Correcting the Apparent Mutation Rate Acceleration at Shorter Time Scales under a Jukes–Cantor Model

Christopher Tuffley,*1 W. Timothy J. White,2 Michael D. Hendy,2 and David Penny3

1Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand
2Department of Mathematics and Statistics, University of Otago, Dunedin, New Zealand
3Institute of Molecular Biosciences, Massey University, Palmerston North, New Zealand

*Corresponding author: E-mail: c.tuffley@massey.ac.nz.
Associate editor: Barbara Holland

Abstract

At macroevolutionary time scales, and for a constant mutation rate, there is an expected linear relationship between time and the number of inferred neutral mutations (the “molecular clock”). However, at shorter time scales, a number of recent studies have observed an apparent acceleration in the rate of molecular evolution. We study this apparent acceleration under a Jukes–Cantor model applied to a randomly mating population, and show that, under the model, it arises as a consequence of ignoring short-term effects due to existing diversity within the population. The acceleration can be accounted for by adding the correction term $h_0e^{-4ht}/3$ to the usual Jukes-Cantor formula $p(t) = \frac{1}{2}(1 - e^{-4ht/3})$, where $h_0$ is the expected heterozygosity in the population at time $t = 0$. The true mutation rate $\mu$ may then be recovered, even if $h_0$ is not known, by estimating $\mu$ and $h_0$ simultaneously using least squares. Rate estimates made without the correction term (i.e., incorrectly assuming the population to be homogeneous) will result in a divergent rate curve of the form $\mu_{div} = \mu + C/t$, so that the mutation rate appears to approach infinity as the time scale approaches zero. Although our quantitative results apply only to the Jukes–Cantor model, it is reasonable to suppose that the qualitative picture that emerges also applies to more complex models. Our study, therefore, demonstrates the importance of properly accounting for any ancestral diversity, because it may otherwise play a dominant role in rate overestimation.

Key words: multiscale, short-term rate acceleration, ancestral polymorphism, divergent rate curve, Jukes–Cantor model.

Introduction

It is well known since Kimura (1968, 1983) that over time scales of tens and hundreds of millions of years, the longer term rate of molecular evolution, $k$, is expected to be virtually the same as the short-term rate of neutral mutations $\mu$, that is, $k \approx \mu$. However, at shorter time scales, a number of recent studies (e.g., García-Moreno 2004; Ho et al. 2005; Penny 2005; Millar et al. 2008; Henn et al. 2009; although there were some earlier indications [Fitch and Atchley 1985]) have observed an apparent acceleration in the rate of molecular evolution. This has led to debate as to the underlying causes of the apparent acceleration, and raised important questions as to how long it persists and how to correct for it.

A number of factors that may contribute to this apparent rate acceleration have been proposed, and we refer the reader to Ho et al. (2011) for a recent review. One such factor is the effect of ancestral polymorphism (Peterson and Masel, 2009; Charlesworth, 2010). Indeed, it is clear that, if ancestral polymorphism is present and not properly accounted for, then an apparent rate acceleration will inevitably be seen. Consider a comparison between two sequences drawn from a population at time $t = 0$, which is incorrectly assumed to be homogeneous. Any differences between the sequences due to polymorphism will appear to be mutations that have occurred in zero time, leading to an apparent infinite substitution rate at time zero. Moreover, on sufficiently short time scales, the average difference due to polymorphism will dominate any difference introduced by substitution, and to first-order the expected difference between sequences at times 0 and $t$ will be given by $p(t) = C + \mu' t$. (Here $\mu'$ will differ slightly from the mutation rate $\mu$, because some mutations will act to decrease the initial polymorphism.) Naively dividing by $t$ to recover $\mu$, under the assumption that the first-order approximation is given by $p(t) = \mu t$, then gives an apparent divergent rate curve $\mu' + C/t$, suggesting that unaccounted ancestral polymorphism will contribute a term of the form $C/t$ to rate estimates. This a priori estimate is in sharp contrast with the use of exponential decay curves of the form $k + \mu e^{-\lambda t}$ (e.g., by Ho et al. 2005) to fit estimated rates.

The effect of ancestral polymorphism on sequence divergence has been studied quantitatively by Peterson and Masel (2009), who calculate the expected divergence between two sequences as a function of time under the Moran model with selection. Their model provides a reasonable fit to data from Genner et al. (2007) and Henn et al. (2009), confirming that the observed rate acceleration is consistent with an underlying constant mutation rate. However, their model does not admit a simple closed form analytic expression, and this makes it difficult to correct for the apparent rate acceleration or quantify its duration.

