8. Shinomura, Y., Eng, J., Rattan, S. C. & Yalow, R. S. *Comp. Biochem. Physiol.* **96B**, 239-242 (1990).
9. Beintema, J. J. & Neuteboom, B. *J. molec. Evol.* **19**, 145-152 (1983).
10. Sarkar, G., Koeberl, D. D. & Sommer, S. S. *Genomics* **6**, 133-143 (1990).
11. Fan, Z.-W., Eng, J., Shaw, G. & Yalow, R. S. *Peptides* **9**, 429-431 (1988).
12. Smith, A. I. *et al. J. Endocr.* **115**, R5-R8 (1987).
13. Wolfe, P. B. & Cebra,, J. J. *Molec. Immun.* **17**, 1493-1505 (1980).
14. Eng, J., Du, B.-H., Raufman, J.-P. & Yalow, R. S. *Peptides* **7**, Suppl. 1, 17-20 (1986).
15. Easteal, S. *Genetics* **124**, 165-173 (1990).
16. Li, W.-H., Gouy, M., Sharp, P. M., O'Huigin, C. & Yang, Y.-W. *Proc. natn. Acad. Sci. U.S.A.* **87**, 6703-6707 (1990).
17. Fitch, W. M. *Am. Nat.* **16**, 111-120 (1980).
18. Felsenstein, J. *Syst. Zool.* **34**, 152-161 (1985).
19. Li, W.-H. & Gouy, M. *Meth. Enzym.* **183**, 645-659 (1990).
20. Fitch, W. M. & Beintema, J. J. *Molec. Biol. Evol.* **7**, 438-443 (1990).
21. Sarich, V. M. & Wilson, A. C. *Science* **158**, 1200-1203 (1973).
22. Shoshani, J., Goodman, M., Czelusniak, J. & Braunitzer, G. in *Evolutionary Relationships among Rodents: A Multidisciplinary Analysis* (eds Luckett, W. P. & Hartenberger, J.-L.) 191-210 (Plenum, New York, 1985).
23. Yu, J.-H., Eng, J., Rattan, S. & Yalow, R. S. *Peptides* **10**, 1195-1197 (1989).
24. Li, W.-H., Wu, C.-I. & Luo, C.-C. *Molec. Biol. Evol.* **2**, 150-174 (1985).
25. Felsenstein, J. *Syst. Zool.* **27**, 401-410 (1978).

ACKNOWLEDGEMENTS. We thank M. Gouy, R. Honeycutt, H. Mendelssohn, T. Mendelson, D. Eilam, Z. Sever, and L. Jacobs for help and suggestions and J. Felsenstein for sending us the PHYLIP package. This study was supported by NIH (W.-H.L.) and the US-Israel Binational Science Foundation (D.G.).

Adaptive protein evolution at the *Adh* locus in *Drosophila*

John H. McDonald & Martin Kreitman

Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey 08544, USA

PROTEINS often differ in amino-acid sequence across species. This difference has evolved by the accumulation of neutral mutations by random drift, the fixation of adaptive mutations by selection, or a mixture of the two. Here we propose a simple statistical test of the neutral protein evolution hypothesis based on a comparison of the number of amino-acid replacement substitutions to synonymous substitutions in the coding region of a locus. If the observed substitutions are neutral, the ratio of replacement to synonymous fixed differences between species should be the same as the ratio of replacement to synonymous polymorphisms within species. DNA sequence data on the *Adh* locus (encoding alcohol dehydrogenase, EC 1.1.1.1) in three species in the *Drosophila melanogaster* species subgroup do not fit this expectation; instead, there are more fixed replacement differences between species than expected. We suggest that these excess replacement substitutions result from adaptive fixation of selectively advantageous mutations.

Consider a set of alleles from more than one species and assume that there is no recombination. The alleles are then connected by a single phylogenetic tree, which can be divided into two parts: between-species branches and within-species branches. Within-species branches connect all the alleles within each species to their most recent common ancestor. Between-species branches connect these common ancestors to the common ancestor of the whole phylogeny. A mutation on a between-species branch will appear in all the descendant alleles and thus will be a fixed difference between species, whereas a mutation on a within-species branch will be a polymorphism within a species.

