

The Molecular Clock and the Relationship between Population Size and Generation Time



Lin Chao; David E. Carr

Evolution, Vol. 47, No. 2 (Apr., 1993), 688-690.

Stable URL:

<http://links.jstor.org/sici?sici=0014-3820%28199304%2947%3A2%3C688%3ATMCA%3E2.0.CO%3B2-%23>

Evolution is currently published by Society for the Study of Evolution.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/ssevol.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

THE MOLECULAR CLOCK AND THE RELATIONSHIP BETWEEN POPULATION SIZE AND GENERATION TIME

LIN CHAO¹ AND DAVID E. CARR²

¹Department of Zoology and ²Department of Botany, University of Maryland, College Park, MD 20742 USA

Key words.—Generation time, molecular clock, population size.

Received January 21, 1992. Accepted August 14, 1992.

A rule of thumb in population genetics is that population size (*N*) and generation time (*g*) across taxa are inversely related. This relationship played a prominent role in the development of the neutral theory of molecular evolution because it was used by Ohta to explain why a molecular clock is compatible with the theory (Ohta, 1977, 1987; Kimura, 1979, 1983; Ohta and Tachida, 1990). However, such a relationship has never been documented. Nei and Graur (1984) failed to find a significant correlation in the only known attempt to test the relationship. They did not, however, test for a correlation between log(*N*) and log(*g*), which is the relationship required to produce a molecular clock in Ohta's original model (Ohta, 1977; Kimura, 1979, 1983). We have reanalyzed Nei and Graur's data set and report a significant negative correlation between log(*N*) and log(*g*).

deleterious" mutations, a hypothesis that Kimura (1979, 1983) embraced and extended. Ohta and Kimura showed that if *u* is instead the rate of slightly deleterious mutations per generation and the selection coefficients (*s*) of the mutations are distributed as a gamma function, mutations are fixed at rate per generation of

$$k_g \propto \frac{u_g}{N^\beta}, \tag{2}$$

where the constant β is a parameter of the function. Populations with small *N* have a higher k_g because they fix more mutations through genetic drift. Ohta's model fits the neutral theory because it treats a deleterious mutation as "effectively" neutral when $s \leq 1/2N$.

To make a molecular clock, Ohta and Kimura assumed u_g to be constant across taxa and considered *g* and *N* to be inversely related across the same taxa as $g = C/N^\alpha$, where α and *C* are constants. Mutations are then fixed at a rate per year of

$$k_y = \frac{k_g}{g} \propto \frac{u_y}{CN^{\beta-\alpha}}. \tag{3}$$

If $\beta - \alpha$ is small, k_y is approximately constant per year and a clock emerges even if *g* varies across taxa.

Kimura (1979, 1983) interpreted β to be an inverse measure of the degree of physiological homeostasis. With strong homeostasis, $\beta = 0$ and mutations are neutral. Ohta assumed $\beta = 1$, which transforms the gamma distribution into an exponential one. Because too few mutations are effectively neutral with an exponential distribution when *N* is large, Kimura believed that $\beta = 0.5$ was more reasonable. To make k_y constant, both Ohta and Kimura anticipated that $\alpha = \beta$ and predicted, respectively, that α should be about 1 and 0.5.

Kimura (1987, 1991) recently returned to emphasizing his original neutral model (eq. 1) and the absolute-time constancy of *u*. Ohta (1987; Ohta and Tachida, 1990) retained her slightly deleterious model, but modified it in response to criticisms by incorporating slightly advantageous mutations into a "nearly neutral" model. Her new model still predicts an inverse relationship between k_g and *N* (cf eq. 2), but it lacks an analytical solution and the exact relationship is not known.

