SHORT NOTE

Estimation of the upper limit of the mutation rate and mean heterozygous effect of deleterious mutations

A. CABALLERO*

Departamento de Bioquímica, Genética e Inmunología, Facultad de Biología, Universidad de Vigo, 36310 Vigo, Spain

(*Received 7 November 2006 and in revised form 19 December 2006*)

Summary

Deng *et al.* have recently proposed that estimates of an upper limit to the rate of spontaneous mutations and their average heterozygous effect can be obtained from the mean and variance of a given fitness trait in naturally segregating populations, provided that allele frequencies are maintained at the balance between mutation and selection. Using simulations they show that this estimation method generally has little bias and is very robust to violations of the mutation–selection balance assumption. Here I show that the particular parameters and models used in these simulations generally reduce the amount of bias that can occur with this estimation method. In particular, the assumption of a large mutation rate in the simulations always implies a low bias of estimates. In addition, the specific model of overdominance used to check the violation of the mutation–selection balance assumption is such that there is not a dramatic decline in mean fitness from overdominant mutations, again implying a low bias of estimates. The assumption of lower mutation rates and/or other models of balancing selection may imply considerably larger biases of the estimates, making the reliability of the proposed method highly questionable.

1. Introduction

Deng *et al.* (2006) have recently proposed the estimation of an upper limit to the diploid genomic rate and the average heterozygous effect of spontaneous deleterious mutations from the mean and variance of a fitness trait in naturally segregating populations. The basis of the method is the following. In the case of outbreeding populations, the diploid genomic rate of deleterious mutations (U) for a fitness trait in a naturally segregating population with deleterious allele frequencies at mutation–selection balance (MSB), assuming multiplicative fitness between loci, is

$$U = -\ln\left(\frac{\overline{W}}{W_{max}}\right),\qquad(1)$$

where $\overline{W}$ is the estimate of the population mean for the fitness trait and W_{max} is the maximum fitness, or that of a hypothetical genotype free of mutations.

Because W_{max} is unknown, a value of 1 can be assumed for traits such as viability, implying that there are no causes of mortality other than the accumulation of deleterious mutations. In this situation, $-\ln\overline{W}$ becomes an upper limit of the rate of mutation,

$$U \leqslant -\ln\overline{W},\qquad(2)$$

i.e. if $W_{max} = 1$ this estimator is unbiased under MSB but, if $W_{max} < 1$, the estimate of U will be biased upwards.

Analogously, from the variance of the population for a fitness trait (σ_W^2), an estimate of the per generation fitness decline from heterozygous mutations ($U\overline{hs}$, where $\overline{hs}$ is the average heterozygous effect of mutations) can be obtained from

$$U\overline{hs} = \ln\left(\frac{\sigma_W^2}{\overline{W}^2} + 1\right).\qquad(3)$$

This estimator is, again, dependent on the assumption of MSB, but is not affected by the value considered

* Corresponding author. Tel: +34 986812568. Fax: +34 986812556. e-mail: armando@uvigo.es

Table 1. *Estimates of the diploid genomic upper limit rate of mutations* (U) *assuming different true values of U and different values of* W_{max}, *the fitness of a hypothetical genotype free of deleterious mutations.* $\overline{W}$ *is the average fitness of the population for the corresponding values of true U and* W_{max}, *using equations* (1) *and* (2) *and assuming that deleterious mutations are at mutation–selection balance frequencies in an infinite population*

True U	W_{max}	$\overline{W}$	Estimated upper limit of U
1·00	1·0	0·368	1·00
	0·8	0·294	1·22
	0·6	0·221	1·51
0·20	1·0	0·819	0·20
	0·8	0·655	0·42
	0·6	0·491	0·71
0·10	1·0	0·905	0·10
	0·8	0·724	0·32
	0·6	0·543	0·61
0·03	1·0	0·970	0·03
	0·8	0·776	0·25
	0·6	0·582	0·54

for W_{max}. Analogous estimators can be obtained for the case of naturally segregating inbred populations (Deng *et al.*, 2006).

By means of computer simulations, Deng *et al.* (2006) tried to quantify the amount of bias in the estimated upper limit of U when $W_{max} < 1$, both under MSB and for a model of balancing selection where overdominant mutations for the fitness trait under study are considered in addition to the dominant mutations. They concluded that the estimation of U and $U\overline{hs}$ is not greatly biased. For example, the estimated upper limit of U was generally no more than twice the true value under the different simulation cases (see table 1 of Deng *et al.*, 2006). Here I point out that some of the models and parameters used by Deng *et al.* (2006) are such that a possible overestimation of U and $U\overline{hs}$ is minimized, and that other models and parameters may change the conclusions dramatically.

