Fluctuating Selection Models and Mcdonald-Kreitman Type Analyses

Toni I. Gossmann[1,2], David Waxman[3], Adam Eyre-Walker[1]*

1 School of Life Sciences, University of Sussex, Brighton, United Kingdom, 2 Department of Animal and Plant Sciences, University of Sheffield, Sheffield, United Kingdom, 3 Centre for Computational Systems Biology, Fudan University, Shanghai, China

Abstract

It is likely that the strength of selection acting upon a mutation varies through time due to changes in the environment. However, most population genetic theory assumes that the strength of selection remains constant. Here we investigate the consequences of fluctuating selection pressures on the quantification of adaptive evolution using McDonald-Kreitman (MK) style approaches. In agreement with previous work, we show that fluctuating selection can generate evidence of adaptive evolution even when the expected strength of selection on a mutation is zero. However, we also find that the mutations, which contribute to both polymorphism and divergence tend, on average, to be positively selected during their lifetime, under fluctuating selection models. This is because mutations that fluctuate, by chance, to positive selected values, tend to reach higher frequencies in the population than those that fluctuate towards negative values. Hence the evidence of positive adaptive evolution detected under a fluctuating selection model by MK type approaches is genuine since fixed mutations tend to be advantageous on average during their lifetime. Never-the-less we show that methods tend to underestimate the rate of adaptive evolution when selection fluctuates.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: a.c.eyre-walker@sussex.ac.uk

Introduction

The McDonald-Kreitman (MK) test [1], and its derivatives [2–5] use the contrast between the levels of polymorphism and substitution at neutral and selected sites to infer the presence of adaptive evolution in the divergence between species. Modified versions of the MK test allow one to quantify α, the proportion of nonsynonymous differences between species due to adaptive evolution [2–5]. The MK test has been widely applied to a number of species and estimates of α vary substantially from limited evidence ($\alpha \approx 0$ to 10%) in humans [6–8] and many plant species [9] to more than 50% in Drosophila [3,10], some plants with large effective population size [11,12] and bacteria [10].

The MK test framework implicitly assumes that selection pressures are constant. However, the environment for most organisms is constantly changing due to fluctuations in physical factors, such as temperature, and biotic factors, such as the prevalence of competitor species and the density and genotype frequencies of other conspecific individuals. This is likely to lead to changes in the strength of selection acting upon a mutation through time [13]; in the extreme this might mean that a mutation is advantageous at one time-point, but deleterious at another. Despite the likelihood that selection fluctuates through time there is relatively little evidence that this is the case. This is probably because measuring the strength of selection is difficult and detecting fluctuating selection requires analyses over several years. However, analyses of data from several species have suggested that some polymorphisms are subject to fluctuating selection ([14–17], reviewed by [13]). In these examples there are changes in the frequency of mutations that appear to be too great to be explained by either random genetic drift or migration. In most of these analyses the mean strength of selection acting upon a mutation appears to be close to zero. However, this might be a sampling artifact, a mutation subject to fluctuating selection in which the average selection coefficient is non-zero is more likely to be lost or fixed.

Fluctuating selection is likely to be more prevalent than the few well documented examples suggest and Bell [13] has argued that fluctuating selection might help resolve why most traits show substantial heritability, even though selection on a short time-scale often appears to be quite strong. Despite the likelihood that the strength of selection varies most work in theoretical population genetics has assumed that the strength of selection is constant through time (see [18–25] for exceptions). Huerta-Sanchez et al. [24] have investigated how fluctuating selection affects the allele frequency distribution, and hence the site frequency spectrum (SFS), and the probability of fixation. They showed that although the expected strength of selection is zero, fluctuating selections leads to an increase in the probability of fixation, a decrease in diversity and a change in the SFS. Huerta-Sanchez et al. [24] therefore suggested that fluctuating selection might generate artifactual evidence of adaptive evolution in MK type analyses. Here we investigate whether this indeed the case and analyse the average strength of selection of mutations contributing to divergence and polymorphism.

Materials and Methods

We use the basic two-allele model investigated by Huerta-Sanchez et al. [24] in which the strength of selection varies each generation, with the expected strength of selection acting upon

Table 1. α estimates for different fluctuating conditions with a expected mean fitness of zero.

