THE ROYAL SOCIETY

Received 5 December 2001
Accepted 28 January 2002
Published online 22 April 2002

Evolution of functionally conserved enhancers can be accelerated in large populations: a population-genetic model

Ashley J. R. Carter and Günter P. Wagner*

Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06520-8106, USA

The evolution of *cis*-regulatory elements (or enhancers) appears to proceed at dramatically different rates in different taxa. Vertebrate enhancers are often very highly conserved in their sequences, and relative positions, across distantly related taxa. In contrast, functionally equivalent enhancers in closely related *Drosophila* species can differ greatly in their sequences and spatial organization. We present a population-genetic model to explain this difference. The model examines the dynamics of fixation of pairs of individually deleterious, but compensating, mutations. As expected, small populations are predicted to have a high rate of evolution, and the rate decreases with increasing population size. In contrast to previous models, however, this model predicts that the rate of evolution by pairs of compensatory mutations increases dramatically for population sizes above several thousand individuals, to the point of greatly exceeding the neutral rate. Application of this model predicts that species with moderate population sizes will have relatively conserved enhancers, whereas species with larger populations will be expected to evolve their enhancers at much higher rates. We propose that the different degree of conservation seen in vertebrate and *Drosophila* enhancers may be explained solely by differences in their population sizes and generation times.

Keywords: enhancer evolution; fitness valley; compensatory mutation; population size

1. INTRODUCTION

In the same manner that coding sequences can often be highly conserved, the nucleotide sequences of vertebrate enhancer regions are often highly conserved, even in distantly related taxa (Aparicio *et al.* 1995; Chiu & Hamrick 2002). For example, in a comparison of the 200 base pair (bp) early enhancer of *Hoxc8* in 29 species of mammals (including representatives from nine eutherian orders and one marsupial), the complete nucleotide sequences of this region were 90% similar across all taxa (Shashikant & Ruddle 1996; Shashikant *et al.* 1998). Additionally, five experimentally characterized *cis*-acting elements in this region were 100% conserved between the taxa, with the exception of five baleen whale species that shared a 4 bp deletion in a single element. This level of conservation is typical among vertebrates, and has been developed into a method for the identification of putative *cis*-regulatory elements (Aparicio *et al.* 1995; Tagle *et al.* 1988; Gumucio *et al.* 1993; Sumiyama *et al.* 2001). Other examples, which demonstrate the high evolutionary conservation of enhancer sequences and organization, include: the 5′ region of the *SRY* (sex-determining, region Y) gene in mammals (Margarit *et al.* 1998); *cis*-regulatory elements of various actin orthologues in vertebrate taxa, ranging from humans to teleost fishes (Liu *et al.* 2000); the locus control region of the β-globin genomic domain in mammals (Hardison *et al.* 1997); and the *cis*-regulatory elements of the *Pax-6* gene between humans and quails (Plaza *et al.* 1999).

Conversely, recent studies of closely related invertebrate species have revealed that enhancers with conserved functions can vary greatly in their sequences, and in the arrangement and number of transcription-factor binding sites. Most notably, in a comparison of functionally equivalent *even-skipped* (*eve*) stripe 2 enhancers in four *Drosophila* species (*melanogaster*, *yakuba*, *erecta* and *pseudoobscura*), many substitutions in binding sites for bicoid, hunchback, Kruppel and giant, as well as large differences in the overall size of the enhancer region, were found (Ludwig *et al.* 1998). The enhancer regions of three of these species have an additional, functionally important, bicoid binding site not present in *D. pseudoobscura* or another species, *D. picticornis*. An examination of the sequence and expression patterns of the *D. pseudoobscura esterase-5B* and *D. melanogaster esterase 6* genes (Tamarina *et al.* 1997) concluded that the conservation of expression patterns need not be accompanied by preservation of the corresponding *cis*-regulatory elements. Studies of *Drosophila* glucose dehydrogenase (*Gld*) expression in the *melanogaster* subgroup (Ross *et al.* 1994) have demonstrated that one species, *D. teissieri*, lacks three elements in its enhancer region that are necessary for expression in the ejaculatory ducts of *D. melanogaster*, and also lacks *Gld* expression in this domain. Surprisingly, *D. erecta* and *D. yakuba* also lack these elements, but they retain the expression patterns observed in the non-*teissieri* species (Stern 2000), suggesting the presence of as yet undiscovered compensatory mechanisms.

These somewhat paradoxical observations indicate that enhancer sequence conservation is high in vertebrate taxa but can be comparatively low, for example, in *Drosophila* and some other invertebrate taxa, such as ascidians (Takahashi *et al.* 1999), houseflies (Hancock *et al.* 1999) and *Tribolium castaneum* (Hancock *et al.* 1999). There is currently no explanation for this difference. We present a

*Author for correspondence (gunter.wagner@yale.edu).

DOI 10.1098/rspb.2002.1968

population-genetic model that seeks to explain this paradox. The model is based on the assumption that the evolution of functional *cis*-enhancer elements is due to pairs of deleterious mutations that, in combination, compensate for their individual deleterious effects.