As noted by Gillespie (1992), the absolute-time dependency of the molecular clock was not predicted by the neutral theory. Instead, Kimura and Ohta modified the theory to fit the observations. Thus, the inverse relationship between *g* and *N* is critical for the present

THE MOLECULAR CLOCK AND THE NEUTRAL THEORY

The rate of amino acid or nucleotide substitution in different taxa provides a measure of the time the taxa diverged. A property of this molecular meter is its approximate dependency on absolute and not generation time (Zuckerkindl and Pauling, 1965; Zuckerkindl, 1987). This dependency is the empirical basis of the molecular clock.

The existence of a molecular clock has been difficult to explain. A clock driven by selection for beneficial mutations requires that the product of many population parameters remains constant across taxa (Kimura, 1983, 1987). Kimura's (1983) neutral theory of evolution provides a simpler explanation, but it has its own difficulties. As originally conceived (Kimura, 1968), the neutral theory proposes that a diploid population of size *N* with a neutral mutation rate *u* produces mutations at a rate $2Nu$. The probability that a neutral mutation is eventually fixed is $1/2N$. Thus, the substitution rate is

$$k = 2Nu \frac{1}{2N} = u. \tag{1}$$

But, if *u* is per generation, so is *k*, and there should not be a molecular clock across taxa with differing values of *g*, even if *u* is constant.

Kimura and Ohta first attempted to explain this discrepancy between the neutral theory and a molecular clock by assuming that *u* across taxa was approximately constant per year and not generation (Kimura, 1983). Later, Ohta (1977) introduced her model of "slightly

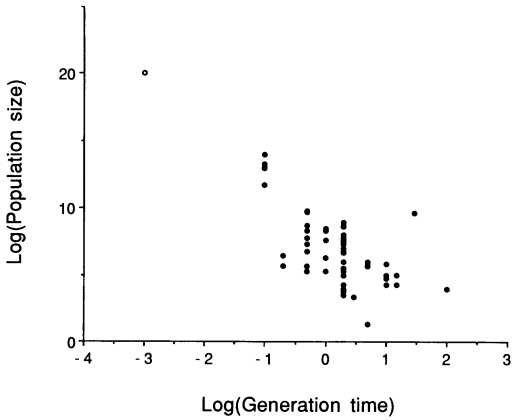


FIG. 1. The relationship between population size and generation time (years). The outlier denoted by an open circle is *E. coli*. Logarithms are base 10.

conception of the theory. To Gillespie, the absence of any correlation between g and N would be "a serious blow to the theory and could quite reasonably be used to reject it altogether."

POPULATION SIZE AND GENERATION TIME

Important as the inverse relationship between N and g is to the neutral theory, it has not been critically examined. Ohta and Kimura (Ohta, 1977, 1987; Ohta and Tachida, 1990; Kimura, 1979, 1983) accepted it as biologically reasonable, but pointed to the need for data. The only attempt at a test was by Nei and Graur (1984), who collected information on species from a variety of higher taxa and failed to find any significant correlation in their data. However, Nei and Graur tested only the relationship between generation time and N , $\log(N)$ and \sqrt{N} . They did not test for a correlation between $\log(N)$ and $\log(g)$, which may have been more appropriate given the assumption of $g = C/N^\alpha$ in Ohta's slightly deleterious models.

We have thus reexamined Nei and Graur's data (Fig. 1; Table 1). As they found, the Pearson's product-moment correlation coefficient (Sokal and Rohlf, 1981) between N and g in their sample of 77 species was not significant. However, the correlation between $\log(N)$ and $\log(g)$ was negative and highly significant, and it remained so even when a prominent outlier, *Escherichia coli*, was removed from the analysis.