β	α_{True}	α^{*}_{True}	α (MK[a])	α (MK[b])	α (MK[c])
1	0.72	0.25	0.09	0.22	0.24
2	0.76	0.46	0.15	0.26	0.30
3	0.85	0.71	0.33	0.39	0.52
4	0.92	0.73	0.43	0.50	0.73
10	0.97	0.91	0.61	0.68	0.81
20	0.99	0.98	0.70	0.74	0.85
30	0.99	0.99	0.77	0.81	0.89
50	1.00	1.00	0.83	0.86	0.92
100	1.00	1.00	0.90	0.93	0.95

[a][2], [b][5],[c][26].
Estimates of adaptive divergence, α, for polymorphism and divergence simulated under varying random fluctuating selection. Three different MK type tests were used. The intensity of the fluctuation is denoted by β.
doi:10.1371/journal.pone.0084540.t001

each allele being zero. Huerta-Sanchez et al. [24] show that this model may be summarized in terms of the parameter $\beta = 2N_e \mathbf{Var}(s)$ where N_e is the effective population size and $\mathbf{Var}(s)$ is the variance in the strength of selection. Furthermore, they demonstrate that this model behaves identically in terms of the SFS and probability of fixation to one in which the strength of selection is autocorrelated between generations; the autocorrelation simply increases the value of β. We investigated the model of Huerta-Sanchez et al. [24] by simulation so that we can track the strength of selection acting upon each mutation as it segregates in the population. In our haploid simulation we introduce a new mutation at a frequency of $1/N$, where N is the population size, at a site that is monomorphic; in our simulations we used a population size of $N = 2000$. The strength of selection acting upon the two alleles is then drawn from a normal distribution with a mean of 1 and a standard deviation of σ. Using the frequency of the new mutation and the strength of selection acting upon the two alleles, we calculated the expected frequency of the new mutation in the next generation, $\hat{f}$, and generated the actual frequency, f, as a number drawn from a binomial distribution with sample size N

and probability $\hat{f}$. If the mutation is lost or fixed a new mutation is introduced and the simulation repeated, otherwise new selection strengths are sampled and another generation repeated. For each value of β we simulated the evolution of 100,000 mutations. We used the simulated data to infer the expected SFS for a population sample of 20 chromosomes (similar results were obtained for other sample sizes). For each mutation that is fixed, or that reaches any arbitrary frequency f, we can calculate the mean strength of selection that has acted upon that mutation up to the time that we sample it.

Results and Discussion

We investigated the impact of fluctuating selection pressures using the model of Huerta-Sanchez et al. [24] (see above for a description of the model). We use simulated data to infer the expected SFS for a population sample and estimated α using several commonly used methods. First we applied the method of Fay et al. [2]:

$$\alpha = 1 - \frac{D_s P_n}{D_n P_s} \qquad (1)$$

where D_n, D_s, P_n and P_s are the numbers of nonsynonymous and synonymous substitutions and polymorphisms, respectively. This method does not take into account the effect of slightly deleterious mutations, which tend to bias the estimate of α downwards. We therefore applied two methods that attempt to correct for this bias. The first is the method of Eyre-Walker and Keightley [5], and second the method of Schneider et al. [26]. The method of Eyre-Walker and Keightley [5] assumes that advantageous mutations are strongly selected and do not contribute substantially to polymorphism. The method of Schneider et al. [26] does not make this assumption and attempts to infer the proportion of mutations that are advantageous and the strength of selection acting in favour of them. For each method we contrast what happens at sites subject to a certain level of fluctuating selection to those at which there is no fluctuation. In both cases the expected strength of selection is zero; the sites with no fluctuation are therefore evolving neutrally. The simulation is set up such that there is free recombination between sites.