Empirical evidence for the feasibility of this type of mechanism comes from a study of substitutions in the *cis*-regulatory and coding sequences of the alcohol dehydrogenase (*Adh*) locus in 10 *Drosophila* species and two medfly species (Parsch *et al.* 1997). We analysed these sequences for phylogenetic correlations between multiple nucleotide substitutions, looking for pairs of substitutions that may indicate compensatory changes. We found several possible sets of correlated mutations and used site-directed mutagenesis to show that, in the case of one pair, a mutation at site 1756 (in the 3′ untranslated region) compensated for a silent substitution at site 819 (in exon 2) that reduced *Adh* activity by 15%. Ludwig *et al.* (2000) explain the aforementioned observations of sequence divergence in Eve 2 stripe enhancer sequences in closely related *Drosophila* species, by suggesting that this type of evolution may be due to the fixation of a series of slightly deleterious mutations by random drift, and subsequent selection for compensatory mutations (Kimura 1983; Ohta & Tachida 1990). This suggestion can account for enhancer sequence divergence while maintaining conserved function, but also predicts that the rate of sequence evolution should decrease with increasing population size. However, if we make the reasonable assumption that vertebrates have smaller effective population sizes than those of *Drosophila*, it cannot explain why enhancer sites in vertebrates evolve much more slowly than in *Drosophila*. One possibility is that, for some undiscovered reason, compensatory mutations are more likely in other taxa than in vertebrates, and the realized rate of sequence evolution is therefore higher in these groups than in vertebrates. We are, however, not aware of any evidence at present that would suggest this to be the case. In the model presented below, we also assume that there are no fundamental differences in the biochemistry of transcriptional regulation between invertebrates and vertebrates, but focus instead on the possible role of effective population size differences between these groups.

These authors [editor's error in preparation]

2. THE MODEL

The model presented here differs from previous models of compensatory mutations in several respects. The largest departure is our incorporation of two distinct pathways that the population can use to reach the alternative functional genotype. In the first pathway, one of the deleterious alleles fixes by random drift and the compensatory mutation occurs afterwards. This pathway becomes increasingly less probable for larger effective population sizes, because the fixation of the deleterious mutation becomes increasingly more difficult due to the increase of the factor *Ns*. This pathway is the one usually considered. In the second, less commonly considered pathway the compensatory mutation occurs while the population is still segregating a number of copies of the initial deleterious allele. Although this pathway may also be expected to be slower in larger populations, due to selection being more efficient against the individuals carrying the deleterious allele (larger *Ns* again), we show that this is often not the case. Figure 1 illustrates the fitness landscape and the set of alleles that we are considering, and figure 2 illustrates the two paths that a population can take. We make no statement as to the precise nature of the mutations we consider. Deleterious mutations that affect the positions or organization of loci, and whose fitness effects can be compensated fall into the scope of our model as well.

For formulating the model, we derive algebraically simple and intuitive equations for the mean time to fixation of the compensatory mutation. These equations require simplifying assumptions, but do make accurate predictions when compared with the results of individual-based computer simulations. The first pathway from figure 2 consists of two serial segments—genesis and fixation of the deleterious allele—followed by genesis and fixation of the final compensatory allele. As each separate expected waiting time exhibits an exponential distribution, the sum of the expected times for the events in each pathway will be the expected time for that pathway. The expected time to fixation of the compensatory mutation by this pathway is therefore

$$t_1 = \frac{1}{2N2\mu p_{\mathrm{del}}} + \frac{1}{2N\mu p_{\mathrm{adv}}}, \tag{2.1}$$

where N is the effective population size, μ is the mutation rate, p_{del} is the probability of fixation of the first, deleterious, mutation and p_{adv} is the probability of fixation of the second, advantageous, mutation. In equation (2.1), we omit the time required to fix the mutations, as these times are far shorter than the waiting times for mutations destined for fixation.

The second pathway consists of the genesis of a double mutant, without first fixing the deleterious initial mutation, and the time required for that allele to reach fixation in the population.

$$t_2 = \frac{1}{2N2\mu\,\alpha\mu p_{\mathrm{comp}}} + t_{\mathrm{comp}}, \tag{2.2}$$

where the parameters N and μ are the same as in equation (2.1) and α represents the absolute number of deleterious alleles that segregate in a population prior to loss per mutant generated, p_{comp} is the probability of fixation of the double mutant against the mainly wild-type background and t_{comp} is the mean time, conditional upon fixation, required to fix the double mutant against the mainly wild-type background (we include the fixation time here as it may become quite large for larger populations in which the waiting time becomes smaller). The formulae for α and t_{comp} are very complicated and are presented in algebraic forms in Appendix A. This equation omits the effects of multiple independently derived double mutants arising simultaneously, but our individual-based simulations show that any inaccuracy caused by this assumption is negligible.

The two pathways discussed above must be combined to predict the total rate of evolution by compensatory mutations. The preceding theory has been presented in the form of expected times; we will now consider the rate of fixation of these pairs of compensatory mutations. This shift will allow us to combine the pathways in a more straightforward manner, and compare our results with more familiar rates of evolution. In general, if there are

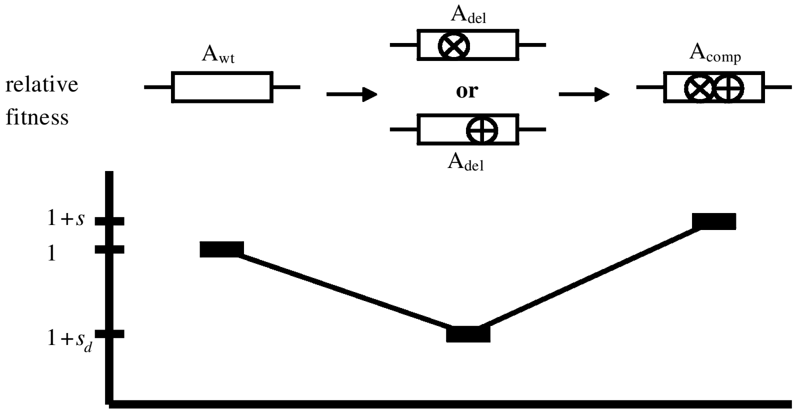

Figure 1. Schematic diagram of the fitness landscape. This single-step fitness valley represents a case in which either of two initial mutations is individually deleterious, whereas in combination they rescue the fitness or increase it. The mutations considered in the model need not be nucleotide substitutions, as this figure may seem to imply, but can also involve positional changes of elements or reorganization events. The example presented in detail omits recombination and assumes both intermediate alleles are equivalent in fitness effects, allowing us to label the three haplotypes A_{wt}, A_{del} and A_{comp}, where A_{wt} is the initial (wild-type) haplotype, A_{del} is either of the deleterious haplotypes and A_{comp} is the (compensatory) double mutant. The actual model uses diploid individuals, and the fitnesses of the genotypes may, therefore, differ.