We considered also whether the observed correlation could have been inflated by pseudoreplication (Harvey and Pagel, 1991). Nearly 50% of the species in the data set were from three genera, and species within a higher taxon may not behave as independent samples. To minimize possible pseudoreplication, we examined the correlation across genera and families in the data. A concern that arose was how to represent the mean N of species within a genus or family. If the correlation between $\log(N)$ and $\log(g)$ is the same across species within a higher taxon and across the higher taxa, then the geometric mean is the best representation. If variation in N among species within a higher taxon reflects the history of any one species, then all species will eventually be bottlenecked to the smallest N and the

TABLE 1. Relationship between population size and generation time across species, genera and families. r is the Pearson product-moment correlation coefficient. α was estimated by a model I regression of $\log(N)$ on $\log(g)$. The variable with the least error variance relative to the variation among taxa was treated as the independent variable. Because $\log(g) - \alpha \log(N) = \log(C)$, the regression estimates $1/\alpha$ as the slope. When species were grouped by genera or families, g for the species was averaged as the geometric mean and N was either averaged as the geometric and harmonic mean or represented by the largest N . Genera and families were determined according to Nowak and Paradiso (1983) and Nelson (1984). Unless noted otherwise, all values and averages for N and g were log-transformed before analysis.

Across	r	α
Species ($n = 77$)		
All species		
N and $g\#$	-0.039 NS	
$\log(N)$ and $\log(g)$	-0.713***	0.33
Without <i>E. coli</i>	-0.609***	0.38
Genera ($n = 32$)		
Geometric mean	-0.748***	0.35
Harmonic mean	-0.692***	0.39
Maximum value	-0.750***	0.33
Families ($n = 23$)		
Geometric mean	-0.780***	0.35
Harmonic mean	-0.713***	0.39
Maximum value	-0.783***	0.33

Not log-transformed; *** $P < 0.001$; NS, not significant for hypothesis $r = 0$.

harmonic mean is better. Finally, if species with small N go extinct more frequently and most new species are derived from species with large N , then the largest N , and not any mean, is more reasonable.

Because our current understanding of phyletic biology cannot discriminate among the above three alternatives for representing N within a higher taxon, we used them all. We found that regardless of our choice, the correlation between $\log(N)$ and $\log(g)$ across genera and families was highly significant (Table 1). Considering the uncertainties of estimating N , it is remarkable that our correlations were as robust as they were. Through a regression of $\log(N)$ on $\log(g)$ (Sokal and Rohlf, 1981), we also estimated α to be closer to Kimura's value of 0.5 (Table 1). Thus, Ohta's assumption of an inverse relationship between $\log(N)$ and $\log(g)$ for her slightly deleterious model is satisfied.

DISCUSSION

Recent findings show that the rate of substitution is less clock-like and more negatively correlated with generation time for synonymous mutations than for non-synonymous mutations (Li et al., 1985, 1987; Gillespie, 1989). Additionally, the best molecular clocks are based on substitution rates for second codon positions and amino acids (Zuckerkandl, 1987). As Li et al. (1987) point out, these findings support both Kimura's original neutral model and Ohta's explanation for the mo-

lecular clock. Ironically, the support for the neutral model requires that u is constant per generation and not per year. If so, synonymous mutations exhibit a generation time effect because they are more neutral and are fixed at a rate equal to eq. 1. On the other hand, because non-synonymous mutations are more subject to purifying selection, their substitution rate follows Ohta's model and is thus more clock-like.

These results also argue against a germline effect as the reason for a molecular clock. The germline argument is that although generation time for rodents is about 100 times shorter than that in humans, the number of germline DNA replications per year is only 10 times higher than that in humans (Li and Tanimura, 1987). As a result, a molecular clock exists because germline replication rates are more similar per year. However, the observation that synonymous substitution rates show a generation time effect argues that a germline effect does not provide a sufficient correction. Furthermore, a germline effect predicts that synonymous and non-synonymous substitution rates should be equally prone (or immune) to a generation time effect and the observation is that non-synonymous substitution rates are more clock-like than synonymous rates.

