C are in excess of 0.7 (e.g., Ohta 1995; Eyre-Walker et al. 2002), implying that the majority of amino-acid altering mutations are deleterious. There is nothing obviously unusual with respect to *A. thaliana* in this regard. Wright et al. (2002) and S. Wright (pers. comm.) have recently investigated constraint in the protein-coding genes of two species of *Arabidopsis*, *A. lyrata* (an outcrosser) and *A. thaliana* (a natural inbreeder), using an outgroup to infer lineage-specific constraint. Estimates for C are 0.88 in both species, despite their different systems of mating; C is likely to underestimate the fraction of amino-acid mutations that are deleterious due to fixation of advantageous amino-acid mutations and purifying selection acting at synonymous sites (Eyre-Walker et al. 2002).

How can the above observations be reconciled with Shaw et al.’s (2002) conclusion that 50% of mutations are advantageous for fitness in *A. thaliana*? First, it is possible that the traits examined by Shaw et al. have intermediate optima and

are under stabilizing selection, so are not genuine major fitness components. Such an effect has been observed in *C. elegans*, in which effects of EMS mutagenesis on productivity late in life are approximately symmetrically distributed, whereas effects on early productivity, a trait more closely related to population growth rate, are strongly skewed downwards (Keightley et al. 2000). Similar effects could also explain the apparently bidirectional effects of EMS mutagenesis on fruit number in *A. thaliana* (Camara and Pigliucci 1999).

Second, the length of the Shaw et al. experiment (17 generations) may have been insufficient to reveal a significant change in mean phenotype, given the level of replication in the MA and control populations. Consider the MA experiment in *C. elegans* carried out over 60 generations by Keightley and Caballero (1997). For lifetime reproductive output, the difference in control and MA-line means was nonsignificant at generation 32, although there was significant among-MA-line variance. By generation 60, the among-MA-line variance was somewhat greater, but the change in overall mean was still nonsignificant. However, the downward skew of the MA-line means in generation 60 provided a strong signature of the effect of MA: seven MA lines had a mean productivity of more than two control standard deviations (SD) below the control mean, whereas the best performing MA line had mean productivity of only 1.5 SD above the control mean. The 214 generation experiment of Vassilieva et al. (2000) clarifies the slowly emerging pattern—as the period of MA extended, progressively lower fitness classes accumulated, whereas the frequencies of the highest fitness classes observed in the controls progressively diminished. Contrary to the suggestions of Shaw et al. (2002), there is no evidence for improved fitness in any periodically bottlenecked *C. elegans* line.

A difficulty with the analysis in Shaw et al. (2002) is the absence of an evaluation of alternative models for the distribution of mutational effects. Indeed, from results reported in their Table 2, it appears that a constant-effects model cannot be rejected for at least one of the two traits (fruit number). Moreover, it was not possible to compare the fit of models with different proportions of positive and negative mutations, because the parameter specifying the location of the mutational-effect distribution had to be fixed prior to the application of MCMCML to avoid “numerical instability.” The latter behavior suggests that the model is overparameterized, and that there is insufficient information in the data from this MA experiment to estimate the proportion of advantageous mutations, which seems to be corroborated by the fact that the standard errors of V_m estimates in this study are about half of the estimates themselves (Shaw et al. 2000).

Although Shaw et al. (2002, p. 460) imply that previous analyses have failed to allow for bidirectional mutations, methods to estimate the proportion of beneficial mutations in MA experiments have been developed and applied (Keightley 1994; Keightley and Caballero 1997; García-Dorado 1997; Keightley and Ohnishi 1998). For example, by fitting a reflected gamma distribution model to the data from a *C. elegans* MA experiment, the ML estimate for the proportion of positive mutations was found to be zero, with an upper limit of 28% (Keightley and Caballero 1997). These earlier approaches assumed a reflected gamma distribution of mutational effects, in preference to the “displaced gamma” pre-

ferred by Shaw et al. (2002). Whereas the reflected gamma distribution has a potential limitation in that it can generate a singularity for mutations with effects near zero, a shifted gamma distribution has the undesirable effect of truncating the distribution of mutations below a lower limit. No model can be expected to incorporate all of the subtleties of biological phenomena, but well-established statistical methodology exists for evaluating the justification for alternative models.

In summary, the vast majority of mutations are deleterious. This is one of the most well-established principles of evolutionary genetics, supported by both molecular and quantitative-genetic data. This provides an explanation for many key genetic properties of natural and laboratory populations.