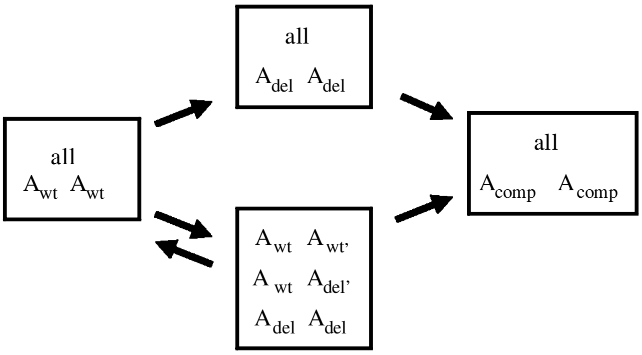

Figure 2. Diagram of possible paths for population. The two possible paths that a population may take to fix a pair of compensatory mutations are shown. Each box indicates the genotypes that a population in that state is segregating, while the arrows indicate the transitions discussed in the text. In the first pathway (top), the population fixes the deleterious intermediate allele (A_{del}) and then generates the compensatory allele (A_{comp}) and fixes that. In the second pathway (bottom), the deleterious allele arises and, while segregating in the mostly wild-type population (A_{wt}), it mutates and gives rise to a compensatory genotype that then fixes. The overall rate of transition must take both pathways into account.

two independent ways to generate a final result, with mean expected times of t_1 and t_2, the overall rate, R, can be closely approximated by the following equation,

$$R = \left(\frac{1}{t_1} + \frac{1}{t_2}\right). \tag{2.3}$$

For the remainder of this study, we concern ourselves exclusively with the case in which the mutations are tightly linked, as is the case in *cis*-regulatory elements, the intermediate alleles are completely recessive and the final compensatory allele is slightly advantageous and dominant to the other alleles. We further simplify the formula by making the reasonable assumption that the second pathway only occurs for large population sizes, allowing us to approximate the probability of fixation of an advantageous allele by $2s$ for this pathway (Kimura 1962). Using these assumptions the formula estimating the total rate of fixation of pairs of compensatory mutations by either pathway is derived in Appendix A and given by:

$$R = \left(\frac{1}{4N\mu \dfrac{1-\mathrm{e}^{-s_d}}{1-\mathrm{e}^{-2Ns_d}}} + \frac{1}{4N\mu \dfrac{1-\mathrm{e}^{-s+s_d}}{1-\mathrm{e}^{-2N(s-s_d)}}} \right)^{-1} + \left(\frac{1}{8Ns\mu^2 \sqrt{\dfrac{2N\pi}{-s_d}}} + t_{\mathrm{comp}} \right)^{-1}, \tag{2.4}$$

where s_d is the selective disadvantage of the intermediate allele (see figure 1 for details).

3. MODEL AND SIMULATION RESULTS

In figure 3, the behaviour of equation (2.4) is shown and compared with the results of individual-based computer simulations as described in Appendix A. The qualitative shape of equation (2.4) in figure 3 is similar for a wide range of μ, s and s_d values. The model consistently predicts the deep trough of evolutionary rate seen in figure 3, with the rates for smaller, and larger, populations at orders of magnitude larger than those in the trough. The interpretation of this shape is fairly straightforward, for smaller populations to the left of the trough, pathway 1 is favoured, while to the right, the second pathway predominates, but intermediate populations generate conditions such that neither pathway can operate effectively. The reason for the increase in evolutionary rate for large populations is the larger number of recessive, deleterious alleles that can segregate in a population of large size, as compared with a small population (see Appendix A), creating a larger target for the compensatory mutations. The decline in rate for even larger populations is entirely due to the increased time required to fix the final haplotype in these populations.

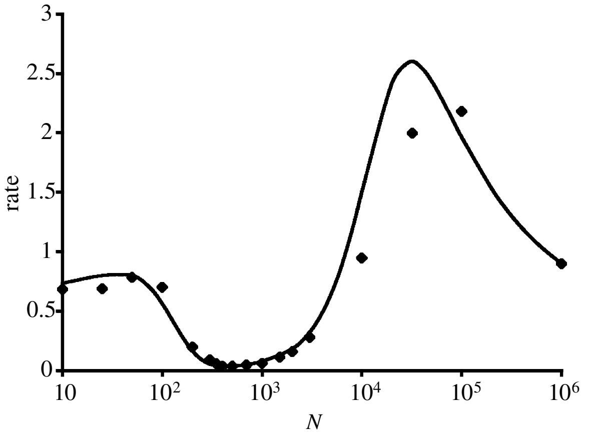

Figure 3. Plot of the rate of fixing a double mutation versus the population size (*N*). The per-generation rate of fixing a double mutation is divided by the mutation rate for the purpose of comparison to the rate of neutral evolution. Points represent the mean time from 100 trials (10 for $N = 10^6$) of an individual-based computer simulation. The line indicates the theoretical prediction made by equation (2.4) divided by μ. For these trials, $s = 0.001$, $s_d = -0.01$, $\mu = 10^{-5}$ and the deleterious allele is completely recessive.