Nucleotide substitutions in a coding region can also be divided into replacement substitutions and synonymous substitutions. Of M possible mutations in a coding region, let M_r be the number of possible neutral replacement mutations and M_s be the number of possible neutral synonymous mutations. All remaining mutations, $M-(M_r+M_s)$, are deleterious. Let μ be the nucleotide mutation rate per nucleotide site, so that the mutation rate for any one of the three possible mutations at a site is $\mu/3$. Under the neutral theory, the expected number of fixed replacement substitutions in a set of alleles is $T_b(\mu/3)M_r$, where T_b is the total time on the between-species branches. The expected number of fixed synonymous substitutions is $T_b(\mu/3)M_s$. For a particular phylogeny and mutation rate, the number of replacement substitutions is independent of the number of synonymous substitutions. Therefore, the expected ratio of replacement to synonymous fixed substitutions is $T_b(\mu/3)M_r : T_b(\mu/3)M_s$, which reduces to $M_r : M_s$. If T_w is the total time on the within-species branches of the phylogeny, the expected ratio of replacement to synonymous polymorphisms is $T_w(\mu/3)M_r : T_w(\mu/3)M_s$, which also reduces to $M_r : M_s$. Thus, if protein evolution occurs by neutral processes, the ratio of replacement to synonymous fixed substitutions should be the same as the ratio of replacement to synonymous polymorphisms. A G-test of independence[1] can be used to test this null hypothesis.

This test does not require many assumptions. Unlike most tests of the neutral theory, it does not require that populations have reached equilibrium. Recombination could cause different segments of a locus to have different phylogenies, but this would only cause a serious bias in the above test if the time to a common ancestor of a genomic segment was correlated with the ratio of neutral replacement to synonymous sites in that segment. There is no reason to expect such a correlation. Similarly, variation in nucleotide mutation rate among regions of a locus would only be important if nucleotide mutation rate was correlated with the ratio of neutral replacement to synonymous sites, which does not seem likely. We have assumed, for simplicity, that all nucleotide mutations are equally likely. A difference in mutation rate between transitions and transversions, coupled with the greater proportion of transitions that are synonymous, would change the ratio of replacement to synonymous substitutions. But the ratio for fixed substitutions would still equal the ratio for polymorphisms. If there are more synonymous sites than replacement sites with two or three possible neutral mutations, multiple substitutions at a single nucleotide site will make the ratio of replacement to synonymous fixed differences a slight overestimate of $M_r : M_s$. For closely related species the effect is trivially small and even in the most extreme case of complete saturation of neutral mutable sites with substitutions, the overestimation from multiple substitutions is not large enough to account for the results presented below.

We have compared DNA sequences of the coding region of the *Adh* locus in three species in the *Drosophila melanogaster* species subgroup (Table 1). There is substantial evidence that the threonine/lysine polymorphism in the alcohol dehydrogenase of *D. melanogaster* is subject to natural selection[2,3], but there has been no evidence to indicate whether any of the replacement differences between species were adaptive. We limited the comparison to closely related species to minimize the number of multiple mutations at single nucleotide sites. The ratio of replacement substitutions to synonymous substitutions that are fixed between species is significantly greater than the ratio of replacement to synonymous polymorphisms (Table 2). About 29% of the fixed differences between species are replacement substitutions, but only 5% of the polymorphisms are replacement substitutions. This is not what would be expected if the observed replacement and synonymous substitutions were selectively neutral. It would be expected, however, if most of the observed replacement substitutions were due to adaptive fixation of selectively advantageous mutations. A substitution which becomes fixed by selection will be a polymorphism for less time than a substitution that becomes fixed by random drift[4] and so an adaptive substitution will be less likely to appear as a polymorphism than will a neutral substitution.

An alternative explanation for the excess number of fixed replacement differences could be a combination of many slightly deleterious replacement mutations and population sizes which have recently undergone dramatic expansion in all three species. In the ancestral small populations, mildly deleterious replacement mutations are effectively neutral and would therefore accumulate as fixed differences. In recently expanded populations, slightly deleterious replacements are selected against and

TABLE 1 Variable nucleotides from the coding region of the *Adh* locus in *D. melanogaster*, *D. simulans* and *D. yakuba*