LITERATURE CITED

- Bataillon, T. 2000. Estimation of spontaneous genome-wide mutation rate parameters: whither beneficial mutations? *Heredity* 84:497–501.
- Camara, M. D., and M. Pigliucci. 1999. Mutational contributions to genetic variance-covariance matrices: an experimental approach using induced mutations in *Arabidopsis thaliana*. *Evolution* 53:1692–1703.
- Chavarrias, 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.
- Elena, S. F., and R. E. Lenski. 1997. Test of synergistic interactions among deleterious mutations in bacteria. *Nature* 390:395–398.
- Eyre-Walker, A., P. D. Keightley, N. G. C. Smith, and D. J. Gaffney. 2002. Quantifying the slightly deleterious model of molecular evolution. *Mol. Biol. Evol.* 19:2142–2149.
- Fay, J., G. J. Wycoff, and C.-I. Wu. 2001. Positive and negative selection on the human genome. *Genetics* 158:1227–1234.
- Frankham, R. 1990. Are responses to artificial selection for reproductive fitness characters consistently asymmetrical? *Genet. Res.* 56:35–42.
- Fry, J. D., P. D. Keightley, S. L. Heinsohn, and S. V. Nuzhdin. 1999. New estimates of rates and effects of mildly deleterious mutation in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* 96:574–579.
- García-Dorado, A. 1997. The rate and effects distribution of viable mutation in *Drosophila*: Minimum distance estimation. *Evolution* 51:1130–1139.
- 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., and A. Caballero. 1997. Genomic mutation rates for lifetime reproductive output and lifespan in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA* 94:3823–3827.
- 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.
- Keightley, P. D., E. K. Davies, A. D. Peters, and R. G. Shaw. 2000. Properties of EMS-induced mutations affecting life history traits in *Caenorhabditis elegans* and inferences about bivariate distributions of mutation effects. *Genetics* 156:143–154.
- Keightley, P. D., and T. A. Bataillon. 2000. Multi-generation maximum likelihood analysis applied to mutation accumulation experiments in *Caenorhabditis elegans*. *Genetics* 154:1193–1201.
- Kibota, T. T., and M. Lynch. 1996. Estimate of the genomic mutation rate deleterious to overall fitness in *E. coli*. *Nature* 381:694–696.
- Li, W.-H. 1997. *Molecular Evolution*. Sinauer, Sunderland, MA.
- Lynch, M., J. Blanchard, D. Houle, T. Kibota, S. Schultz, L. Vas-

- silieva, and J. Willis. 1999. Perspective: spontaneous deleterious mutation. *Evolution* 53:645–663.
- Mackay, T. F. C., R. Lyman, and M. S. Jackson. 1992. Effects of P elements on quantitative traits in *Drosophila melanogaster*. *Genetics* 130:315–332.
- Mitchell, J. A., and M. J. Simmons. 1977. Fitness effects of EMS-induced mutations on the X chromosome of *Drosophila melanogaster*. II. Hemizygous fitness effects. *Genetics* 87:775–783.
- Mukai, T. 1970. Viability mutations induced by ethyl methanesulfonate in *Drosophila melanogaster*. *Genetics* 65:335–348.
- Ohnishi, O. 1977. Spontaneous and ethyl methanesulfonate-induced mutations controlling viability in *Drosophila melanogaster*. II. Homozygous effect of polygenic mutations. *Genetics* 87:529–545.
- Ohta, T. 1995. Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory. *J. Mol. Evol.* 40:56–63.
- Schultz, S. T., M. Lynch, and J. H. Willis. 1999. Spontaneous deleterious mutation in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* 96:11393–11398.
- Shaw, F. J., C. J. Geyer, and R. G. Shaw. 2002. A comprehensive model of mutations affecting fitness and inferences for *Arabidopsis thaliana*. *Evolution* 56:453–463.
- Shaw, R. G., D. L. Byers, and E. Darmo. 2000. Spontaneous mutational effects on reproductive traits of *Arabidopsis thaliana*. *Genetics* 155:369–378.
- Temin, R. G. 1978. Partial dominance of EMS-induced mutations affecting viability in *Drosophila melanogaster*. *Genetics* 89:315–340.
- Vassilieva, L. L., and M. Lynch. 1999. The rate of spontaneous mutation for life-history traits in *Caenorhabditis elegans*. *Genetics* 151:119–129.
- Vassilieva, L. L., A. M. Hook, and M. Lynch. 2000. The fitness effects of spontaneous mutations in *Caenorhabditis elegans*. *Evolution* 54:1234–1246.
- Wloch, D. M., K. Szafraniec, R. H. Borts, and R. Korona. 2001. Direct estimate of the mutation rate and the distribution of fitness effects in the yeast *Saccharomyces cerevisiae*. *Genetics* 159:441–452.
- Wright, S. I., B. Lauga, and D. Charlesworth. 2002. Rates and patterns of molecular evolution in inbred and outbred *Arabidopsis*. *Mol. Biol. Evol.* 19:1407–1420.
- Yang, H. P., A. Y. Tanikawa, W. A. Van Voorhies, J. C. Silva, and A. S. Kondrashov. 2001. Whole-genome effects of ethyl methanesulfonate-induced mutation on nine quantitative traits in outbred *Drosophila melanogaster*. *Genetics* 157:1257–1265.
- Zeyl, C., and J. A. G. M. DeVisser. 2001. Estimates of the rate and distribution of fitness effects of spontaneous mutation in *Saccharomyces cerevisiae*. *Genetics* 157:53–61.

Corresponding Editor: J. Fry