Using the derivative of equation (2.4) with respect to *N*, we can determine the population values that exhibit minimal and maximal rates of evolution due to compensatory mutations. Thus, for each set of parameters we can obtain two values of *N*, representing the population sizes that have the minimum and maximum rates. Figure 4 shows these pairs of population sizes for a wide range of *s* and s_d values.

From figure 4 we can see that there are two distinct domains of population sizes. In the range $10 \leqslant N \leqslant 3000$, the rate of fixation of the compensatory alleles is the lowest, while in the range $10\,000 \leqslant N \leqslant 400\,000$, the rate increases to a maximum. Examination of the rates reveals that the maximal rates of evolution are generally greater than the neutral rates. We have used a mutation rate of 10^{-5} for all of the examples shown so far. Figure 5 shows plots of the evolution rate predicted by equation (2.4) (relative to the neutral rate of evolution) for a wide range of mutation rates and selective parameters. While smaller mutation rates obviously slow the transition down for any given population, a rate trough is still visible for moderate population sizes, and a rate maximum often exceeds the neutral rate for larger populations. Decreasing the mutation rate increases the population size with maximum rate, and widens the range of population sizes with low rates of evolution.

It is striking to note that even when the recessive deleterious intermediate allele is lethal, in its homozygous state, and the final fitness advantage is only $s = 0.001$, these pairs of mutations will fix at rates exceeding that of neutral mutations for large populations. Also striking is the magnitude of the differences in relative rate between populations of different size. The slopes of the plots in figure 5 indicate that a change in population size of only 100-fold corresponds to a change in the evolutionary rate of *ca*. 1000-fold. Estimates of effective population sizes (Kondrashov 1995; Nei & Graur 1984) of many vertebrates species are *ca*. 10^4–10^5, while invertebrates are estimated to have effective population sizes of *ca*. 10^6–10^7, values that would lead this model to predict 1000-fold differences, or more, in the rate of fixation of pairs of compensatory mutations of the nature considered above. Furthermore, invertebrate species with up to 100-fold larger populations also tend to have generation times that are significantly shorter than vertebrates (easily 10–100 times smaller), amplifying this rate difference when measured in units of physical time. Therefore, if we consider the rates in chronological terms, the difference in the rate of compensatory evolution between invertebrate and vertebrate species can be expected to easily exceed hundreds of thousands.

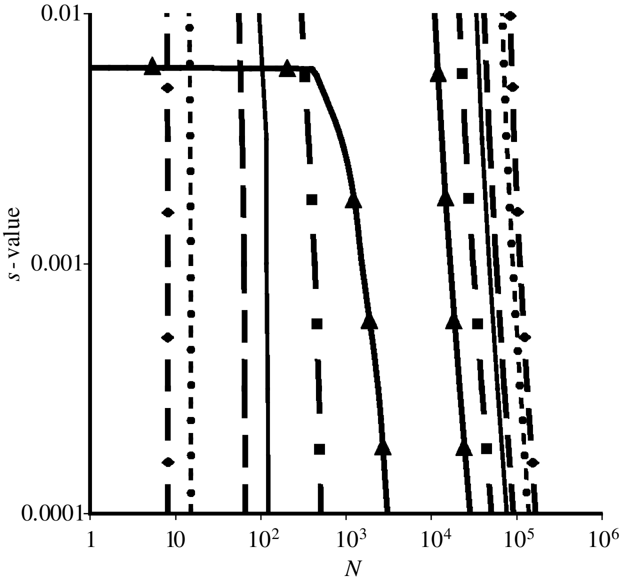

Figure 4. Plot of the population sizes when the derivative of equation (2.4) is zero. The vertical axis represents the increasing relative fitness of the compensatory allele, the horizontal axis is population size. Each value of s_d is represented by a pair of lines. The lower population size for each set of s_d and *s* values corresponds to the point of maximum time required, the higher population size for each set corresponds to the point of minimum time to fix the compensatory mutation. The mutation rate, μ, is 10^{-5}. Diamonds, $s_d = 1$; circles, $s_d = 0.5$; dashed lines, $s_d = 0.1$; solid lines, $s_d = 0.05$; squares, $s_d = 0.01$; triangles, $s_d = 0.001$.

If the assumptions regarding recessivity and linkage are relaxed, we must modify our predictions. Less recessive alleles (with the same s_d) leave the region to the left of the trough relatively unchanged as p_{del} varies little with recessivity (within an order of magnitude for relevant parameter combinations, results not shown). On the other hand, less recessive alleles slow the fixation rate for populations larger than this population, as fewer copies of each deleterious mutation are segregated prior to loss, since α is more sensitive to heterozygote fitness. The rate still increases with population size, but less quickly. If the mutations are not tightly linked then the second pathway becomes much slower as double mutants are destroyed by recombination (Christiansen *et al.* 1998), effectively reducing p_{comp}. The qualitative shape of the plots seen in figure 5 is still seen for moderate recombination rates (recombination rate, $c < 1/N$, results not shown), but with a shallower incline after the trough. In the case we seek