Initially we ran our simulations with the strength of selection changing every generation. Our simulations demonstrate that
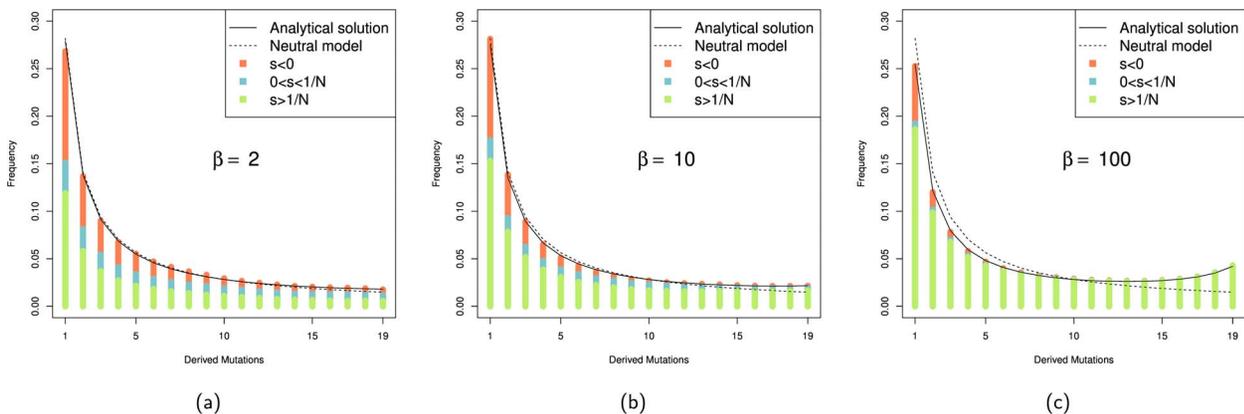

Figure 1. SFS generated under fluctuating conditions ($β = 2$, 10 and 100) with mean selective effect of zero. The proportion of mutations with negative mean selection coefficients are shown in red, the proportion of positive mutations are divided into slightly positive ($<1/N$, blue) and strongly positive ($>1/N$, green). The analytical solution is obtained from the equations from [24].
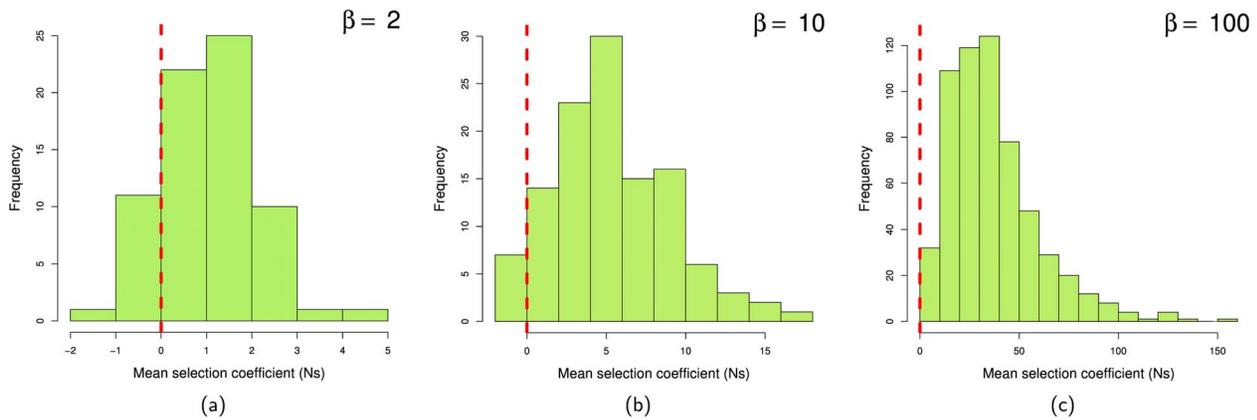doi:10.1371/journal.pone.0084540.g001

Figure 2. Distributions of mean fitness effects of mutations at the time of fixation for fluctuating conditions ($\beta = 2$, 10 and 100) with mean selective effect for all mutations of zero.
doi:10.1371/journal.pone.0084540.g002

fluctuating selection can generate evidence of adaptive evolution; all three of the methods to estimate α yield positive estimates for all values of β (Table 1). The fact that a fluctuating selection model generates evidence of adaptive evolution even when the expected strength of selection is zero suggests that fluctuating selection generates artifactual evidence of positive selection [24]. However, the mean strength of selection experienced by a mutation, that is sampled in a set of DNA sequences, or that spreads to fixation, might not be zero, even though its expected value over all mutations (not just those that fix) is zero; it might be that those mutations that spread to high frequency in the population are those, which just by chance have mean selective values that are positive, whilst those mutations which fluctuate to negative values are lost from the population. To investigate this we tracked the mean strength of selection of each mutation at each frequency up to when it was lost or fixed. From this analysis it is evident that the vast majority of mutations that contribute to the SFS are positively selected, except at very low frequencies and when fluctuations in the strength of selection are quite weak (Figure 1). The bias towards positive mean strengths of selection is even more extreme for those mutations that become fixed (Figure 2).