		D. melanogaster	*D. simulans*	*D. yakuba*		
	Con.	a b c d e f g h i j k l	a b c d e f	a b c d e f g h i j k l		
781	G	T T T T T T T T T T T T	- - - - - -	- - - - - - - - - - - -	Repl.	Fixed
789	T	- - - - - - - - - - - -	- - - - - -	C C C C C C C C C C C C	Syn.	Fixed
808	A	- - - - - - - - - - - -	- - - - - -	G G G G G G G G G G G G	Repl.	Fixed
816	G	T T T T - - - - - - - T	T T T T T T	- - - - - - - - - - - -	Syn.	Poly.
834	T	- - - - - - - - - - - -	C C - - - C	- - - - - - - - - - - -	Syn.	Poly.
859	C	- - - - - - - - - - - -	- - - - - -	G G G G G G G G G G G G	Repl.	Fixed
867	C	- - - - - - - - - - - -	- - - - - -	G G G G G A G G G G G G	Syn.	2 Poly.
870	C	T T T T T T T T T T T T	- - - - - -	- - - - - - - - - - - -	Syn.	Fixed
950	G	- - - - - - - - - - - -	- A - - - -	- - - - - - - - - - - -	Syn.	Poly.
974	G	- - - - - - - - - - - -	T - T T T T	- - - - - - - - - - - -	Syn.	Poly.
983	T	- - - - - - - - - - - -	- - - - - -	C C C C C C C C C C C C	Syn.	Fixed
1019	C	- - - - - - - - - - - -	- - - - - -	- - - - A - - - - - - -	Syn.	Poly.
1031	C	- - - - - - - - - - - -	- - - - - -	- - - - - - - - A - - -	Syn.	Poly.
1034	T	- - - - - - - - - - - -	- - - - - -	- C C C C C - - C - C C	Syn.	Poly.
1043	C	- - - - - - - - - - - -	- - - - - -	- - - - A - - - - - - -	Syn.	Poly.
1068	C	T T - - - - - - - - - -	- - - - - -	- - - - - - - - - - - -	Syn.	Poly.
1089	C	- - - - - - - - - - - -	A A A A A A	- - - - - - - - - - - -	Repl.	Fixed
1101	G	- - - - - - - - - - - -	- - - - - -	A A A A A A A A A A A A	Repl.	Fixed
1127	T	- - - - - - - - - - - -	- - - - - -	C C C C C C C C C C C C	Syn.	Fixed
1131	C	- - - - - - - - - - - -	- - - - - -	- - - - T - - - - - - -	Syn.	Poly.
1160	T	- - - - - - - - - - - -	- - - - - -	C C C C C C C C C C C C	Syn.	Fixed
1175	T	- - - - - - - - - - - -	- - - - - -	C C C C C C C C C C C C	Syn.	Fixed
1178	C	- - - - - - - - - - - -	- - - - - -	- - - - - - - - - - A -	Syn.	Poly.
1184	C	- - - - - - - - - - - -	- - - - - -	G G G G G G G G G G G G	Syn.	Fixed
1190	C	- - - - - - - - - - - -	- - - - - -	- - A - - - - - - - - -	Syn.	Poly.
1196	G	- - - - - - - - - - - -	- - - - - -	T T T - T T T - T - -	Syn.	Poly.
1199	C	- T - - - - - - - - - -	- - - - - -	- - - - - - - - - - - -	Syn.	Poly.
1202	T	- - - - - - - - - - - -	- - - - - -	C C C C C C C C C C C C	Syn.	Fixed
1203	C	- - - - - - - - - - - -	- - T - - -	- - - - - - - - - - - -	Syn.	Poly.
1229	T	- - C C C C C C C C C C	- - - - - -	- - - - - - - - - - - -	Syn.	Poly.
1232	T	- - - - - - - - - - - -	- - - - - -	A A A A A A A A A A A A	Syn.	Fixed
1235	C	- - - - - - - - - - A -	- - - - - -	- - - - - - - - - - - -	Syn.	Poly.
1244	C	- - - - - - - - - - - -	- - - - - -	- - A - - - - - - - - -	Syn.	Poly.
1265	C	- - - - - - - - - - - -	- - - - - -	G G G G G G G G G G G G	Syn.	Fixed
1271	A	- - - - - - - - - - - -	- T - T - -	- - - - - - - - - - - -	Syn.	Poly.
1277	T	- - - - - - - - - - - -	- - - - - -	C C C C C C C C C C C C	Syn.	Fixed
1283	C	A A - - - - - - - - - -	- - - - - -	- - - - - - - - - - - -	Syn.	Poly.
1298	C	- - - - - - - - - - - -	- - - - - -	T T T T T T T T T T T T	Syn.	Fixed
1304	C	- - - - - - - - - - - -	- - - - T -	- - - - - - - - - - - -	Syn.	Poly.
1316	C	- - - - - - - - - - - -	- - T T - -	T T T T T T T T T T T T	Syn.	Poly.
1425	C	A A - - - - - - - - - -	- - - - - -	- - - - - - - - - - - -	Syn.	Poly.
1431	T	C C - - - - - - - - - -	- - - - - -	C C C C C C C C C C C C	Syn.	Poly.
1443	C	- - - - - - G G G G G G	- - - - - -	- - - - - - - - - - - -	Syn.	Poly.
1452	C	- - - - - - T T T T T T T	- - - - - -	- - - - - - - - - - - -	Syn.	Poly.
1490	A	- - - - - - - C C C C C C	- - - - - -	- - - - - - - - - - - -	Repl.	Poly.
1504	C	T T T T T T T T T T T T	- - - - - -	- - - - - - - - - - - -	Syn.	Fixed
1518	C	- - - - - T T T T T T T	- - - - - -	- - - - - - - - - - - -	Syn.	Poly.
1524	T	- - - - - - - - - - - -	- - - - - -	G G G G G G G G G G G G	Syn.	Fixed
1527	C	T T T T T T - - - - - -	- - - - - -	- - - - T - - - - - - -	Syn.	Poly.
1530	G	- - - - - - - - - - - -	- - - - - -	- - - A - - - - - - - -	Syn.	Poly.
1545	T	- - - - - - - - - - - -	- - - - - -	C C C C C C C C C C C C	Syn.	Fixed
1548	C	- - - - - - - - - - - -	- - - - - -	- - - - A - - - - - - -	Syn.	Poly.
1551	C	- - - - - - - - - - - -	- - - T - -	- - - - - - - - - - - -	Syn.	Poly.
1555	C	- - - - - - - - - - - T	- - - - - -	- - - - - - - - - - - -	Repl.	Poly.
1557	C	A A A A A - - - - - - -	- - - - - -	- - - - - - - - - - - -	Syn.	Poly.
1560	G	- - - - - - - - - - - -	- - - A - -	- - - - - - - - - - - -	Syn.	Poly.
1573	G	- - - - - - - - - - - -	- - - - - -	C C C C C C C C C C C C	Repl.	Fixed
1581	C	- - - - - - - - - - - -	- - - - - -	T T T T T T T T T T T T	Syn.	Fixed
1584	C	- - - - - - - - - - - -	- - - - - -	G G G G G G G G G G G -	Syn.	Poly.
1590	C	T T T T T T T T T T T T	T T T - - -	- - - - - - - - - - - -	Syn.	Poly.
1596	G	- - A A - A - - - - - A	- - - - - -	- - - - - - - - - - - -	Syn.	Poly.
1611	A	- - - - - - - - - - - -	- - - - - -	T T T T T T T T T T T T	Syn.	Fixed
1614	C	- - - - - - - - - - - -	- G - - - -	- - - - - - - - T - - -	Syn.	2 Poly.
1635	C	- - - - - - - - - - - -	- - - - - -	- - T - T - - - - - - -	Syn.	Poly.
1657	A	- - - - - - - - - - - -	- - - - - -	T T T T T T T T T T T T	Repl.	Fixed
1665	C	- - - - - - - - - - - -	- - - - - -	- - - - - - - - - T - -	Syn.	Poly.