2. Overestimation of U when the true rate of mutation is low

In the first place, in their computer simulations to assess the bias in the estimation of U and $U\overline{hs}$ for populations at MSB frequencies, Deng *et al.* (2006) assumed a true mutation rate of $U = 1$, and did not estimate the amount of bias for lower values of the true U. The assumption of a large true value for U implies a low relative bias when $W_{max} < 1$ because,

as the amount of decline in mean fitness through deleterious mutations gets larger, the non-genetic decline becomes relatively less relevant. For example, when the true value of U is 1 and W_{max} is, say, 0·6, the estimated U would only be 1·51 times the true U (see table 1 of Deng *et al.*, 2006, and Table 1 of this note). However, if a lower true mutation rate is assumed, the amount of the bias could be considerably larger (Table 1). For example, if again $W_{max} = 0·6$ but the true U is 0·1, the estimated U would be 6 times larger than the true one, and if $U = 0·03$ the estimated U would be 18 times larger than the true one. Therefore, the conclusion drawn by Deng *et al.* (2006) about the generally low bias of the estimates of U because of $W_{max} < 1$ applies only if the true mutation rate is large. To the extent that many estimates of mutation rates obtained for several fitness traits and species are much lower than 1 (see, for example, García-Dorado *et al.*, 2004), the conclusion drawn by Deng *et al.* (2006) about a low impact of $W_{max} < 1$ on the proposed estimation method does not generally hold.

3. Overestimation of U and $U\overline{hs}$ under overdominance

The second point to note refers to the balancing selection model used by Deng *et al.* (2006) to check the robustness of the estimators for a violation of the MSB assumption. For this they used a model described by Li *et al.* (1999), in which a number of mutations, assumed to be overdominant for the fitness trait under study, are considered in the simulations in addition to the recessive mutations. The results show that this model of balancing selection has little or no impact on the estimates of either the upper limit of U or $U\overline{hs}$, irrespective of the assumed value of W_{max} (see tables 1 and 2 of Deng *et al.*, 2006). However, there is a technical flaw in the overdominant simulation model used by Deng *et al.* (2006), at least in relation to the estimation method proposed, as explained in the following.

In addition to recessive mutations (with heterozygous effects $h_i s_i$), Deng *et al.* (2006) considered overdominant loci with genotypes BB, Bb and bb. They referred to n_3, n_4 and n_5 as the numbers of the three genotypes, respectively, obtained from a trinomial distribution with frequencies p^2, $2pq$ and q^2, respectively, where $p = (h_o - 1)/(2h_o - 1)$ (the dominance and selection coefficients assumed for the overdominant loci were constant values of $h_o = -0·2$ and $s_o = 0·03$, respectively). Then, the equation used for obtaining the mean fitness of outcrossing populations on p. 59 of Deng *et al.* (2006) was $W_{max}[\Pi(1 - h_i s_i)](1 - h_o s_o)^{n_3}(1 - s_o)^{n_4}$. Here I assume that there is a typographical error, and the equation should have exponents n_4 and n_5 instead of n_3 and n_4 (an extra bracket is also mistyped in the equation of p. 59). If the exponents are corrected, the equation

agrees with that of Li *et al.* (1999, p. 898, where n_3 and n_4 referred to *Bb* and *bb*). Note that because h_o is negative (-0.2), the fitness of the overdominant heterozygous genotype (*Bb*) is larger than 1, because the usual scaling of the genotypic values by that with the largest fitness value does not seem to be made. These genotypic values larger than 1 for the heterozygote imply that $W_{max} > 1$ for particular loci, despite the fact that a maximum value $W_{max} = 1$ is assumed in their estimator (equation 5 of their paper; equation 2 in this note).

Thus, although overdominance should produce a genetic load, reducing the population mean and biasing upwards the estimated upper limit of *U*, the addition of overdominant loci in the Deng *et al.* (2006) simulations can cause an increase in population mean fitness. This leads to downwardly biased *U* estimates, as can be seen in their table 1 (i.e. see $U = 0.97$ and 0.93 instead of 1.00 for outcrossing populations, mixed dominance and overdominance, $W_{max} = 1$), implying that the population mean is actually larger when overdominant genes are considered ($\overline{W} = 0.378$ and 0.394, respectively) than when they are not ($\overline{W} = 0.368$). For real data, W_{max} can only be assumed to be 1 for viability (and then to a rough and risky approximation), for which overdominance would never increase the value of the optimum genotype above 1. For other fitness traits, scaling to W_{max} in equation (1) cannot be avoided. Thus, the simulation model used by Deng *et al.* (2006) is intrinsically flawed, as the sign of the bias detected is opposite to that expected using the appropriate W_{max} scaling.