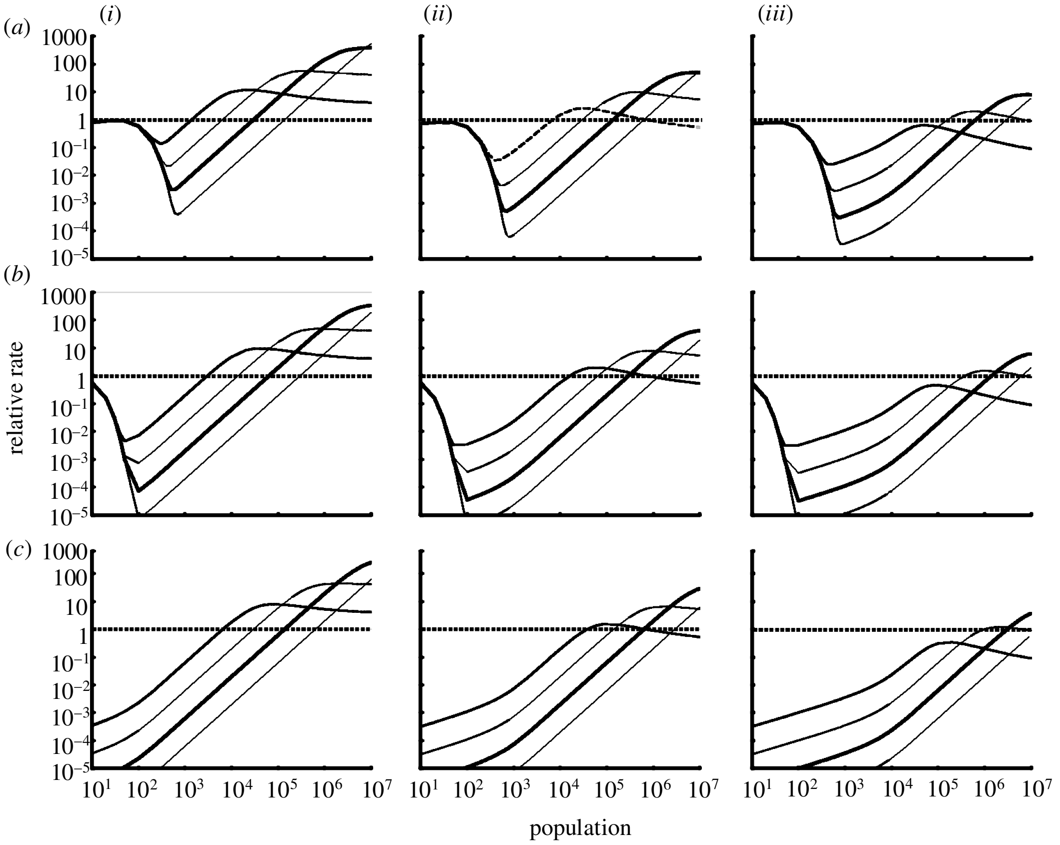

Figure 5. Plots of equation (2.4) divided by μ for varying mutation rates. This relative rate is represented on a log scale ranging from 10^{-5} to 10^3; the dotted lines indicate a rate equal to that of the neutral rate for single, neutral substitutions ($R = \mu/\mu = 1$). Population is on a log scale, ranging from 10^1 to 10^7. Values of s_d for (*a*) −0.01; (*b*) −0.1; (*c*) −1. Values of *s* for columns, (*i*) 0.01; (*ii*) 0.001; (*iii*) 0.0001. Each plot shows four values of μ, 10^{-5}, 10^{-6}, 10^{-7}, 10^{-8}, larger values having faster rates for smaller populations. The dashed uppermost curve in the (*a*)(*ii*) plot corresponds to that in figure 3.

to model, tightly linked enhancer elements within relatively small *cis*-regulatory regions, recombination is likely to be sufficiently low that this effect can be ignored.

4. OTHER STUDIES

There is an extensive literature on models of compensatory mutations. Several researchers (Birky & Walsh 1988; Charlesworth 1994) have examined the so-called Hill–Robertson effect, by which the probability of fixation of a deleterious allele can be assisted by linkage between that locus and a different locus exhibiting positively or negatively selected alleles. The proposed mechanism for this phenomenon is a reduction of the effective population size for the locus of interest due to selection at the linked locus. This mechanism predicts decreased rates of evolution as the absolute population size increases. The Hill–Robertson effect does not, however, explain the difference in the rate of enhancer evolution between *Drosophila* and vertebrate species, assuming that the latter have generally smaller populations than insects.

Stephan (1996) used diffusion techniques to examine the expected time to fix pairs of deleterious and compensatory mutations. Dominance effects were ignored and the final combination of alleles had the same fitness as the initial state. While Stephan focused his attention upon the effects of recombination rather than population size, he did note that a population can move from one fitness peak to another without loss of significant mean population fitness if only a small number of individuals possess the deleterious intermediate genotypes. A similar study by Innan & Stephan (2001) studied the time to fix a pair of compensatory mutations. Their model assumed $\mu < 1/2N$, additive deleterious mutations, a final fitness identical to the initial fitness, and pseudosampling simulations rather than the individual-based ones used in our study. As in the previous study mentioned, the authors focused their attention upon recombination and linkage disequilibrium and did not separate the effects of population size and the factor *Ns*, where *s* is the fitness cost of the deleterious allele. Philips (1996) also used diffusion methods to investigate several models of compensatory evolution and concluded that increasing population size dramatically decreases the rate of evolution, the opposite of the conclusion reached in this study. Gillespie (1999, 2000) has considered models similar to those used by us. He illustrated an example with a qualitatively similar relationship between expected time to fix both alleles and population size to that derived in this paper. In another study, Gillespie (1984) considered a model that is similar to the second pathway we present, but uses effectively additive deleterious alleles instead of recessive ones since he considered the genotypes to be haploid. This analysis led Gillespie to conclude that the expected time to fix the double mutation, even though it does decrease for larger populations, is prohibitively long.