If we track mutations that ultimately become fixed it is evident that those mutations that start off being slightly negative quickly become positive in their mean value (Figure 3). Interestingly those that start off being highly positive tend to decrease in mean selection coefficient as well; this is probably a consequence of

averaging over many selective episodes, and hence approaching the expected value. We also find that the average mean selection coefficient for all mutations that get fixed declines with time. This is because the critical time for an advantageous mutation is when it is rare because it is more likely to be lost. Those mutations that are strongly positively selected at an early stage have more chance of remaining in the population.

Huerta-Sanchez et al. [24] showed that a model in which the strength of selection changes randomly on average every n^{th} generation behaves identically to a model in which the strength selection changes every generation, but with a larger value of β. Therefore as expected we find that mutations that contribute to polymorphism and divergence tend to be positive selected if the strength of selection changes every 10th generation (Figure 4).

From our simulations it is possible to obtain α_{True}, the proportion of substitutions that are on average positively selected during their passage to fixation. However, some of these may have been fixed by random genetic drift because they are effectively neutral; we therefore also calculated the proportion of substitutions, α_{True}^*, in which the mean strength of selection was $> 1/N$. The methods of Fay et al. [2], Eyre-Walker and Keightley [5] and Schneider et al. [26] all consistently underestimate α_{True} and α_{True}^*, although the effect appears to be more severe for α_{True} and when fluctuating conditions are weak. The method of Schneider et al. [26] is better than that of Eyre-Walker and Keightley [5], which is better than that of Fay et al. [2]. It is perhaps not surprising that
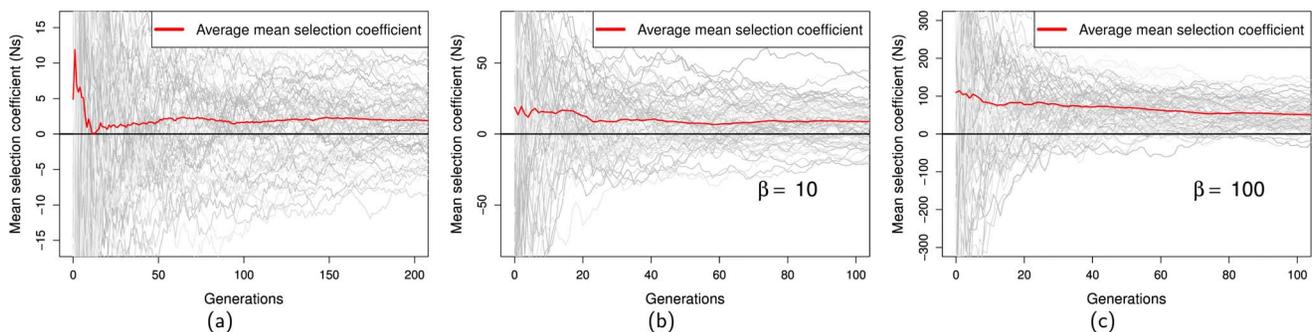