The purpose of this article is to quantify the effect of ancestral polymorphism under a Jukes–Cantor-type model applied to a randomly mating population. This is a more
restrictive setting than that considered by Peterson and Masel, but the benefit of working in this simpler setting is that we are able to obtain simple and explicit analytic expressions for quantities of interest. In particular, we show that, under the model, ancestral polymorphism contributes a term $h_0 e^{-\frac{r}{2} \mu t}$ to the usual Jukes–Cantor formula $p(t) = \frac{3}{4} (1 - e^{-\frac{r}{2} \mu t / 3})$, where $h_0$ is the expected heterozygosity at time $t = 0$, and confirm that comparisons between sequences made under the incorrect assumption $h_0 = 0$ lead to a divergent rate curve of the form $\mu_{div} = \mu + C/t$. This has consequences at long time scales as well as short, because $C/t$ tends to zero comparatively slowly as $t$ tends to infinity (in particular, more slowly than any exponential decay), and so the effects of ancestral polymorphism may be relatively long lived. We show, however, that the true mutation rate $\mu$ may still be recovered from several observations even if the level of heterozygosity at time 0 is unknown, by estimating $h_0$ simultaneously with $\mu$ using least squares. We show further that our correction term also applies to other population structures (e.g., island models) under an appropriate assumption on $h_0$ and that the divergent rate curve $\mu_{div}$ provides a reasonable fit to mutation rate estimates taken from García-Moreno (2004) and Henn et al. (2009).

These results show that ancestral polymorphism, where present and unaccounted for, will be a significant contributing factor to mutation rate overestimation, with the magnitude of the effect approaching infinity as the time scale shrinks to zero. Our principal finding then is that an apparent rate acceleration at short time scales is consistent with—and indeed to be expected under—a constant mutation rate, in the presence of ancestral polymorphism that is not properly taken into account. This finding is in agreement with that of Peterson and Masel (2009).

Ancestral polymorphism is just one of several processes that are thought to contribute to the apparent rate acceleration and is unlikely to be the sole contributing factor. However, it is clearly vital that its effect be quantified and accounted for if we are to obtain meaningful and accurate rate estimates on short time scales and fully resolve the question of the apparent acceleration.

**Methods**

We consider a population of $N$ individuals evolving under a random mating process, with discrete generations, where at each generation the allele of each individual is replaced with a copy of the allele from the previous generation, chosen uniformly at random (Wright, 1931). For simplicity, we restrict our attention to haploid populations. We assume that alleles are $r$-state characters evolving under the $r$-state Jukes–Cantor model (Jukes and Cantor, 1969), with an instantaneous mutation rate of $\mu$ mutations per individual per generation per site. All mutations are assumed to be neutral. We orient time in the forward direction, so that populations at times $t > 0$ descend from the population at $t = 0$. We refer to the population at time 0 as the reference population and the population at a given time $t > 0$ of interest as the contemporary population.

Consider a pair of individuals chosen uniformly at random, one from each of the reference and contemporary populations, and let $P(t)$ be the probability that they have different character states at a fixed site. Suppose that the distribution of character states at that site in the reference population is given by $\pi = (\pi_1, \ldots, \pi_r)$. (Note that $\pi$ as used here is the distribution of states within the population at a given site, rather than the equilibrium distribution of the Jukes–Cantor model, which is the distribution of states across all sites within an individual.) Then

**Theorem.** The probability that uniformly randomly chosen members of the reference and contemporary populations have different states at a given site is given by

$$P(t) = h_0 e^{-\frac{r}{2} \mu t} + \frac{r - 1}{r} (1 - e^{-\frac{r}{2} \mu t}),$$

where

$$h_0 = P(0) = \sum_{i=1}^{r} \pi_i (1 - \pi_i) = \sum_{j=0}^{r-1} \pi_j \pi_{j+1}$$

is the expected heterozygosity at $t = 0$. In particular, for $r = 4$ (as for nucleotides) we get

$$P(t) = h_0 e^{-\frac{3}{4} \mu t} + \frac{3}{4} (1 - e^{-\frac{3}{4} \mu t}).$$

The proof of the theorem is given in the Appendix. Note that the second term in equation (1) is the standard probability under the $r$-state Jukes–Cantor model that a net change takes place at the given site, when the contemporary sequence directly descends from the reference sequence—this second term is the usual form of the equation for longer time periods. Thus, the theorem adds the correction term $h_0 e^{-\frac{r}{2} \mu t / (r - 1)}$ when the sequences being compared are not necessarily direct descendants. This accounts for the variation that is present in the reference population from past mutations that have not yet been either fixed or lost. The two expressions $P(t)$ and the Jukes–Cantor mutation probability are asymptotic, in the sense that

$$\frac{P(t)}{1 - e^{-\frac{r}{2} \mu t}} = 1 + \frac{h_0 e^{-\frac{r}{2} \mu t}}{1 - e^{-\frac{r}{2} \mu t}} \to 1$$
as $t \to \infty$.