In general, the current consensus is that larger populations make crossing fitness valleys more and more unlikely to the point of impossibility, contrary to our results. To our knowledge, the only studies, other than those of Gillespie (1984, 1999), that consider larger populations to be more advantageous than smaller ones for the rate of crossing fitness valleys are those of Stone & Wray (2001) and Hansen *et al.* (2000). Stone & Wray used simulations based on sequence data to predict the time for a population to acquire novel enhancer sequences via neutral mutations, finding that larger populations acquire such sites more quickly than smaller ones. Hansen *et al.* derived a model that considers a locus and its non-expressed duplicate, in which the deactivated locus transfers neutrally accumulated mutations en masse to the primary locus of interest by gene conversion, crossing a multistep fitness valley.

5. CONCLUSION

The main result of this study is that, for recessive, deleterious mutations and slightly advantageous compensatory mutations, the relationship between the rate of sequence evolution and population size is highly nonlinear. In moderately small populations, the rate of evolution, as measured by the mean number of generations required to fix the double mutant, is much slower than the neutral rate. For moderately large populations, the rate of evolution can be several times faster, even exceeding the neutral rate. Hence, differences of one or two orders of magnitude in effective population size can have dramatic implications for the rate of enhancer evolution through compensatory mutations. This difference in relative rate will be magnified by the fact that those populations with very high effective population sizes are also likely to be ones in which generation times are much smaller than populations of moderate size.

Enhancer elements are often very small, and mutations in the same element would be very tightly linked, as our theoretical model assumes. Enhancers are also often clustered in relatively small regions, meaning that changes in position or organization of elements within this region would also tend to be tightly linked. If enhancer elements can evolve through a series of deleterious and compensatory mutations, the observed differences in the rate of enhancer evolution between vertebrates and *Drosophila* may be due to demographic and generation time differences alone, rather than due to, as yet undiscovered, differences in mechanisms of transcriptional regulation. If this explanation is correct, one would predict that the rate of enhancer evolution in *Drosophila* species with small effective population sizes should be lower than in *D. melanogaster* and its kin, and more like that found in vertebrates. Conversely, it is predicted that vertebrate species with very large population sizes should show a higher than neutral rate of enhancer evolution.

The authors thank Reinhard Bürger, Chi-Hua Chiu, Thomas Hansen and Joachim Hermisson for comments on the manuscript. The financial support of National Science Foundation grant #9905403 (G.P.W.), a Department of Defense graduate fellowship (A.J.R.C.) and a Leo F. Rettger graduate fellowship (A.J.R.C.) are gratefully acknowledged.

APPENDIX A

(a) α *and* t$_{\text{comp}}$ *values*

α represents the mean number of deleterious alleles (A_2), per initial mutant allele, that segregate in a population prior to loss. Formulae for α were derived by Li & Nei (1972) and are as follows:

$$\alpha = \left\{ \begin{array}{ll} \dfrac{1}{-hs_d} & h \geqslant 0.3 \\ 2N\sqrt{\dfrac{\pi}{-2Ns_d(1-2h)}}\,e^{A^2}(1-\operatorname{erf}(A)) & 0.3 \geqslant h > 0 \\ \sqrt{\dfrac{2N\pi}{-s_d}} & h = 0 \end{array} \right\}$$

where $A = h\sqrt{\dfrac{-2Ns_d}{1-2h}}$,

$$\operatorname{erf}(A) = \int_0^A \sqrt{\frac{1}{2\pi}}\, e^{-\frac{t^2}{2}} dt$$

where N is the population size, and the fitnesses of the genotypes are as follow: $A_1A_1 = 1$, $A_1A_2 = 1 - hs_d$ and $A_2A_2 = 1 - s_d$. We compared these equations to individual-based computer simulations as described below. Figure 6 represents these comparisons for recessive alleles ($h = 0$). The formulae are reasonably accurate at predicting α. These formulae indicate that α increases significantly with population size N for $h < 0.3$, while even for $h = 0.5$, larger populations (with higher Ns_d values) do not reduce the overall numbers of deleterious alleles that segregate prior to loss. Equations for the number of generations required for fixation of an allele in a population, from an initial mutant copy, were derived by Kimura & Ohta (1969; eqn 39) and are as follows:

For neutral alleles we use the simple relationship $t_{\text{comp}} = 4N$.

For relatively advantageous alleles, A_3, where fitnesses are $w(A_2A_2) = 1$, $w(A_2A_3) = 1 + hs$, $w(A_3A_3) = 1 + s$:

$$t_{\text{comp}} = \int_{\frac{1}{2N}}^{1} \psi(\xi)u(\xi)(1 - u(\xi))d\xi + \frac{1 - u\left(\dfrac{1}{2N}\right)}{u\left(\dfrac{1}{2N}\right)} \int_0^{\frac{1}{2N}} \psi(\xi)u^2(\xi)d\xi,$$

u and ψ are given by the following:

$$u(p) = \frac{\int_0^p G(x)dx}{\int_0^1 G(x)dx},\quad \psi(x) = \frac{2\int_0^1 G(x)dx}{V_{\delta x}G(x)},$$

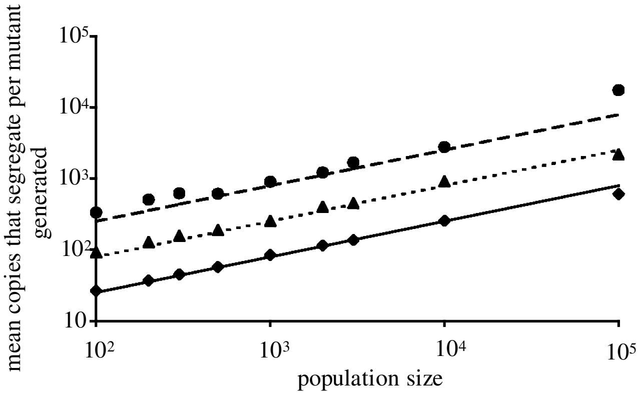

Figure 6. Plots of α. The mean number of deleterious alleles that segregate per initial mutation are shown. Lines represent prediction based upon the equation from Li & Nei (1972; eqn 38), points are simulation data. In all these cases, the deleterious allele was recessive ($h = 0$). Values of relative fitness of the deleterious homozygote are represented by circles (0.99), triangles (0.9) and diamonds (0).

