

# Shaking the 'deleterious mutations' dogma?

Thomas Bataillon

UMR 1097 'Diversité & Génomes des Plantes Cultivées' INRA, Domaine de Melgueil, 34130 Mauguio, France

**Estimating the genome-wide mutation rate and the distribution of mutation effects on fitness is important both for the evolution of sex and the fate of small populations. To date, most studies have suggested that most mutations with phenotypic effects tend to be deleterious to fitness. However, Shaw *et al.*, studying the effect of mutations on fitness in *Arabidopsis thaliana*, now suggest that mutations can often have beneficial effects. These findings have been questioned by Keightley and Lynch.**

Testing the assertion that most mutations that have a phenotypic effect are deleterious to fitness, and trying to quantify more precisely the term 'most', is at the heart of many problems in evolutionary and conservation genetics. For instance, a high genome-wide mutation rate for fitness ( $U$ ) could explain major transitions in evolution, such as the emergence of sex and recombination [1]. From a conservation perspective, the rate of mutation and the distribution of mutational effects on fitness could determine the fate of populations [2]. Information about these mutational properties has been obtained through two types of study. The first type estimate  $U$  based on either the observation of the levels of inbreeding depression in natural populations [3] or on the patterns of molecular divergence between closely related species [4]. These are indirect methods and make the implicit assumption that all mutations are deleterious and provide little information about the distribution of mutation effects [5] (in spite of this, these methods represent an active field of research [6]). By contrast, the so-called 'mutation accumulation' (MA) method yields direct estimates of both  $U$  and the distribution of mutation effects for fitness  $f(s)$ . For this method, a single inbred line is used to establish many sub-lines that diverge for several generations (Fig. 1a). Monitoring the divergence between the MA lines for a given trait (Fig. 1b) provides an estimation of the number of mutations that have accumulated on average in the MA lines and enables to infer the effect of these mutations on the trait of interest to be inferred. In practice, components of fitness (hereafter 'fitness traits'), such as viability or lifetime reproductive success, are studied as a proxy for fitness.

## Are mutations deleterious to fitness? New data and a controversy

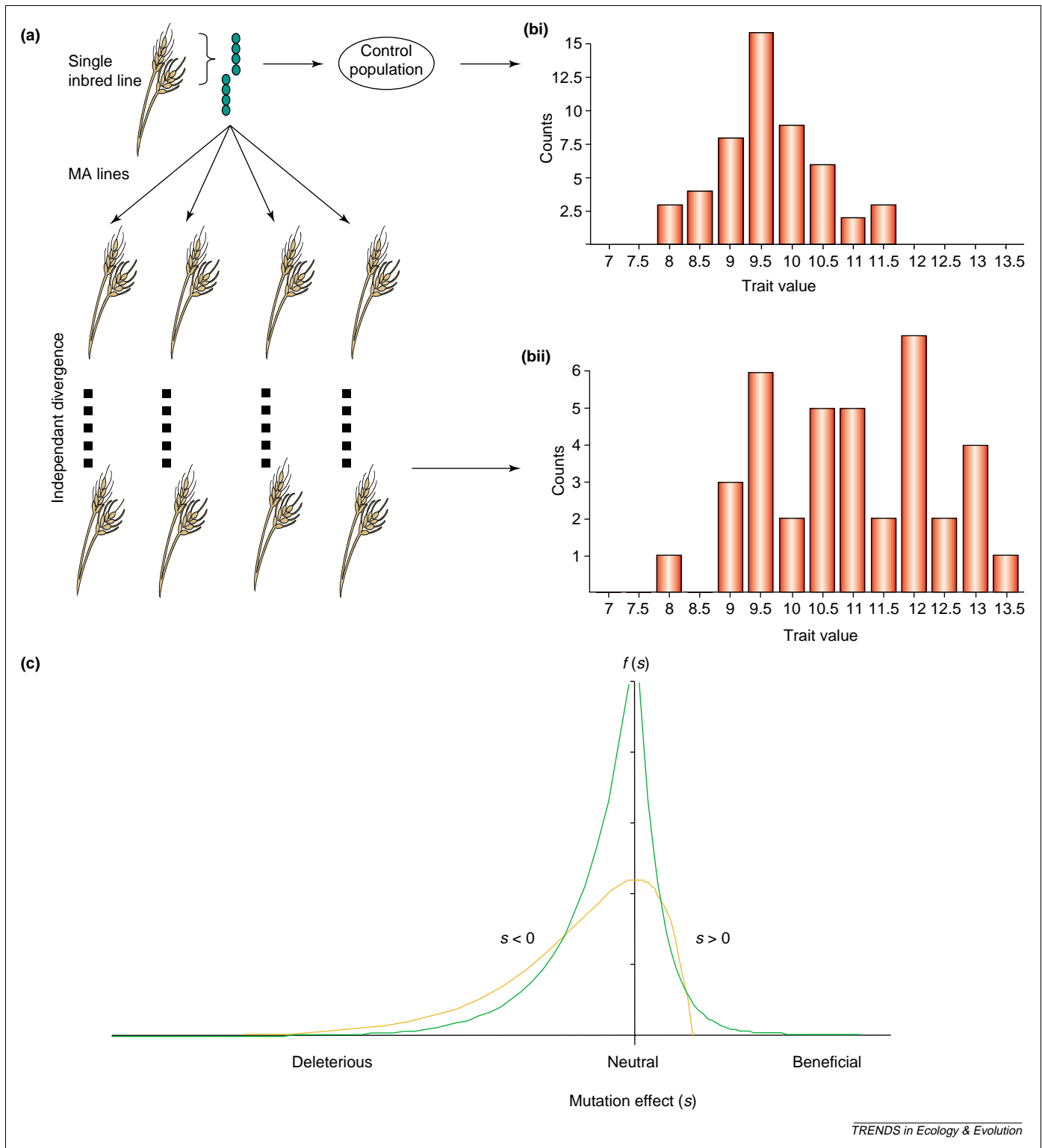
Studies based on MA experiments have often found that the distribution of MA line values for fitness traits after several generations of MA exhibits: (1) a divergence among

