

Opinion piece

On measuring selection in experimental evolution

Distributions of mutation fitness effects from evolution experiments are available in an increasing number of species, opening the way for a vast array of applications in evolutionary biology. However, comparison of estimated distributions among studies is hampered by inconsistencies in the definitions of fitness effects and selection coefficients. In particular, the use of ratios of Malthusian growth rates as ‘relative fitnesses’ leads to wrong inference of the strength of selection. Scaling Malthusian fitness by the generation time may help overcome this shortcoming, and allow accurate comparison of selection coefficients across species. For species reproducing by binary fission (neglecting cellular death), $\ln 2$ can be used as a correction factor, but in general, the growth rate and generation time of the wild-type should be provided in studies reporting distribution of mutation fitness effects. I also discuss how density and frequency dependence of population growth affect selection and its measurement in evolution experiments.

Keywords: mutation fitness effects; experimental evolution; population growth; density dependence; frequency dependence

1. INTRODUCTION

The aim of this opinion piece is to clarify the definition and measurement of selection coefficients in evolution experiments investigating fitness effects of mutations. Such studies, from the accumulation of deleterious mutations under relaxed selection over many generations [1–3] to the identification of spontaneous or induced mutations as they arise [4–7], have generated considerable interest recently (reviewed in [8,9]), notably because of their important bearings on many questions across evolutionary genetics. First, they allow testing theoretical predictions about the genetics of adaptation, such as the size of beneficial mutations that get fixed in a population, or the cost of complexity for adaptation (reviewed in [10]). Second, they uncover the intensity of deleterious mutation effects, enabling quantitative predictions to be made about, e.g. the evolution of recombination [11], of selfing [12], or the extinction of small populations [13]. Third, they allow us to address fitness interactions between genes (epistasis) and the ruggedness of fitness landscapes [14–16], which determines the very nature of the evolutionary process from a genetic standpoint. Here I argue that one of the most popular measurements of fitness in evolution experiments (the ratio of Malthusian parameters) does not have a clear

evolutionary meaning, and should be replaced by a measurement relating more directly to the dynamics of frequency change of mutations.

2. FITNESS AND SELECTION

The primary goal of measuring fitness effects of mutations is to relate them to the evolutionary dynamics of alleles under selection. This is quantified by the selection coefficient s . Consider an asexual (haploid) population consisting of two genotypes, a mutant and a wild-type, with population sizes (or density) M and N , and frequencies p and $(1 - p)$. With continuous growth and no age structure, the selection coefficient can be defined as

$$s = \frac{d}{dt} \ln \left(\frac{p}{1-p} \right) \quad (2.1)$$

[17], which has units of time^{-1} . The frequency of the mutant increases if $s > 0$ and decreases if $s < 0$, at a speed determined by s . Since the ratio of allelic frequencies is also the ratio of numbers (or densities) of each genotype, we may also write

$$s = \frac{d \ln M}{dt} - \frac{d \ln N}{dt} \quad (2.2)$$

In particular, if selection is density independent and there are no interactions between genotypes, then

$$s = r_m - r_w, \quad (2.3)$$

where r is the Malthusian parameter [17] or intrinsic rate of increase of each genotype (m , mutant; w , wild-type).

In practice, r is estimated as the regression slope of log-population size against time in the exponential phase, that is, at low population density (assuming no Allee effects). Note that *selection* can be density independent even if *population growth* is density dependent, but this implies that the genotype which initially grows faster also has a higher carrying capacity (figure 1a, first row; electronic supplementary material, appendix; see also [18,19]). When this is valid, s is constant and from equation (2.1), the ratio of genotypic frequencies increases exponentially in time (figure 1a, second row), resulting in a logistic trajectory of the mutant frequency in time. Moreover, in this case, equation (2.3) holds whether each r is measured from cultures of each genotype in isolation, or in a competition experiment mixing both genotypes. However, when in competition, more statistical significance can be obtained by estimating s directly from equation (2.1) [20].

If selection is density dependent, the selection coefficient changes with population size (figure 1b), so that measuring the growth rate in the exponential phase is not sufficient to estimate s . Similarly, with genotype-by-genotype interactions in fitness, s changes with the relative frequencies of each genotype (figure 1c). Disentangling these two possible causes of changes in s (density dependence and frequency dependence) requires combining a competition experiment with cultures of each genotype in isolation (electronic supplementary material, appendix).

Most population genetic models are formulated in discrete non-overlapping generations rather than

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rsbl.2010.0580> or via <http://rsbl.royalsocietypublishing.org>.