with $G(x) = \exp\left\{-\int_0^x \frac{2M_{\delta\xi}}{V_{\delta\xi}} d\xi\right\}$,

where $V_{\delta x}$ and $M_{\delta x}$ represent the mean change in frequency of the allele, due to drift and selection, respectively, and are given by

$$V_{\delta x} = \frac{x(1-x)}{2N} \text{ and } M_{\delta x} = x^3(2sh - s) + x^2(s - 3sh) + x(sh).$$

(b) *Calculations for sample case*

For the special case described, we make the following substitutions: for relatively deleterious alleles, A_2, where fitnesses are $w(A_1A_1) = 1$, $w(A_1A_2) = 1 + hs_d$, $w(A_2A_2) = 1 + s_d$, $s_d < 0$, we use

$$p_{\text{del}} = \frac{1 - e^{-s_d}}{1 - e^{-2Ns_d}}.$$

This formula is not precisely accurate for all h values, but the difference in p_{del} for varied h is not large (results not shown) and we use this equation for the sake of clarity in our model.

For relatively advantageous dominant alleles, A_3, where fitnesses are $w(A_2A_2) = 1$, $w(A_2A_3) = 1 + s$, $w(A_3A_3) = 1 + s$, we use

$$p_{\text{comp}} = \frac{1 - e^{-2s}}{1 - e^{-4Ns}}.$$

In the first pathway, the fitness advantage is actually $(s - s_d)$ as the compensatory allele fixes against a background of complete fixation of the deleterious single mutation, so we use

$$p_{\text{comp}} = \frac{1 - e^{-2(s-s_d)}}{1 - e^{-4N(s-s_d)}}.$$

(c) *Computer simulations*

All individual-based computer simulations were written and run in C++ on IBM compatible computers using Microsoft Visual Studio 97.

Individual-based computer simulations to obtain mean times to fix a pair of compensatory mutations were performed as follows. The population in each trial was initialized with all N diploid individuals homozygous for the wild-type allele (allele 1). In each generation, every individual allele was given a fixed probability of mutating to the next allele in the sequence 1 → 2 → 3. Mutations are considered irreversible. As we omit recombination, both deleterious alleles are effectively identical and can be collapsed into a single haplotype, for this reason the mutation rate from 1 to 2 is twice that of 2 to 3. After this mutation step, N pairs of individuals were picked (sequentially with replacement and probabilities weighted by their fitness) and mated to produce the members of the next generation. The simulation ended when all individuals were homozygous for the final compensatory mutation (allele 3). The total number of generations that transpired until the end of the simulation was recorded in each trial.

Trials were performed for the nine combinations of $s = 0.1, 0.01, 0.001$ with $s_d = -0.01, -0.1, -1$ for population sizes ranging from 100 to 100 000 and $\mu = 10^{-5}$. In all cases, the computer simulations and theoretical predictions matched very closely, validating the accuracy of the approximations used in the model.

Individual-based computer simulations to obtain the mean number of deleterious alleles that segregate prior to loss were performed as follows. The population in each trial was initialized with all $N - 1$ diploid individuals homozygous for the wild-type allele, and one individual heterozygous for the wild-type allele and the deleterious allele (allele 2). In each generation, N pairs of individuals were picked (sequentially with replacement and probabilities weighted by their fitness) and mated to produce the members of the next generation. The simulation ended when all individuals were homozygous for either the wild-type allele or, in rare cases, the deleterious allele. The total number of copies of the deleterious allele that segregated prior to the end, including the initial mutant in generation zero, of the simulation was recorded in each trial. Results are shown in figure 6.

REFERENCES

Aparicio, S., Morrison, A., Gould, A., Gilthorpe, J., Chaudhuri, C., Rigby, P., Krumlauf, R. & Brenner, S. 1995 Detecting conserved regulatory elements with the model genome of the Japanese puffer fish, *Fugu rubripes*. *Proc. Natl Acad. Sci. USA* **92**, 1684–1688.

Birky, C. & Walsh, J. 1988 Effects of linkage on rates of molecular evolution. *Proc. Natl Acad. Sci. USA* **85**, 6414–6418.

Charlesworth, B. 1994 The effect of background selection against deleterious mutations on weakly selected, linked variants. *Genet. Res.* **63**, 213–227.

Chiu, C.-H. & Hamrick, M. W. 2002 Evolutionary developmental biology in primates; using the limb as a model system. *Evol. Anthropol.* (In the press.)

Christiansen, F., Otto, S. & Bergman, A. 1998 Waiting with and without recombination: the time to production of a double mutant. *Theor. Popul. Biol.* **53**, 199–215.

Gillespie, J. 1984 Molecular evolution over the mutational landscape. *Evolution* **38**, 1116–1129.

Gillespie, J. 1999 The role of population size in molecular evolution. *Theor. Popul. Biol.* **55**, 145–156.

Gillespie, J. 2000 The neutral theory in an infinite population. *Gene* **261**, 11–18.

Gumucio, D., Shelton, D., Bailey, W., Slightom, J. & Goodman, M. 1993 Phylogenetic footprinting reveals unexpected complexity in trans factor binding upstream from the ε-globin gene. *Proc. Natl Acad. Sci. USA* **90**, 6018–6022.