The first column contains the nucleotide positions, numbered according to ref. 6, so that the first base of the coding region of exon 2 is 778, of exon 3 is 942 and exon 4 is 1,417. The second column contains the consensus (con.) nucleotide for each site. Nucleotides identical to the consensus are shown as a dash. Several *D. yakuba* individuals were heterozygous; sites which are underlined contained both the nucleotide shown and the consensus nucleotide. The *D. melanogaster* sequences are six Adh^S (a–f) and five Adh^F (g–k) alleles[6] and one Adh^{FChD} allele[10]. *D. simulans* sequences a and b were also from the literature[11,12]. *D. simulans* sequences c–f were of cloned alleles from flies from Australia, Canada, France and the Congo. They were sequenced by the dideoxy chain termination method directly from supercoiled plasmids[13]. The *D. yakuba* alleles of single flies from Brazzaville, Congo were sequenced directly from products of amplification by the polymerase chain reaction[14,15]. Each of these sequences is from one fly from a separate isofemale line. Every allele was completely sequenced in both directions. Each substitution relative to the consensus nucleotide at a site is classified as either fixed or polymorphic. A substitution that is fixed in one species and polymorphic in another is classified as a polymorphism. Because we do not know whether a substitution that is polymorphic in more than one species represents one or two mutations, it is classified as a single polymorphism. Syn, synonymous; Repl., replacement; Poly., polymorphic.