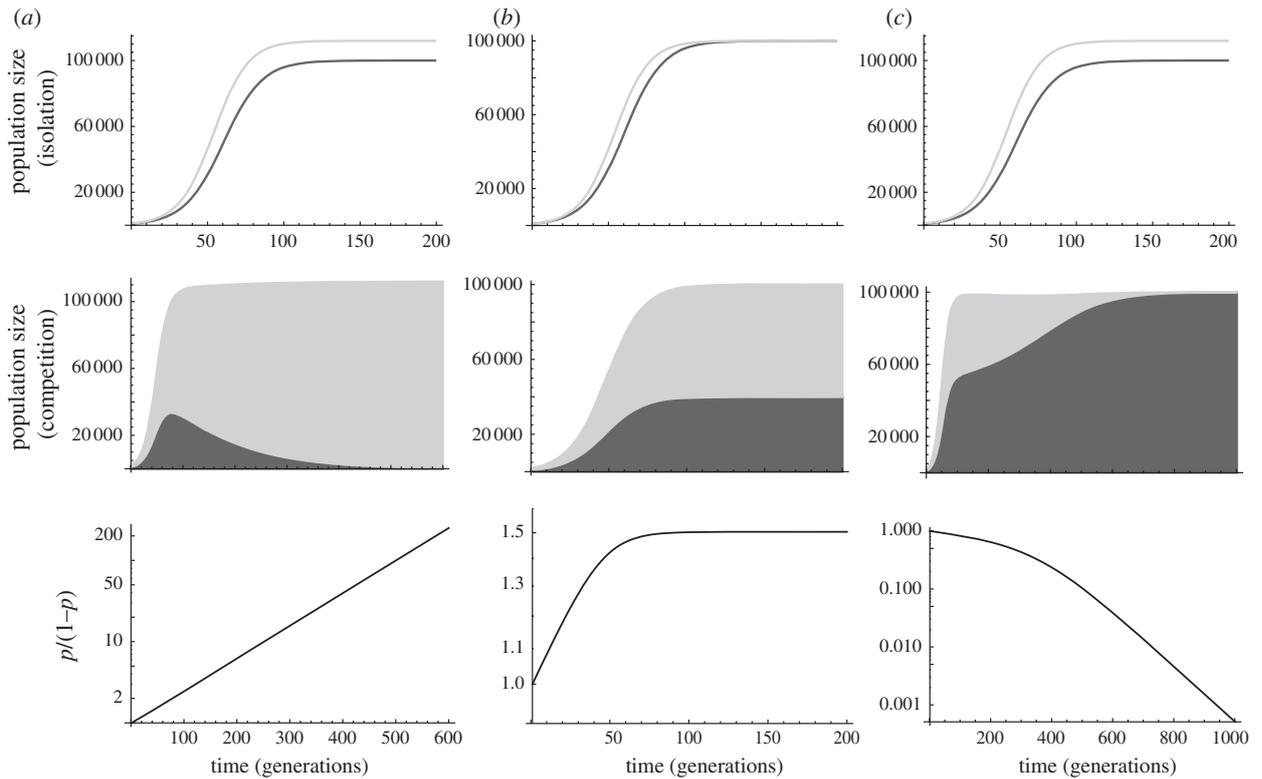


Figure 1. Selection and demography in isolation versus competition. The population sizes of two genotypes grown in isolation (first row: dark grey, wild-type; light grey, mutant) or in competition (second row: dark grey area, wild-type; light grey area, mutant) are shown together with the ratio of genotypic frequencies in competition (third row: note the logarithmic scale on the y -axis) for three demographic scenarios. Scenario (a) leads to frequency- and density-independent selection. Scenario (b) illustrates frequency-independent but density-dependent selection, in the particular case where both genotypes have the same carrying capacity K . In this case, s tends to 0 as the population approaches the carrying capacity. In scenario (c), selection is density-independent but frequency-dependent. Recursions were made from the discrete-generation model in equation (A2) from the electronic supplementary material, appendix, with logistic population growth, using $\lambda_w = 1.08$, $\lambda_m = 1.09$, $K_w = 100\,000$ in all scenarios, and $K_m = K_w \ln(\lambda_m)/\ln(\lambda_w)$ and $\iota_w = \iota_m = 1$ in (a); $K_m = K_w$ and $\iota_w = \iota_m = 1$ in (b); $K_m = K_w \ln(\lambda_m/\lambda_w)$, $\iota_w = 1$ and $\iota_m = 0.98$ in (c).