4. Simulation of the estimation method under overdominance

A model of pure overdominance for the loci controlling fitness traits, such as that considered by Deng *et al.* (2006), is decisively rejected by the data as a general model of maintenance of variation (Charlesworth & Hughes, 2000). Other more widely recognized models of balancing selection imply overdominance for global fitness without overdominant gene action for the studied trait. For example, overdominance for overall fitness might arise from the antagonistic pleiotropic effects of mutations on more than one fitness trait, the so-called marginal overdominance model (Falconer & Mackay, 1996, p. 41). Antagonistic pleiotropy is considered a central part of some evolutionary theories, such as the evolution of senescence, and a plausible cause of maintenance of genetic variation for fitness components (see, for example, Roff, 1997; Charlesworth & Hughes, 2000).

In order to assess the impact of an antagonistic pleiotropic model of mutations on the estimation method proposed by Deng *et al.* (2006), the simulation model and procedure described by Fernández *et al.* (2005) was used to obtain estimates of *U* and $U\overline{hs}$ with the method of Deng *et al.* (2006). In brief, mutations are assumed to be strictly deleterious for the fitness trait studied, and a variable proportion of them are assumed to have a pleiotropic advantageous effect on another fitness trait, generating overdominance on global fitness. Using diffusion approximations and transition matrix methods, the distributions of gene frequencies for non-pleiotropic and pleiotropic mutations in a finite population of size $N = 10^4$ at the mutation–selection–drift balance were obtained. From these distributions the population mean ($\overline{W}$) and variance (σ_W^2) for the fitness trait under study were calculated. The corresponding estimates of the upper limit of *U* and of $U\overline{hs}$, obtained from equations (2) and (3) and assuming $W_{max} = 1$, are presented in Table 2 for two sets of mutational parameters that correspond to the models analysed in tables 1 and 2 of Fernández *et al.* (2005), respectively, scaled to the whole diploid genome.

For the first set of mutational parameters (set A in Table 2), the assumed diploid genomic rate of mutations is $U = 0.03$ with a theoretical rate of decline of $U\overline{hs} = 0.12\%$ per generation, whereas for the second set (set B), $U = 0.2$ and $U\overline{hs} = 0.40\%$. Mutational effects are gamma distributed with shape parameter $\beta = 1$ and 0.263 for sets A and B, respectively. Because it might be expected that antagonistic pleiotropic mutations are those with the smallest homozygous effects (it seems reasonable that mutations largely deleterious for a fitness trait are less likely to be favourable for another), the mean effect of these is only 1/8, 1/4 or 1/2 of the average effect of non-pleiotropic mutations. A variable proportion of overdominant pleiotropic mutations relative to the number of non-pleiotropic segregating mutations was investigated, the number of non-pleiotropic segregating mutations being 232 for set A and 5990 for set B. For example, 1% pleiotropic overdominant mutations with mean effect $\bar{s}/8$ in set B implies that about 60 segregating mutations with average homozygous effect $\bar{s} = 0.05/8 = 0.00625$ affecting the trait under study are pleiotropic, this number (60) being about 1% of the number of non-pleiotropic segregating mutations (5990) with average homozygous effect $\bar{s} = 0.05$. More details of the model and procedure can be found in Fernández *et al.* (2005).

In the absence of overdominant pleiotropic mutations, the estimated *U* and $U\overline{hs}$ are close to their expected values for an infinite population, although somewhat lower because the populations analysed are finite ones (see García-Dorado *et al.*, 2003). As shown by Table 2, the effect of a given proportion of pleiotropic overdominant mutations is clear, producing substantial overestimation of *U* and $U\overline{hs}$. The case with $W_{max} < 1$ (not considered in the table) would increase this overestimation further. In conclusion,

Table 2. *Estimates of the diploid genomic upper limit rate of mutations (U) and the per generation mean heterozygous fitness decline ($U\overline{hs}$) for two sets of mutational parameters, assuming $W_{max}=1$*