Note that the theorem requires that the population at the earlier time be used as the reference population. In particular, if populations at several different times are to be compared, then this should be done by comparing each population against the oldest population, to ensure that all pairwise comparisons made involve the same value of the initial heterozygosity $h_0$. This requirement may be relaxed if there is reason to believe that the heterozygosity $h$ is constant, or nearly so. In general, we expect the distribution $\pi$ of states at a given site (and hence also $h$) to depend on $t$, but as $t \to \infty$ we also expect $\pi$ to approach an equilibrium distribution $\pi_{eq}$, where the rate at which new mutations are introduced balances the rate at which mutations are fixed or lost by the mating
process. If this equilibrium is assumed to have occurred then equation (1) may be written in the form

\[ P(t) = h_\infty e^{-\gamma t} + \frac{r - 1}{r} (1 - e^{-\gamma t}) , \]  

(2)

where \( h_\infty \) is the expected heterozygosity at equilibrium. Under this assumption, any two populations at times \( t_1 \) and \( t_2 \) may be compared, using \( t = |t_1 - t_2| \).

**Recovering the Mutation Rate from Observed Data**

We now demonstrate how the mutation rate \( \mu \) may be estimated from observations of \( P(t) \). If the expected heterozygosity is assumed to have reached equilibrium, then \( h_\infty \) may be substituted for \( h_0 \) throughout.

Equation (1) may be rearranged to the form

\[ P(t) = \frac{r - 1}{r} \left( 1 - \left( 1 - \frac{r}{r - 1} h_0 \right) e^{-\gamma t} \right) . \]  

(3)

If \( h_0 \) is known then we may estimate \( \mu \) as

\[ \mu = - \frac{r - 1}{rt} \log \left( \frac{r - 1 - h_0}{r - 1} P(t) \right) . \]  

(4)

On the other hand, if \( h_0 \) is not known then it will need to be estimated simultaneously with \( \mu \). This may be done using a least squares fit to

\[ \log \left( \frac{r - 1}{r - 1} P(t) \right) = \log(1 - \frac{r}{r - 1} h_0) - \frac{r - 1}{r} \mu t, \]

if the least squares line is \( y = mt + b \), then we recover \( \mu \) and \( h_0 \) as

\[ \mu = - \frac{r - 1}{r} m, \quad h_0 = \frac{r - 1}{r} (1 - e^b). \]

**The Divergent Rate Curve**

A divergent rate curve arises when we either omit or use an incorrect value for \( h_0 \) to estimate \( \mu \). In particular, if we ignore the existing diversity and thus use \( h_0 = 0 \) we get the estimate

\[ \mu_{\text{div}} = - \frac{(r - 1) \log(1 - \frac{r}{r - 1} P(t))}{rt}, \]

\[ = \mu - \frac{(r - 1) \log(1 - \frac{r}{r - 1} h_0)}{rt}, \]  

(5)

which adds the divergent term \( \frac{(1 - \gamma) \log(1 - \frac{r}{r - 1} h_0)}{rt} \) to the estimate earlier. Thus, if the variation present within the reference population is ignored, we obtain a divergent rate curve of the form \( \mu_{\text{div}} = \mu + C/t \), where \( C = \frac{(r - 1)}{rt} \log(1 - \frac{r}{r - 1} h_0) \) is a positive constant independent of \( t \). This rate estimate tends to infinity as \( t \to 0 \) and thus becomes increasingly inaccurate on shorter and shorter time scales.