Figure 3. Average mean selection coefficient over time for fixed mutations under fluctuating conditions of $\beta = 2, 10, 100$. Shown are the first 100 generations (200 in case of $\beta = 2$) of 80 mutations that got fixed. The red line indicates the average mean selection coefficient, trajectories in grayscale indicate mean selection coefficients for individual mutations.
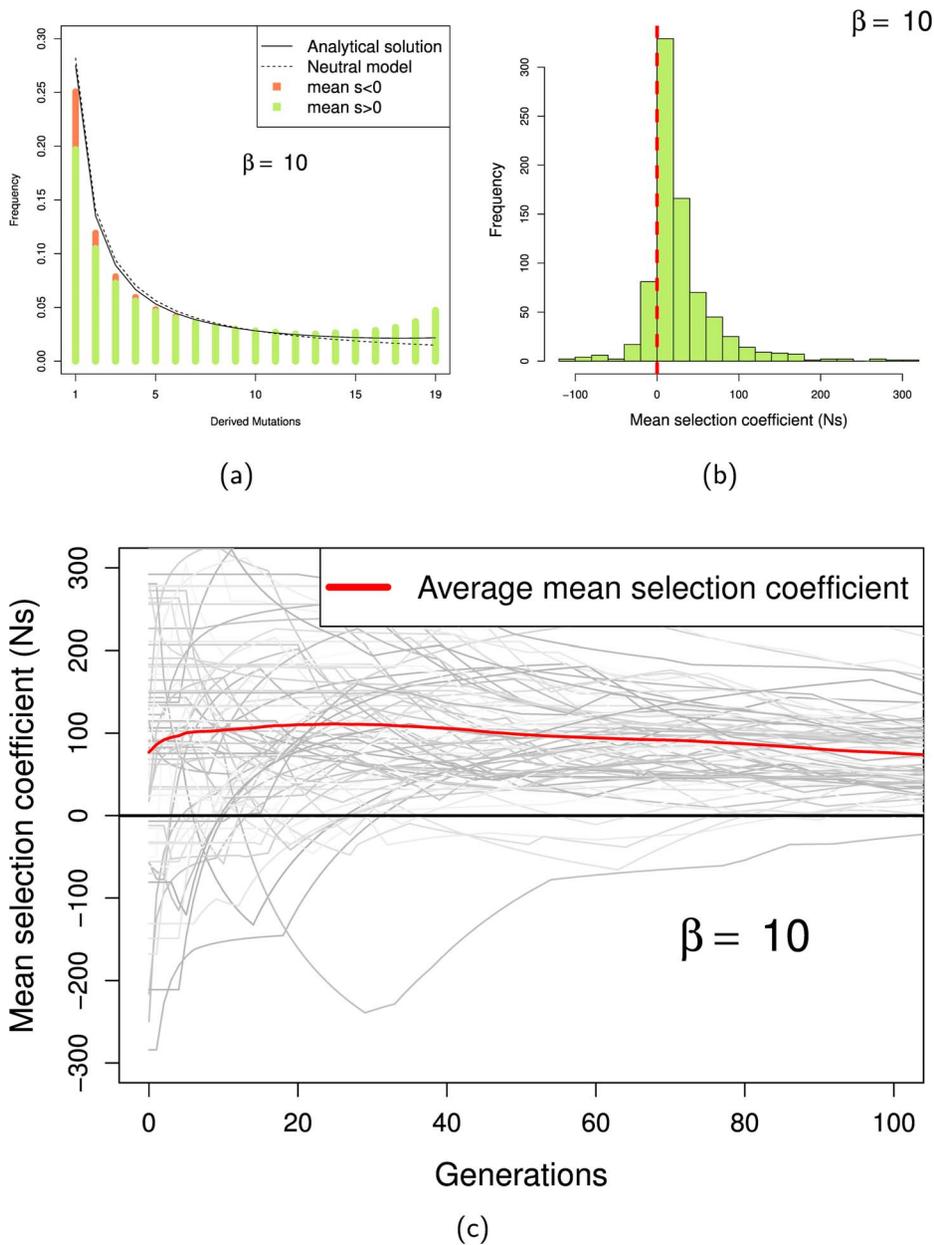doi:10.1371/journal.pone.0084540.g003

Figure 4. SFS and mean fitness effects of mutations under an autocorrelated fluctuating model ($\beta = 10$ and fitness coefficient changes on average every 10th generation). (a) SFS generated under fluctuating conditions. The proportion of mutations with positive and negative mean selection coefficients are shown in green and red, respectively. (b) Distributions of mean fitness effects of mutations at the time of fixation. (c) Average mean selection coefficient over time for fixed mutations.
doi:10.1371/journal.pone.0084540.g004

the method of Schneider et al. [26] performs best since the vast majority of mutations that become fixed have a small average positive selection coefficient and such mutations are likely to contribute to polymorphism, which the other two methods assume does not happen.

We find, in agreement with the suggestion of Huerta-Sanchez et al. [24], that the fluctuating selection does lead to a signature of adaptive evolution. However, we also show that those mutations contributing to polymorphism and divergence are on average positively selected during their lives, even though the expected strength selection is zero. We therefore conclude that the signature of adaptive evolution is genuine. However, it is also evident that

methods to estimate the level of adaptive evolution tend to under-estimate the contribution of mutations subject to fluctuating selection.

Acknowledgments

Author Contributions

Conceived and designed the experiments: TIG DW AEW. Performed the experiments: TIG. Analyzed the data: TIG AEW. Wrote the paper: TIG DW AEW.