Although Ohta's model is the only one that can account for the more clock-like behavior of non-synonymous rates, support for her model was until now weakened by the absence of data for an inverse relationship between N and g . The demonstration of an inverse relationship between $\log(N)$ and $\log(g)$ shows that her slightly deleterious model can produce a reasonable molecular clock. Because her nearly neutral model does not specify an exact relationship between k_g and N , it is not clear how constant a clock is produced by the model if $\log(N)$ and $\log(g)$ are correlated. Most likely, any inverse relationship between N and g makes the substitution rate more clock-like. We hope that this study will encourage the collection of additional data for examining the generality of these results.

ACKNOWLEDGMENTS

We thank D. Gill, W. Stephan, W.-H. Li, J. Gillespie, and T. Ohta for helpful comments. This work was supported in part by the Computer Science Center at UMD, Maryland Agricultural Experiment Station, and by an award from the National Institutes of Health (GMS) to L.C.

LITERATURE CITED

- GILLESPIE, J. H. 1989. Lineage effects and the index of dispersion of molecular evolution. *Mol. Biol. Evol.* 6:636-647.
- . 1992. *The Causes of Molecular Evolution*. Oxford University Press, N.Y., USA.
- HARVEY, P. H., AND M. D. PAGEL. 1991. *The Comparative Method in Evolutionary Biology*. Oxford University Press, N.Y., USA.
- KIMURA, M. 1968. Evolutionary rate at the molecular level. *Nature* 217:624-626.
- . 1979. Model of effectively neutral mutations in which selective constraint is incorporated. *Proc. Natl. Acad. Sci. USA* 76:3440-3444.
- . 1983. *The Neutral Theory of Molecular Evolution*. Cambridge University Press, UK.
- . 1987. Molecular evolutionary clock and the neutral theory. *J. Mol. Evol.* 26:24-33.
- . 1991. Recent development of the neutral theory viewed from the Wrightian tradition of theoretical population genetics. *Proc. Natl. Acad. Sci. USA* 88:5969-5973.
- LI, W.-H., C.-C. LUO, AND C.-I. WU. 1985. Evolution of DNA sequences, pp. 1-94. *In* R. J. MacIntyre (ed.), *Molecular Evolutionary Genetics*. Plenum, N.Y., USA.
- LI, W.-H., AND M. TANIMURA. 1987. The molecular clock runs more slowly in man than in apes and monkeys. *Nature* 326:93-96.
- LI, W.-H., M. TANIMURA, AND P. M. SHARP. 1987. An evaluation of the molecular clock hypothesis using mammalian DNA sequences. *J. Mol. Evol.* 25:330-342.
- NEI, M., AND D. GRAUR. 1984. Extent of protein polymorphism and the neutral mutation theory. *Evol. Biol.* 17:73-118.
- NELSON, J. S. 1984. *Fishes of the World*. John Wiley and Sons, N.Y., USA.
- NOWAK, R. M., AND J. L. PARADISO. 1983. *Walker's Mammals of the World*. Johns Hopkins University Press, Baltimore, MD USA.
- OHTA, T. 1977. Extension to the neutral mutation random drift hypothesis, pp. 148-176. *In* M. Kimura (ed.), *Molecular Evolution and Polymorphism*. National Institute of Genetics, Mishima, Japan.
- . 1987. Very slightly deleterious mutations and the molecular clock. *J. Mol. Evol.* 26:1-6.
- OHTA, T., AND H. TACHIDA. 1990. Theoretical study of near neutrality. I. Heterozygosity and rate of mutant substitution. *Genetics* 126:219-229.
- SOKAL, R. R., AND R. J. ROHLF. 1981. *Biometry*. Freeman, San Francisco, CA USA.
- ZUCKERKANDL, E. 1987. On the molecular clock. *J. Mol. Evol.* 26:34-46.
- ZUCKERANDL, E., AND L. PAULING. 1965. Evolutionary divergence and convergence in proteins, pp. 97-165. *In* V. Bryson and H. J. Vogel (eds.), *Evolving Genes and Proteins*. Academic Press, N.Y., USA.

Corresponding Editor: R. Harrison