continuous ones as above, and in this case, an equivalent to equation (2.1) is

$$s_T = \ln\left(\frac{p'/(1-p')}{p/(1-p)}\right) = \ln\left(\frac{M'/N'}{M/N}\right), \quad (2.4)$$

with primes denoting values in the next generation, and the subscript T indicating that the evolutionary change is taken over a generation. In this case, if selection is density independent and there are no genotype-by-genotype interactions, then

$$s_T = \ln\left(\frac{\lambda_m}{\lambda_w}\right) \approx \frac{\lambda_m}{\lambda_w} - 1, \quad (2.5)$$

with λ_m and λ_w the density-independent components of the per-generation multiplicative growth rate of each genotype (electronic supplementary material, appendix). With overlapping (but discrete) generations, such that the population is structured by ages, or if life stages can be identified (such as larva, juvenile, adult), equation (2.5) is also valid for asexuals (and can be used as a weak-selection approximation for sexuals) if λ is the leading eigenvalue of the projection matrix of transition rates (juvenile/adult survival and fecundity) [21,22, p. 61].

Hence, with discrete generations, the ratio λ_m/λ_w of the fitness of the mutant to that of the wild-type

(or ‘relative fitness’, unitless) determines the evolutionary dynamics. Note that λ is necessarily positive (and the population size decreases if $\lambda < 1$), while r can be negative for a decreasing population. Because a negative r cannot be measured starting from a small population, some authors have pooled all mutants with (unmeasured) negative growth rates into the class $r=0$ (sometimes labelled as ‘lethals’), thus artificially inflating this class [6,7,23].

Although equations (2.3) and (2.5) are classical results in population genetics [22,24], a persistent tradition in experimental evolution has been to measure relative fitness in continuously growing populations as r_m/r_w , that is, as a ratio of Malthusian parameters [6,7,23,25–28]. Accordingly, the selection coefficient of a mutation is often inferred from demographic parameters as $s_r = r_m/r_w - 1$, mirroring the definition in discrete-time models from equation (2.5) [7,9,29,30]. However this measure does not directly relate to the evolutionary dynamics of mutations, even for density-independent selection with no interactions between genotypes. Notably, since $s_r = s/r_w$, it would predict faster evolution per unit time in a population with the smaller growth rate r_w of the wild-type, while the actual evolutionary dynamics is the same if s is equal. This has important consequences for any application of these measures to questions in

evolutionary biology where the magnitude of fitness effects of mutations is critical [11–13,31].

3. COMPARING SELECTION COEFFICIENTS

The motivation for presenting results in terms of ‘relative Malthusian fitness’ in experimental evolution has been to obtain a dimensionless parameter [20], allowing for comparison between genotypes and species with different generation times. For instance, a selection coefficient measured in h^{-1} (say, for *Escherichia coli*) is not comparable to one measured in $days^{-1}$ (say, for *Caenorhabditis elegans*). The important question for comparative purposes is how fast evolution occurs per generation. A dimensionless measurement of selection that allows comparison between studies carried out over different time scales is then

$$s_T = (r_m - r_w)T, \quad (3.1)$$

where T is the generation time [22, p. 178]. Electronic supplementary material, table S1, shows an overview of some of the relevant literature where fitness is measured in continuous time (that is, not from survival and fecundity per generation). In six out of nine articles, the selection coefficient per unit time s , or per generation s_T , is not provided and cannot be estimated, because only ratios of Malthusian fitness are given, not the fitness of the wild-type r_w . In only one article is r_w given, so s_T can be estimated using the generation time reported in another study for the same organism (yeast). In two cases, the selection coefficient that is provided is directly s_T , since it is a difference of Malthusian fitnesses measured as growth rates per generation, not per unit time. Overall, the available data make it difficult to compare selection coefficients across species and studies.

Generation times can also vary among mutants. This may even be the main cause of genetic variation in growth rates (per unit time) for some organisms and life cycles [32]. In this case, T should be replaced by T_w , thus measuring how fast evolution occurs relative to the generation time of the wild-type (or ‘resident’) genotype [30]. In particular, selection coefficients are often measured for unicellular organisms that reproduce by binary fission. In this case, neglecting cellular deaths, the number of generations (division events) of the wild-type per unit time is simply $r_w/\ln 2$, and the generation time of the wild-type is $\ln 2/r_w$. A correct measure of the dynamics of selection per generation for species reproducing by binary fission (neglecting cellular death) is then

$$s_{T,\text{bin}} = \frac{r_m - r_w}{r_w} \ln 2. \quad (3.2)$$

Note that if the focus of the study is fixation probabilities rather than the (deterministic) evolutionary dynamics, then T_m should be used as a scaling factor instead of T_w , thus measuring the expected lifetime reproductive output of the mutant. This should affect the shape of the distribution of s , since a different scaling factor would be used for each mutant. None of the studies reported in electronic supplementary material, table S1, provided information about variation in generation times among mutants. Besides, comparison of

expected fixation probabilities across species based solely on measures of selection coefficients from frequency changes is not to be recommended anyway, since for a given s , fixation probabilities can differ widely depending on how stochasticity affects offspring numbers [33].