Mean effect[a]	Parameter set A			Parameter set B		
	% ODM[b]	Estimated upper limit of U	Estimated $U\overline{hs}$ (%)	% ODM[b]	Estimated upper limit of U	Estimated $U\overline{hs}$ (%)
—	0	0·027	0·11	0	0·171	0·22
$\overline{s}/8$	5·4	0·184	0·52	1·0	0·344	0·46
	8·6	0·267	0·69	2·1	0·521	0·71
	17·0	0·525	1·36	3·2	0·699	0·94
	29·0	0·857	2·14	6·3	1·233	1·68
$\overline{s}/4$	2·1	0·109	0·57	0·2	0·248	0·47
	3·2	0·189	0·99	1·1	0·577	1·47
	7·5	0·379	2·02	2·2	0·958	2·57
	10·7	0·596	3·23	3·3	1·391	4·00
$\overline{s}/2$	2·1	0·190	2·16	0·2	0·332	1·36
	4·3	0·354	4·22	1·2	1·041	6·57
	8·6	0·757	9·57	2·4	1·894	12·64
	12·9	1·179	15·17	3·6	2·789	18·81

Population size: $N=10^4$. Mutational effects were obtained from a gamma distribution and dominance coefficients obtained as a function of the mutational effects as described by Fernández *et al.* (2005). Set A: True $U=0·03$, scale parameter for the gamma distribution $\beta=1$, mean homozygous effect $\overline{s}=0·2$, mean dominance coefficient $\overline{h}=0·2$. Set B: $U=0·20$, $\beta=0·263$, $\overline{s}=0·05$, $\overline{h}=0·4$.
[a] Mean effect of overdominant pleiotropic segregating mutations for the trait under study.
[b] Proportion of overdominant pleiotropic mutations segregating in the genome relative to the number of non-pleiotropic segregating mutations (232 for set A and 5990 for set B). Average standard errors of the estimates are 0·8 % of the estimated values for U and 2·4 % of the estimated values for $U\overline{hs}$.

balancing selection through antagonistic pleiotropic effects on fitness components may imply substantial upward biases in the estimates of the mutation rate and the per generation mean heterozygous decline in fitness estimated by the method of Deng *et al.* (2006).

5. Discussion

The first source of bias discussed in this note is important if the true rate of mutation is low relative to the overestimation involved in assuming a value of $W_{max}=1$. Most estimates of U obtained from mutation-accumulation (MA) experiments suggest values of U lower than about 0·2 (e.g. García-Dorado *et al.*, 1999, 2004; Fry & Heinsohn, 2002; Charlesworth *et al.*, 2004; Baer *et al.*, 2005; Joseph & Hall, 2004; Schoen, 2005) so the overestimation problem could become relevant. However, MA experiments provide lower-limit values of U, as many mutations of very small effect can pass undetected in the laboratory. Recent genome-wide interspecies comparisons or analysis of large amounts of the genome have revealed that deleterious mutation rates can be around 1 (Denver *et al.*, 2004; Halligan & Keightley, 2006), for which the overestimation implied in assuming $W_{max}=1$ would be small (Deng *et al.*, 2006; Table 1). However, this presumably large number of mutations undetected in MA experiments must have exceedingly

low values of *s*, as it has been deduced that only mutations with effects smaller than about 0·001 are undetected in such experiments (García-Dorado *et al.*, 2004). This may add another caveat to the estimation method of Deng *et al.* (2006). Equation (1), based on the assumption of a balance between selection and deleterious mutation, may not estimate reliably the rate of mutations with exceedingly small deleterious effects. The reason is that the lower the effect of mutations, the more time is necessary for a large population to reach MSB equilibrium, and the longer this time, the more unlikely the mean fitness of the population is to depend exclusively on the balance between deleterious mutation and selection, because favourable mutations cannot then be ignored (García-Dorado *et al.*, 2004).

The second source of bias discussed (balancing selection) may imply strong overestimation of U and $U\overline{hs}$ irrespective of the true value of U (Table 2). It must be noted, however, that the relative contribution of balancing selection to standing variation is generally unknown and may depend on the particular fitness trait and population studied. Thus, although strict MSB is unlikely to be the only or main source of genetic variation in most natural populations (Charlesworth & Hughes, 2000), some analyses suggest that there are populations with genetic variation for fitness components close to their expectations

under MSB (see, for example, Kusakabe & Mukai, 1984; Rodríguez-Ramilo *et al.*, 2004).

Deng *et al.* (2006) applied their method to data from three species of *Daphnia*, obtaining an average estimate of the upper limit of U for survivorship of 0·73, and estimates of $U\overline{hs}$ for fecundity ranging between 1% and 8% (with an average of 3·6%). The conclusion was that the mutation pressure is very high in natural populations of *Daphnia*. The estimates of U could be in agreement with the large mutation rates inferred from molecular studies, but the high average heterozygous effect ($U\overline{hs} = 3·6\%$) strongly disagrees with the results from MA experiments. The highest per generation decline in fitness observed from MA experiments is of the order of 1% (more often around 0·1%) for homozygotes ($U\overline{s}$; see, for example, García-Dorado *et al.*, 1999) and, assuming partial recessivity of deleterious mutations (say $\overline{h} \approx 0·2$; Charlesworth & Hughes, 2000; García-Dorado & Caballero, 2000), the corresponding rate of decline for heterozygotes ($U\overline{hs}$) should be smaller than 1% (and even 0·1%). Therefore, the results obtained by Deng *et al.* (2006) for $U\overline{hs} = 3·6\%$ in *Daphnia* seem to be extremely large in comparison with most previous estimates, and it is possible that the discussed sources of bias may be inflating the estimates substantially.