**Results and Discussion**

Figure 1A shows results of simulations over microevolutionary time scales and exhibits the phenomena described above. Most importantly, if we do not correct for existing genetic diversity, there is the apparent acceleration at shorter times, even though the basic neutral mutation rate \( \mu \) is kept constant. This first main point then is that the apparent increase in "rate" (the divergent rate curve) is obtained at shorter periods. However, it is then important that, using equation (4) and the value of \( h_0 \), the value of \( \mu \) can still be estimated accurately at these shorter intervals. This is certainly as expected; it has always been assumed that the mutation rate \( \mu \) was basically constant, that it was independent of time. Thus, the conclusion from figure 1A is that by explicitly considering genetic variability in the ancestral (reference) population—either by using an a priori estimate of it (blue curve) or by estimating it simultaneously using least squares (dashed black line)—it is possible to recover the mutation rate \( \mu \). In practice, it may be that knowledge of \( \mu \) (from longer term studies) may also be important in understanding population structure and its change over time.

**The Simulations**

The simulations were run using the statistical package R (R Development Core Team, 2011). Each simulation begins with a haploid population containing 70 individuals having allele A and 30 individuals having allele B, giving a true \( h_0 \) of 0.42. There are four allele types in total, and mutations from any given allele to any of the other three take place at equal rates, as in the Jukes–Cantor model for DNA substitutions. This population is then evolved for 1,000 generations. In each generation, reproduction is simulated according to the Wright-Fisher process, whereby each of the 100 individuals making up the new population is chosen randomly with replacement from the previous generation; each of these new individuals is then subjected to mutation at the rate of \( \mu = 0.001 \) mutations per individual per generation. For the final step in each generation, a pair of individuals—one from the current population and one from the initial population—is picked randomly and compared; if they have different alleles, a counter for that generation is increased by 1. The entire simulation was repeated 2,000 times, with counters accumulating across runs; consequently the final counter value for generation \( i \), after dividing by 2,000, is an estimate of the probability that an individual picked randomly from the initial population differs from an individual picked randomly from generation \( i \). These relative frequencies are used to estimate the mutation rate parameter \( \mu \) in several different ways, as we now describe.

**Estimating \( \mu \)**

Three different approaches to estimating \( \mu \) are shown in figure 1A. The dashed pink curve shows the result of estimating \( \mu \), while incorrectly assuming \( h_0 = 0 \) (the assumption made by most previous studies); the solid blue curve shows the result of estimating \( \mu \) assuming \( h_0 \) to be its true value, 0.42. The difference is plain, particularly at early times when the probability of observing a difference is dominated by the heterozygosity of the initial population. To demonstrate the limited accuracy achievable using the incorrect assumption that \( h_0 = 0 \), the \( \mu \) curve that would be estimated from an infinite number of observations (instead of 2,000) is shown as a dashed green line—it is scarcely better, indicating that
sampling error is not the problem. The fact that the pink curve tracks the dashed green curve also indicates that, as expected, the simulation fits our theoretical divergent curve \( \mu_{dp} \), equation (5). Note that the erratic behavior of the blue curve near 0 is due to numerical instability, because the calculations there involve logs and ratios of numbers close to zero.

Both of the above estimation procedures can estimate \( \mu \) using a probability estimate from a single time point but assume \( h_0 \) to be known a priori. Where this is not the case, \( h_0 \) can be estimated simultaneously with \( \mu \) using least squares, which requires probability estimates from multiple time points. The result of estimating a single value for \( \mu \) using all 1,001 time points is shown as a horizontal dashed black line in figure 1A.

**Sparingly sampled data**

To test our work on more realistic sparsely sampled data, the mutation rate estimates of figure 1A were repeated using 20 randomly chosen time points from data generated from only
200 repetitions of the simulation. The results are plotted in figure 1B; note that two of the magenta points at the left hand end, and one of the blue points, do not appear in the plot because they are off the scale. The data are noisy due to the smaller number of repetitions, but in all other respects, the same phenomena are apparent as in figure 1A.

**Other Population Structures**

Our corrected formula extends to cases where the population consists of multiple subpopulations that interact by arbitrary migration rates, provided that the initial distribution of states is the same for each subpopulation. Figure 1C shows the accurate recovery of \( p \) and \( h_0 \) for a simulated population consisting of two islands, each containing 70 individuals having allele A and 30 individuals having allele B, which mix sporadically and asymmetrically: in each generation, the probability that an individual in subpopulation 1 comes from subpopulation 2 is just 1%, whereas the probability that an individual in subpopulation 2 comes from subpopulation 1 is 10%. Adding this limited kind of population structure does not alter the fact that the probability distribution of states for the ancestor of an individual chosen randomly at time \( t \) remains equal to that for an individual chosen randomly from the initial population, so the probability \( p_{\text{diff}} \) that the ancestral and reference states differ is still given by \( h_0 \).