References

1. McDonald JH, Kreitman M (1991) Adaptive protein evolution at the Adh locus in Drosophila. Nature 351: 652–654.
2. Fay JC, Wyckoff GJ, Wu CI (2001) Positive and negative selection on the human genome. Genetics 158: 1227–1234.
3. Smith NGC, Eyre-Walker A (2002) Adaptive protein evolution in Drosophila. Nature 415: 1022–1024.
4. Bierne N, Eyre-Walker A (2004) The genomic rate of adaptive amino acid substitution in Drosophila. Mol Biol Evol 21: 1350–1360.
5. Eyre-Walker A, Keightley PD (2009) Estimating the rate of adaptive molecular evo- lution in the presence of slightly deleterious mutations and population size change. Mol Biol Evol 26: 2097–2108.
6. Sequencing C, Consortium A (2005) Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437: 69–87.
7. Zhang L, Li WH (2005) Human SNPs reveal no evidence of frequent positive selection. Mol Biol Evol 22: 2504–2507.
8. Boyko AR, Williamson SH, Indap AR, Degenhardt JD, Hernandez RD, et al. (2008) Assessing the evolutionary impact of amino acid mutations in the human genome. PLoS Genet 4: e1000083.
9. Gossmann TI, Song BH, Windsor AJ, Mitchell-Olds T, Dixon CJ, et al. (2010) Genome wide analyses reveal little evidence for adaptive evolution in many plant species. Mol Biol Evol 27: 1822–1832.
10. Charlesworth J, Eyre-Walker A (2006) The rate of adaptive evolution in enteric bacteria. Mol Biol Evol 23: 1348–1356.
11. Slotte T, Foxe JP, Hazzouri KM, Wright SI (2010) Genome-wide evidence for efficient positive and purifying selection in Capsella grandiora, a plant species with a large effective population size. Mol Biol Evol 27: 1813–1821.
12. Strasburg JL, Kane NC, Raduski AR, Bonin A, Michelmore R, et al. (2011) Effective population size is positively correlated with levels of adaptive divergence among annual sunowers. Mol Biol Evol 28: 1569–1580.
13. Bell G (2010) Fluctuating selection: the perpetual renewal of adaptation in variable environments. Philos Trans R Soc Lond B Biol Sci 365: 87–97.
14. Fisher R, Ford E (1947) The spread of a gene in natural conditions in a colony of the moth Panaxia dominula L. Heredity 1: 143–174.
15. Mueller LD, Barr LG, Ayala FJ (1985) Natural selection vs. random drift: evidence from temporal variation in allele frequencies in nature. Genetics 111: 517–554.
16. Lynch M (1987) The consequences of fluctuating selection for isozyme polymorphisms in Daphnia. Genetics 115: 657–669.
17. O'Hara RB (2005) Comparing the effects of genetic drift and uctuating selection on genotype frequency changes in the scarlet tiger moth. Proc Biol Sci 272: 211–217.
18. Kimura M (1954) Process Leading to Quasi-Fixation of Genes in Natural Populations Due to Random Fluctuation of Selection Intensities. Genetics 39: 280–295.
19. Gillespie J (1973) Natural selection with varying selection coefficientsa haploid model. Genet Res 21: 115–120.
20. Gillespie J (1991) The causes of molecular evolution, volume 2. Oxford University Press, USA.
21. Jensen L (1973) Random selective advantages of genes and their probabilities of fixation. Genet Res 21: 215–219.
22. Karlin S, Levikson B (1974) Temporal fluctuations in selection intensities: Case of small population size. Theoretical Population Biology 6: 383–412.
23. Takahata N, Ishii K, Matsuda H (1975) Effect of temporal uctuation of selection coefficient on gene frequency in a population. Proc Natl Acad Sci U S A 72: 4541–4545.
24. Huerta-Sanchez E, Durrett R, Bustamante CD (2008) Population genetics of poly- morphism and divergence under fluctuating selection. Genetics 178: 325–337.
25. Waxman D (2011) A unified treatment of the probability of fixation when population size and the strength of selection change over time. Genetics 188: 907–913.
26. Schneider A, Charlesworth B, Eyre-Walker A, Keightley PD (2011) A method for inferring the rate of occurrence and fitness effects of advantageous mutations. Genetics 189: 1427–1437.