MA lines that is best explained by the independent fixation of mutations; and (2) more skew towards low trait values of the MA lines (relative to the mean of the controls) as further generations are considered. The latter observation is consistent with the view that, on average, new mutations reduce fitness components [7]. In a recent paper, Shaw *et al.* [8] used an MA experiment to study the effect of spontaneous mutations on two fitness traits in *Arabidopsis thaliana*: the number of fruits produced and the number of seeds per fruit. The authors found that, compared with previous studies, these fitness traits had a distribution of MA line values that are scattered much more symmetrically around the value of the controls (Fig. 1b) [9].

Characterizing the evolution of line values provides information about how much heritable variation arose through mutation during the experiment but gives no direct insight into genome-wide mutation rates or the effects of mutations underlying such variation. Typically, the number of mutations fixed in a MA line is neither known nor directly observable. Therefore, to gain more insight into the mutational processes generating such heritable variation, researchers use sophisticated statistical methods to find an appropriate 'model'; that is, an underlying  $U$  and a distribution of  $f(s)$  that best fits the observed data (i.e. the distribution of MA line values for a fitness trait and its evolution through time). Historically, the method used to make inferences about  $U$  and the mean fitness effect of mutations ( $s$ ) ignored the fact that different mutations can have different effects. Recent improvements include the incorporation of variability in mutation effects for a single generation [10,11] or the information that comes from observing several generations (at the cost of ignoring variable mutation effects) [12]. To fit the data, it is necessary to assume some distribution family for the phenotypic effect of mutations. Reflected  $\Gamma$  distributions (Fig. 1c) allowing for both deleterious and favourable mutations have already been used [11]. But these distributions have the *a priori* undesirable mathematical properties of having a discontinuity near zero, which could well be the 'grey zone' in which most mutation effects lie.

The analysis by Shaw *et al.* [8] represents a significant improvement. They use a sophisticated numerical method to evaluate the likelihood of a model that incorporates all the phenotypic information that is available from several generations as well as incorporating variable mutation effects. The distribution used to fit the data is also a  $\Gamma$  distribution but is 'displaced' to avoid the discontinuity in zero (Fig. 1c). Shaw *et al.*'s analysis suggests that the underlying distribution of mutation effects is approximately symmetrical, with roughly equal numbers of

Corresponding author: Thomas Bataillon (Thomas.Bataillon@ensam.inra.fr).



**Fig. 1.** Principle of a mutation accumulation (MA) experiment. (a) A MA experiment is begun using seeds (green) from a single inbred line to establish several sub-lines (MA lines) whilst the remaining seeds are kept as controls. Natural selection is greatly relaxed by ensuring that each MA line is represented by exactly one individual in the next generation. MA lines are maintained by strict selfing. Mutations accumulate independently in each MA line for several generations, after which (b) the MA lines and controls are assayed for fitness traits in a common environment. The empirical distribution of the mean phenotypic values for a trait in both the controls (bi) and MA lines (bii) is then used to infer the genome-wide mutation rate ( $U$ ) and the distribution of mutation effect  $f(s)$  on the trait. (c) Two alternative families of distributions  $f(s)$  are used to fit mutation effects in the statistical analysis of MA experiments: a reflected  $\Gamma$  (green line; exhibits a discontinuity in zero) and a displaced  $\Gamma$  (orange line).

mutations increasing and decreasing the phenotypic value of the MA lines for both fitness traits.

In a reply to this study, Keightley and Lynch [13] raise some doubts about the analysis and discuss possible artefacts in fitting the data. They claim that there might

be little information about the distribution of mutation effects that can actually be extracted from the study by Shaw *et al.* The drawback of all these sophisticated methods, in spite of their very sound theoretical statistical basis, is that fitting the model to the data is

computationally demanding and requires a lot of time and, therefore, it is still hard to evaluate the statistical performances of such methods (through Monte Carlo simulation and analysis of many data sets). In spite of these potential technical caveats, there is reasonable evidence that, for both fitness traits, the observed distributions of phenotypic values of MA lines in the experiment by Shaw *et al.* show a significant divergence between the lines but no significant change in mean or skewness after 17 generations of MA. Ultimately, the question should be solved by re-analyzing their data using different families of distributions  $f(s)$  to describe the effect of mutations and to assess the robustness of their finding.