If the initial state distributions differ across subpopulations, then the probability that the ancestor of a randomly chosen contemporary individual has a particular state is no longer constant, so \( h_0 \) in equation (1) must be replaced with a function of time, \( p_{\text{diff}}(t) \). Figure 1D shows results from a simulation of one such scenario. There are two islands, each containing 100 individuals: the first contains 70 individuals having allele A and 30 individuals having allele B, and the second contains 25 individuals with each of the four alleles. Migration rates are as for figure 1C. The blue curve, which shows an attempt to estimate \( p \) using equation (4) and assuming that \( h_0 = 0.6675 \) (the probability that two individuals chosen randomly from the initial population have different states), produces a divergent rate curve because this more general population structure violates the assumption that \( p_{\text{diff}} \) is constant and equal to \( h_0 \).

Despite the fact that \( p_{\text{diff}} \) is nonconstant when the subpopulations have different initial state distributions, our approach is still useful here. Under reasonable conditions on migration probabilities, the probability that the ancestor of a randomly chosen individual has a given state converges toward an equilibrium value as \( t \to \infty \), and so \( p_{\text{diff}} \) will approach an equilibrium value also. In particular, when estimating \( p \) and \( h_0 \) simultaneously via least squares, only the estimate of \( h_0 \) is affected by different initial state distributions across subpopulations; as expected, the dashed black line depicting the least squares estimate of \( p \) in figure 1D is close to the true value of \( p \). As \( p \) is usually the parameter of interest, this means that our model can still be used for inference with these more general population structures. Note that any edge-weighted graph describing mating probabilities within a population can be represented as a multi-island model in which each island contains a single individual.

**The Divergent Rate Curve and Real Data**

Under the simple Jukes–Cantor model considered here, we have established that rate estimates that do not properly account for ancestral polymorphism will overestimate the rate, leading to a divergent rate curve of the form \( \hat{\mu} = \mu + C/t \). We do not typically expect real data to fit the assumptions of the Jukes–Cantor model, but the heuristic argument of the introduction suggests that even so we might still expect unaccounted ancestral polymorphism to contribute a term of the form \( C/t \) to rate estimates. To test this, we used orthogonal least squares to fit curves of the form \( y = \mu + C/t \) to data sets from García-Moreno (2004, fig. 1) and Henn et al. (2009, table 2/fig. 2).

Orthogonal least squares minimizes the sum of the squared distances from each data point to the approximating curve. This measure of fit may be appropriate when the observed values of both the dependent and independent variables are subject to error, as is the case here. In contrast, ordinary least squares minimizes the sum of the squared vertical distances and so treats the observed values of the independent variable as being known exactly. For the problem at hand, when \( t \) is close to 0, this will lead to small errors in \( t \) being treated as very large errors in \( y \), with the result that such points will have a disproportionate influence on the fitted curve. This makes orthogonal least squares a better choice for the problem under consideration.

The curves of best fit were calculated using MATLAB, and the results can be seen in figures 2 and 3, respectively. In both cases, we see good qualitative agreement between the data points and the fitted curve.

**Conclusion**

The Jukes–Cantor model is one of the oldest and simplest substitution models, and the benefit of studying ancestral
polymorphism in this simple setting is that we are able to obtain explicit analytic expressions for quantities of interest, including the expected divergence between two sequences, and the uncorrected rate curve $\mu_{uc}$. Moreover, many more complex and realistic models include the Jukes–Cantor model as a special case, so although our precise quantitative findings are only supported under the limited model considered here, there are clear qualitative implications for these more general models.

Given the rapid advance in DNA sequencing technology, we expect a large increase in sequences that have diverged in recent, intermediate, and longer times. It is therefore essential to be able to generalize results to a full range of divergence times—it is no longer appropriate to consider separately shorter term microevolutionary and longer term macroevolutionary studies. Our results show that a significant contributing factor to the apparent acceleration at shorter times is the pre-existing diversity within a population, and this can be estimated for a wide range of population sizes and structures (Charlesworth and Charlesworth, 2010). By considering the expected diversity in a population, we show how the real rate of neutral mutation can be estimated at shorter times, using either an appropriate estimate of genetic diversity or data from multiple time points. As the time scales of traditional phylogenetics and population genetics draw closer together, we anticipate seeing an increasing emphasis on the careful handling of genetic diversity in phylogenetic analyses; the work presented here can provide a starting point for the exploration of more general models involving more complex substitution processes, time-varying population sizes, and other effects.