would not appear as polymorphisms. This explanation requires that many replacement mutations have selection coefficients against them that make them neutral in each small population but selected against in each expanded population. It also requires that each population has been large for long enough that slightly deleterious polymorphisms have been eliminated by selection, but not for so long that slightly advantageous back-mutations have replaced the slightly deleterious fixed substitutions. This slightly deleterious model requires so many assumptions about selection coefficients, population sizes, and the times of population expansions that we prefer the simpler explanation, the occasional fixation of an adaptive mutation.

The rate of neutral synonymous mutation at the *Adh* locus in *Drosophila* has been estimated to be about 8×10^{-9} per synonymous site per year[5]. There are about 192 effectively synonymous sites in the coding regions of the *Adh* locus[6], so that the sum of the lengths of the between-species branches of the phylogeny connecting *D. melanogaster*, *D. simulans* and *D. yakuba*, with 17 synonymous substitutions, is about 11 million years (ignoring multiple hits for this rough calculation). If all seven fixed replacement substitutions were adaptive, there was an average of one adaptive fixation every 1.6 million years. It has been suggested that for a plausible value of genetic load, one mutation could be fixed by selection every 300 generations[7].

TABLE 2 Number of replacement and synonymous substitutions for fixed differences between species and polymorphisms within species

	Fixed	Polymorphic
Replacement	7	2
Synonymous	17	42

A *G*-test of independence (with the Williams correction for continuity)[1] was used to test the null hypothesis, that the proportion of replacement substitutions is independent of whether the substitutions are fixed or polymorphic. $G=7.43$, $P=0.006$.

More realistic models of selection suggest that adaptation could occur much more quickly[8,9], but if we accept the conservative figure and assume there are 10 generations per year in *Drosophila*, it would mean that 50,000 loci could be undergoing adaptive evolution as quickly as *Adh*. Thus, the rate of adaptive protein evolution seen at the *Adh* locus may not be unusually high, and selective fixation of adaptive mutations may be a viable alternative to the clocklike accumulation of neutral mutations as an explanation for most protein evolution. □

Received 20 March; accepted 13 May 1991.

1. Sokal, R. R. & Rohlf, F. J. *Biometry* (Freeman, San Francisco, 1981).
2. Lewontin, R. C. *A. Rev. Genet.* **19,** 81-102 (1985).
3. Kreitman, M. & Hudson, R. R. *Genetics* **127,** 565-582 (1991).
4. Nei, M. *Molecular Evolutionary Genetics* (Columbia University Press, New York, 1987).
5. Sharp, P. M. & Li, W.-H. *J. molec. Evol.* **28,** 398-402 (1989).
6. Kreitman, M. *Nature* **304,** 412-417 (1983).
7. Haldane, J. B. S. *J. Genet.* **57,** 511-524 (1957).
8. Maynard Smith, J. *Nature* **219,** 1114-1116 (1968).
9. Sved, J. A. *Am. Nat.* **102,** 283-293 (1968).
10. Collet, C. *J. molec. Evol.* **27,** 142-146 (1988).
11. Bodmer, M. & Ashburner, M. *Nature* **309,** 425-430 (1984).
12. Cohn, V. H. & Moore, G. P. *Molec. Biol. Evol.* **5,** 154-166 (1988).
13. Zagursky, R. J., Baumeister, K., Lomas, N. & Berman, M. L. *Gene Analyt. Techn.* **2,** 89-94 (1985).
14. Saiki, R. K., Bugawan, T. L., Horn, G. T., Mullis, K. B. & Erlich, H. A. *Nature* **324,** 163-166 (1988).
15. Higuchi, R. G. & Ochman, H. *Nucleic Acids Res.* **17,** 5865 (1989).

ACKNOWLEDGEMENTS. We thank S. El-Gogary for cloning the *D. simulans* alleles and J. R. David for the isofemale lines of *D. yakuba*. This work was supported by the NIH and the NSF. DNA sequences have been deposited in GenBank, accession numbers X57361-X57364 (*D. simulans*) and X57365-X57376 (*D. yakuba*).