### In search of the 'right' fitness trait

Beyond the intricacies of model fitting, a mutational model including both negative-effect and positive-effect mutations is needed to explain the observed evolution of the number of fruits and number of seeds per fruit in *A. thaliana*. A key question posed by Keightley and Lynch [13] is to what extent fitness in natural populations of *A. thaliana* is related to the reproductive output of plants measured in greenhouse conditions. These 'fitness' traits could actually be under stabilizing selection in nature. Keightley and Lynch also note that a recent study using indirect molecular methods suggests that mutations in *A. thaliana* are mostly deleterious [14]. Solving this apparent paradox will require studies of these 'fitness' traits in natural populations.

The results reported by Shaw *et al.* have potentially important consequences for population conservation and for our view of the effect of mutations on adaptive evolution. The reply from Keightley and Lynch [13], as well as Shaw *et al.*'s subsequent response [15], illustrates the potential pitfalls that are inherent to these difficult yet crucial studies. Clearly, further empirical studies and methodological developments are needed to provide a more

comprehensive view of spontaneous mutations and their effect on fitness, not only in the laboratory, but also in the wild.

### Acknowledgements

I thank Frank Shaw, Ruth Shaw and Peter Keightley for comments.

### References

- 1 Barton, N.H. and Charlesworth, B. (1998) Why sex and recombination? *Science* 281, 1986–1990
- 2 Lynch, M. *et al.* (1995) Mutation accumulation and the extinction of small populations. *Am. Nat.* 146, 489–518
- 3 Deng, H.W. and Lynch, M. (1997) Inbreeding depression and inferred deleterious-mutation parameters in *Daphnia*. *Genetics* 147, 147–155
- 4 Eyre-Walker, A. and Keightley, P.D. (1999) High genomic deleterious mutation rates in hominids. *Nature*, 397, 344–347
- 5 Bataillon, T. (2000) Estimation of spontaneous genome-wide mutation rate parameters: whither beneficial mutations? *Heredity* 84, 497–501
- 6 Eyre-Walker, A. *et al.* (2002) Quantifying the slightly deleterious mutation model of molecular evolution. *Mol. Biol. Evol.* 19, 2142–2149
- 7 Lynch, M. *et al.* (1999) Perspective: spontaneous deleterious mutations. *Evolution* 53, 645–663
- 8 Shaw, F.H. *et al.* (2002) A comprehensive model of mutations affecting fitness and inferences for *Arabidopsis thaliana*. *Evolution* 56, 453–463
- 9 Shaw, R.G. *et al.* (2000) Spontaneous mutational effects on reproductive traits of *Arabidopsis thaliana*. *Genetics* 155, 369–378
- 10 Garcia-Dorado, A. (1997) The rate and effects distribution of viability mutation in *Drosophila*: minimum distance estimation. *Evolution* 51, 1130–1139
- 11 Keightley, P. (1994) The distribution of mutation effects on viability in *Drosophila melanogaster*. *Genetics* 138, 1315–1322
- 12 Keightley, P.D. and Bataillon, T.M. (2000) Multigeneration maximum-likelihood analysis applied to mutation-accumulation experiments in *Caenorhabditis elegans*. *Genetics* 154, 1193–1201
- 13 Keightley, P.D. and Lynch, M. (2003) Towards a realistic model of mutations affecting fitness. *Evolution* 57, 683–685
- 14 Wright, S.I. *et al.* (2002) Rates and patterns of molecular evolution in inbred and outbred *Arabidopsis*. *Mol. Biol. Evol.* 19, 1407–1420
- 15 Shaw, R.G. *et al.* (2003) What fraction of mutations reduce fitness? Reply to Keightley and Lynch. *Evolution* 57, 686–689

0169-5347/03/\$ - see front matter © 2003 Elsevier Science Ltd. All rights reserved.  
doi:10.1016/S0169-5347(03)00128-9

## Can you bank on GenBank?

### D. James Harris

Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO\UP), ICETA, Campus Agrario de Vairão, 4485-661 Vila do Conde, Portugal

Sequencing mitochondrial DNA (mtDNA) is now a routine laboratory procedure. Most journals insist that published sequences be submitted to data bases such as GenBank, where they are publicly available. But quality control of the raw data often depends solely on the original scientists. So just how reliable are the sequences in the data bases? According to a new paper by Forster in the *Annals of Human Genetics*, more than half of all published human mtDNA studies contain mistakes, a level so high that geneticists could be drawing

incorrect conclusions in population and evolutionary studies. Much greater controls are needed, both from journals and from individual scientists. Fortunately, some new methods for detecting errors using phylogenetic networks have recently been proposed. How effective these are remains to be tested.

Genbank is the genetic sequence data base of the National Institute of Health, an annotated collection of all publicly available DNA sequences – currently >28 000 000 000 bases from >250 000 species. Nobody working in the field of molecular biology can be unaware of its immense value.

Corresponding author: D. James Harris (james@mail.icav.up.pt).