Appendix: Proof of the Main Result

The individual chosen from the contemporary population descends from a unique member of the population reference, and this ancestor is equally likely to be any member of the reference population. We may therefore calculate the probability that the contemporary and reference states differ by considering two cases: the ancestral state agrees with the reference state but evolves to disagree; and the ancestral state disagrees with the reference state and evolves so as to remain in disagreement at time $t$.

Under the $r$-state Jukes–Cantor model, the probability that there is a net transition from state $i$ to state $j \neq i$ in time $t$ is given by

$$p(t) = \frac{1}{r} \left(1 - e^{-\frac{r}{C_{25}} t}\right).$$

The probability that the ancestral state agrees with the reference state but evolves to disagree is therefore given by

$$\sum_i \pi_i^2 (r-1)p(t) = (r-1)p(t) \sum_i \pi_i^2,$$

whereas the probability that both the ancestral and contemporary states differ from the reference is given by

$$\sum_i \sum_{j \neq i} \pi_i \pi_j (1-p(t)) = (1-p(t)) h_0. \tag{7}$$

Observe that

$$1 = \left( \sum_i \pi_i \right)^2 = \sum_i \pi_i^2 + \sum_i \sum_{j \neq i} \pi_i \pi_j = h_0 + \sum_i \pi_i^2,$$

so $\sum_i \pi_i^2 = 1 - h_0$. Hence summing equations (6) and (7) we get

$$p(t) = (1 - h_0)(r-1)p(t) + h_0(1-p(t))$$

$$= (r-1)p(t) + h_0(1-p(t)) - (r-1)p(t)$$

$$= \frac{r-1}{r} (1 - e^{-\frac{r}{C_{25}} t}) + h_0 e^{-\frac{r}{C_{25}} t},$$

as claimed.

Acknowledgments

Figure 2 is adapted from Figure 1 of García-Moreno (2004), which is copyright the Journal of Avian Biology, published by John Wiley and Sons. Figure 3 is adapted from Figure 2 of Henn et al. (2009), and is reused by permission of Oxford University Press, on behalf of the Society for Molecular Biology and Evolution. The authors thank Jaime García-Moreno and Brenna Henn for permission to reuse these figures, and Jaime García-Moreno for sending us the data plotted in figure 2.

References

Charlesworth B, Charlesworth D. 2010. Elements of evolutionary genetics. Greenwood Village (CO): Roberts and Co.

Charlesworth D. 2010. Don’t forget the ancestral polymorphisms. Heredity 105:509–510.

Fitch WM, Atchley WR. 1985. Evolution in inbred strains of mice appears rapid. Science 228:1169–1175.

García-Moreno J. 2004. Is there a universal mtDNA clock for birds? J Avian Biol. 35:465–468.

Gnener MJ, Seehausen O, Lunt DH, Joyce DA, Shaw PW, Carvalho GR, Turner GF. 2007. Age of cichlids: new dates for ancient lake fish radiations. Mol Biol Evol. 24:1269–1282.
Henn BM, Gignoux CR, Feldman MW, Mountain JL. 2009. Characterizing the time dependency of human mitochondrial DNA mutation rate estimates. Mol Biol Evol. 26:217–230.

Ho SYW, Lanfear R, Bromham L, Phillips MJ, Soubrier J, Rodrigo AG, Cooper A. 2011. Time-dependent rates of molecular evolution. Mol Ecol. 20:3087–3101.

Ho SYW, Phillips MJ, Cooper A, Drummond AJ. 2005. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. Mol Biol Evol. 22:1561–1568.

Jukes TH, Cantor CR. 1969. Evolution of protein molecules. In: Munro HN, Allison JB, editors. Mammalian protein metabolism. New York: Academic Press. p. 21–123.

Kimura M. 1968. Evolutionary rate at the molecular level. Nature 217: 624–626.

Kimura M. 1983. The neutral theory of molecular evolution. Cambridge (UK): Cambridge University Press.

Millar CD, Dodd A, Anderson J, Gibb GC, Ritchie PA, Baroni C, Woodhams MD, Hendy MD, Lambert DM. 2008. Mutation and evolutionary rates in adélie penguins from the antarctic. PLoS Genetics 4(10):e1000209. doi:10.1371/journal.pgen.1000209.

Penny D. 2005. Evolutionary biology: relativity for molecular clocks. Nature 436:183–184.

Peterson GI, Masel J. 2009. Quantitative prediction of molecular clock and $K_a/K_s$ at short timescales. Mol Biol Evol. 26:2595–2603.

R Development Core Team. 2011. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing.

Wright S. 1931. Evolution in Mendelian populations. Genetics 16:97–159.