Hancock, J., Shaw, P., Benneton, F. & Dover, G. 1999 High sequence turnover in the regulatory regions of the developmental gene hunchback in insects. *Mol. Biol. Evol.* **16**, 253–265.

Hansen, T., Carter, A. & Chiu, C.-H. 2000 Gene conversion may aid adaptive peak shifts. *J. Theor. Biol.* **207**, 495–511.

Hardison, R., Slightom, J., Gumucio, D., Goodman, M., Stojanovic, N. & Miller, W. 1997 Locus control regions of mammalian β-globin gene clusters: combining phylogenetic analyses and experimental results to gain functional insights. *Gene* **205**, 73–94.

Innan, H. & Stepan, W. 2001 Selection intensity against deleterious mutations in RNA secondary structures and rate of compensatory nucleotide substitutions. *Genetics* **159**, 389–399.

Kimura, M. 1962 On the probability of fixation of mutant genes in a population. *Genetics* **47**, 713–719.

Kimura, M. 1983 *The neutral theory of molecular evolution*. New York: Cambridge University Press.

Kimura, M. & Ohta, T. 1969 The average number of generations until fixation of a mutant gene in a finite population. *Genetics* **61**, 763–771.

Kondrashov, A. 1995 Contamination of the genome by very slightly deleterious mutations: why have we not died 100 times over? *J. Theor Biol.* **175**, 583–594.

Li, W.-H. & Nei, M. 1972 Total number of individuals affected by a single deleterious mutation in a finite population. *Am. J. Hum. Genet.* **24**, 667–679.

Liu, T., Wu, J. & He, F. 2000 Evolution of *cis*-acting elements in 5′ flanking regions of vertebrate actin genes. *J. Mol. Evol.* **50**, 22–30.

Ludwig, M. Z., Patel, N. H. & Kreitman, M. 1998 Functional analysis of eve stripe 2 enhancer evolution in *Drosophila*: rules governing conservation and change. *Development* **125**, 949–958.

Ludwig, M., Bergman, C., Patel, N. & Kreitman, M. 2000 Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature* **403**, 564–567.

Margarit, E., Guillen, A., Rebordosa, C., Vidal-Taboada, J., Sanchez, M., Ballesta, F. & Oliva, R. 1998 Identification of conserved potentially regulatory sequences of the *SRY* gene from 10 different species of mammals. *Biochem. Biophys. Res. Commun.* **245**, 370–377.

Nei, M. & Graur, D. 1984 Extent of protein polymorphism and the neutral mutation theory. *Evol. Biol.* **17**, 73–118.

Ohta, T. & Tachida, H. 1990 Theoretical study of near neutrality. I. Heterozygosity and rate of mutant substitution. *Genetics* **126**, 219–229.

Parsch, J., Tanda, S. & Stephan, W. 1997 Site-directed mutations reveal long-range compensatory interactions in the *Adh* gene of *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **94**, 928–933.

Philips, P. 1996 Waiting for a compensatory mutation: phase zero of the shifting balance-process. *Genet. Res.* **67**, 271–283.

Plaza, S., Saule, S. & Dozier, C. 1999 High conservation of *cis*-regulatory elements between quail and human for the *Pax-6* gene. *Dev. Genes Evol.* **209**, 165–173.

Ross, J. L., Fong, P. & Cavener, D. 1994 Correlated evolution of the *cis*-regulatory elements and developmental expression of the *Drosophila Gld* gene in seven species from the subgroup *melanogaster*. *Dev. Genet.* **15**, 38–50.

Shashikant, C. S. & Ruddle, F. H. 1996 Combinations of closely situated *cis*-acting elements determine tissue-specific patterns and anterior extent of early *Hoxc8* expression. *Proc. Natl Acad. Sci. USA* **93**, 12 364–12 369.

Shashikant, C. S., Kim, C. B., Borbely, M. A., Wang, W. C. & Ruddle, F. H. 1998 Comparative studies on mammalian *Hoxc8* early enhancer sequence reveal a baleen whale-specific deletion of a *cis*-acting element. *Proc. Natl Acad. Sci. USA* **95**, 15 446–15 451.

Stephan, W. 1996 The rate of compensatory evolution. *Genetics* **144**, 419–426.

Stern, D. 2000 Perspective: evolutionary developmental biology and the problem of variation. *Evolution* **54**, 1079–1091.

Stone, J. R. & Wray, G. A. 2001 Rapid evolution of *cis*-regulatory sequences via local point mutations. *Mol. Biol. Evol.* **18**, 1764–1770.

Sumiyama, K., Kim, C.-B. & Ruddle, F. 2001 An efficient *cis*-element discovery method using multiple sequence comparisons based on evolutionary relationships. *Genomics* **71**, 260–262.

Tagle, D. A., Koop, B. F., Goodman, M., Slightom, J. L., Hess, D. L. & Jones, R. T. 1988 Embryonic ε and γ globin genes of a prosimian primate (*Galago crassicaudatus*). Nucleotide and amino acid sequences, developmental regulation and phylogenetic footprints. *J. Mol. Biol.* **203**, 439–455.

Takahashi, H., Mitani, Y., Satoh, G. & Satoh, N. 1999 Evolutionary alterations of the minimal promotor for notochord-specific brachyury expression in ascidian embryos. *Development* **126**, 3725–3734.

Tamarina, N., Ludwig, M. & Richmond, R. 1997 Divergent and conserved features in the spatial expression of the *Drosophila pseudoobscura* esterase-5 B gene and the esterase-6 gene of *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **94**, 7735–7741.

As this paper exceeds the maximum length normally permitted, the authors have agreed to contribute to production costs.