4. CONCLUSION

Although distributions of mutation fitness effects are being measured in a growing number of organisms, lack of consistency between estimates still prevents proper comparison of selection coefficients for different species, even when selection is density and frequency independent. In particular, the use of relative Malthusian fitness results in overestimating the per-generation strength of selection (in terms of evolutionary dynamics) by a factor $(\ln 2)^{-1} \approx 1.44$ for species that reproduce by binary fission (neglecting cellular deaths), and by an undetermined factor in other cases. It is worthwhile noting that this bias has no effect on the shape of the distribution and on the coefficient of variation of selection coefficients of mutations originating from a given genotype, an important parameter relating to the effective number of phenotypic traits under selection and to the ‘cost of complexity’ for adaptation [9,34]. Apart from this, the fitness effects of mutations across species should be re-examined in the light of the present argument, with particular attention to comparisons between species that reproduce by binary fission and others.

I thank Thomas Lenormand, Guillaume Martin, Arpat Ozgul, Lilia Perfeito, Gabriel Perron, Olivier Tenaillon and an anonymous reviewer for helpful discussions and comments on an earlier draft. This work was supported by a Newton International Fellowship by the Royal Society.

Luis-Miguel Chevin*

Division of Biology, Imperial College London, Silwood Park, Buckhurst Road, Ascot SL5 7PY, UK

*l.chevin@imperial.ac.uk

- Charlesworth, B., Borthwick, H., Bartolome, C. & Pignatelli, P. 2004 Estimates of the genomic mutation rate for detrimental alleles in *Drosophila melanogaster*. *Genetics* **167**, 815–826. (doi:10.1534/genetics.103.025262)
- Kibota, T. T. & Lynch, M. 1996 Estimate of the genomic mutation rate deleterious to overall fitness in *E. coli*. *Nature* **381**, 694–696. (doi:10.1038/381694a0)
- Mukai, T., Chigusa, S. I., Crow, J. F. & Mettler, L. E. 1972 Mutation rate and dominance of genes affecting viability in *Drosophila melanogaster*. *Genetics* **72**, 335.
- Imhof, M. & Schlötterer, C. 2001 Fitness effects of advantageous mutations in evolving *Escherichia coli* populations. *Proc. Natl Acad. Sci. USA* **98**, 1113–1117. (doi:10.1073/pnas.98.3.1113)
- Perfeito, L., Fernandes, L., Mota, C. & Gordo, I. 2007 Adaptive mutations in bacteria: high rate and small effects. *Science* **317**, 813–815. (doi:10.1126/science.1142284)
- Sanjuan, R., Moya, A. & Elena, S. F. 2004 The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proc. Natl Acad. Sci. USA* **101**, 8396–8401. (doi:10.1073/pnas.0400146101)