Acknowledgements

References

Baer, C. F., Shaw, F., Steding, C., Baumgartner, M., Hawkins, A., Houppert, A., Mason, N., Reed, M., Simonelic, K., Woodard, W. & Lynch, M. (2005). Comparative evolutionary genetics of spontaneous mutations affecting fitness in rhabditis nematodes. *Proceedings of the National Academy of Sciences of the USA* **102**, 5785–5790.

Charlesworth, B. & Hughes, K. A. (2000). The quantitative genetics of life history traits. In *Evolutionary Genetics: From Molecules to Morphology* (ed. R. S. Singh & C. B. Krimbas), pp. 369–392. Cambridge: Cambridge University Press.

Charlesworth, B., Borthwick, H., Bartolomé, C. & Pignatelli, P. (2004). Estimates of the genomic mutation rate for detrimental alleles in *Drosophila melanogaster*. *Genetics* **167**, 815–826.

Deng, H.-W., Li, J., Pfrender, M. E., Li, J.-L. & Deng, H. (2006). Upper limit of the rate and per generation effects of deleterious genomic mutations. *Genetical Research* **88**, 57–65.

Denver, D. R., Morris, K., Lynch, M. & Thomas, W. K. (2004). High mutation rate and predominance of insertions in the *Caenorhabditis elegans* nuclear genome. *Nature* **430**, 679–682.

Falconer, D. S. & Mackay, T. F. C. (1996). *An Introduction to Quantitative Genetics*, 4th edn. Harlow: Longman.

Fernández, B., García-Dorado, A. & Caballero, A. (2005). The effect of antagonistic pleiotropy on the estimation of the average coefficient of dominance of deleterious mutations. *Genetics* **171**, 2097–2112.

Fry, J. D. & Heinsohn, S. L. (2002). Environmental dependence of mutational parameters for viability in *Drosophila melanogaster*. *Genetics* **161**, 1155–1167.

García-Dorado, A. & Caballero, A. (2000). On the average coefficient of dominance of deleterious spontaneous mutations. *Genetics* **155**, 1991–2001.

García-Dorado, A., López-Fanjul, C. & Caballero, A. (1999). Properties of spontaneous mutations affecting quantitative traits. *Genetical Research* **74**, 341–350.

García-Dorado, A., Caballero, A. & Crow, J. F. (2003). On the persistence and pervasiveness of a new mutation. *Evolution* **57**, 2644–2646.

García-Dorado, A., López-Fanjul, C. & Caballero, A. (2004). Rates and effects of deleterious mutations and their evolutionary consequences. In *Evolution of Molecules and Ecosystems* (ed. A. Moya & E. Font), pp. 20–32. Oxford: Oxford University Press.

Halligan, D. L. & Keightley, P. D. (2006). Ubiquitous selective constraints in the *Drosophila* genome revealed by a genome-wide interspecies comparison. *Genome Research* **16**, 875–884.

Joseph, S. B. & Hall, D. W. (2004). Spontaneous mutations in diploid *Saccharomyces cerevisiae*: more beneficial than expected. *Genetics* **168**, 1817–1825.

Kusakabe, S. & Mukai, T. (1984). The genetic structure of natural populations of *Drosophila melanogaster*. XVII. A population carrying genetic variability explicable by the classical hypothesis. *Genetics* **108**, 393–408.

Li, J.-L., Li, J. & Deng, H.-W. (1999). The effect of overdominance on characterizing deleterious mutations in large natural populations. *Genetics* **151**, 895–913.

Rodríguez-Ramilo, S. T., Pérez-Figueroa, A., Fernández, B., Fernández, J. & Caballero, A. (2004). Mutation–selection balance accounting for genetic variation for viability in *Drosophila melanogaster* as deduced from an inbreeding and artificial selection experiment. *Journal of Evolutionary Biology* **17**, 528–541.

Roff, D. A. (1997). *Evolutionary Quantitative Genetics*. New York: Chapman and Hall.

Schoen, D. J. (2005). Deleterious mutation in related species of the plant genus *Amsinckia* with contrasting mating systems. *Evolution* **59**, 2370–2377.