- 7 Wloch, D. M., Szafraniec, K., Borts, R. H. & Korona, R. 2001 Direct estimate of the mutation rate and the distribution of fitness effects in the yeast *Saccharomyces cerevisiae*. *Genetics* **159**, 441–452.
- 8 Eyre-Walker, A. & Keightley, P. D. 2007 The distribution of fitness effects of new mutations. *Nat. Rev. Genet.* **8**, 610–618. (doi:10.1038/nrg2146)
- 9 Martin, G. & Lenormand, T. 2006 A general multivariate extension of Fisher's geometrical model and the distribution of mutation fitness effects across species. *Evolution* **60**, 893–907.
- 10 Orr, H. A. 2005 The genetic theory of adaptation: a brief history. *Nat. Rev. Genet.* **6**, 119–127. (doi:10.1038/nrg1523)
- 11 Otto, S. P. & Lenormand, T. 2002 Resolving the paradox of sex and recombination. *Nat. Rev. Genet.* **3**, 252–261. (doi:10.1038/nrg761)
- 12 Lande, R. & Schemske, D. W. 1985 The evolution of self-fertilization and inbreeding depression in plants. 1. Genetic models. *Evolution* **39**, 24–40. (doi:10.2307/2408514)
- 13 Lynch, M. & Gabriel, W. 1990 Mutation load and the survival of small populations. *Evolution* **44**, 1725–1737. (doi:10.2307/2409502)
- 14 Gros, P. A., Le Nagard, H. & Tenaillon, O. 2009 The evolution of epistasis and its links with genetic robustness, complexity and drift in a phenotypic model of adaptation. *Genetics* **182**, 277–293. (doi:10.1534/genetics.108.099127)
- 15 Martin, G., Elena, S. F. & Lenormand, T. 2007 Distributions of epistasis in microbes fit predictions from a fitness landscape model. *Nat. Genet.* **39**, 555–560. (doi:10.1038/ng1998)
- 16 Weinreich, D. M., Delaney, N. F., Depristo, M. A. & Hartl, D. L. 2006 Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**, 111–114. (doi:10.1126/science.1123539)
- 17 Fisher, R. A. 1930 *The genetical theory of natural selection*. Oxford, UK: Oxford University Press.
- 18 Charlesworth, B. 1971 Selection in density-regulated populations. *Ecology* **52**, 469–474. (doi:10.2307/1937629)
- 19 Prout, T. 1980 Some relationships between density-independent selection and density-dependent population growth. *Evol. Biol.* **13**, 68.
- 20 Lenski, R. E., Rose, M. R., Simpson, S. C. & Tadler, S. C. 1991 Long-term experimental evolution in *Escherichia coli*. 1. Adaptation and divergence during 2000 generations. *Am. Nat.* **138**, 1315–1341. (doi:10.1086/285289)
- 21 Caswell, H. 2001 *Matrix population models: construction, analysis, and interpretation*. Sunderland, MA: Sinauer Associates.
- 22 Charlesworth, B. 1994 *Evolution in age-structured populations*. Cambridge, UK: Cambridge University Press.
- 23 Fudala, A. & Korona, R. 2009 Low frequency of mutations with strongly deleterious but nonlethal fitness effects. *Evolution* **63**, 2164–2171. (doi:10.1111/j.1558-5646.2009.00713.x)
- 24 Crow, J. F. & Kimura, M. 1970 *An introduction to population genetics theory*. New York, NY: Harper & Row.
- 25 Bantinaki, E., Kassen, R., Knight, C. G., Robinson, Z., Spiers, A. J. & Rainey, P. B. 2007 Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. III. Mutational origins of wrinkly spreader diversity. *Genetics* **176**, 441–453. (doi:10.1534/genetics.106.069906)
- 26 Barrick, J. E., Yu, D. S., Yoon, S. H., Jeong, H., Oh, T. K., Schneider, D., Lenski, R. E. & Kim, J. F. 2009 Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* **461**, 1243–1247. (doi:10.1038/nature08480)
- 27 de Visser, J. A. & Lenski, R. E. 2002 Long-term experimental evolution in *Escherichia coli*. XI. Rejection of non-transitive interactions as cause of declining rate of adaptation. *BMC Evol. Biol.* **2**, 19. (doi:10.1186/1471-2148-2-19)
- 28 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*. *Genetica* **103**, 349–358.
- 29 Kassen, R. & Bataillon, T. 2006 Distribution of fitness effects among beneficial mutations before selection in experimental populations of bacteria. *Nat. Genet.* **38**, 484–488. (doi:10.1038/ng1751)
- 30 Sanjuan, R. 2010 Mutational fitness effects in RNA and single-stranded DNA viruses: common patterns revealed by site-directed mutagenesis studies. *Phil. Trans. R. Soc. B* **365**, 1975–1982. (doi:10.1098/rstb.2010.0063)
- 31 Haldane, J. B. S. 1927 A mathematical theory of natural and artificial selection. Part V. Selection and mutation. *Proc. Camb. Phil. Soc.* **23**, 838–844. (doi:10.1017/S0305004100015644)
- 32 Wahl, L. M. & DeHaan, C. S. 2004 Fixation probability favors increased fecundity over reduced generation time. *Genetics* **168**, 1009–1018. (doi:10.1534/genetics.104.029199)
- 33 Johnson, T. & Gerrish, P. 2002 The fixation probability of a beneficial allele in a population dividing by binary fission. *Genetica* **115**, 283–287. (doi:10.1023/A:1020687416478)
- 34 Orr, H. A. 2000 Adaptation and the cost of complexity. *Evolution* **54**, 13–20. (doi:10.1111/j.0014-3820.2000.tb00002.x)