Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 4548–4551, May 1993
Genetics

Amino acid substitution at the *Adh* locus of *Drosophila* is facilitated by small population size

(nearly neutral theory/synonymous vs. nonsynonymous substitutions)

TOMOKO OHTA

National Institute of Genetics, Mishima 411, Japan

Communicated by Motoo Kimura, January 25, 1993

ABSTRACT **The number of amino acid replacement substitutions and that of synonymous substitutions are examined by using DNA sequences of the *Adh* locus of *Drosophila*. The ratio of replacement to synonymous substitutions is higher in sequence comparisons between species than in polymorphisms within species. The ratio for the between-species comparisons is highest in the Hawaiian group and lowest in the *obscura* group. These observations suggest that amino acid substitutions are facilitated by small population size. The result is in accord with the nearly neutral theory of molecular evolution.**

Based on the neutral theory of molecular evolution (1), and by using available DNA sequences, it is now possible to examine various theories of molecular evolution. McDonald and Kreitman (2) examined the relative numbers of amino acid replacement and synonymous substitutions at the *Adh* locus (encoding alcohol dehydrogenase) in three species of the *Drosophila melanogaster* group and noted that the relative numbers should be the same for within- and between-species comparisons under selective neutrality. They have shown that there are more replacement substitutions for between-species comparisons than the neutral prediction. From this, they suggest that the excess replacement substitutions result from the adaptive fixation of advantageous mutations.

There are now more data to examine the pattern of evolution at the *Adh* locus if their claim of the adaptive fixation is valid. In particular, it is desirable to determine if there is any relationship between the number of amino acid replacement substitutions and population size. This is important because speciation in *Drosophila* is often associated with a bottleneck (3, 4), and the excess amino acid substitutions for between-species comparisons would suggest that amino acid substitution occurs while the population size is small. Although rejected by McDonald and Kreitman (2), the relationship between amino acid substitution and population size is in accord with the prediction of the slightly deleterious or the nearly neutral theory of molecular evolution (5, 6) and needs to be examined in more detail. In this respect, the data of Hawaiian *Drosophila* are especially interesting, because there is evidence for severe bottlenecks in this group (3, 4).

Herein, I examine the numbers of amino acid replacement and synonymous substitutions in *Drosophila* for the Hawaiian and the *obscura* groups. The results indicate that amino acid substitutions are facilitated by small population size. The sequence divergence of an *Adh* locus of some mammalian species is also studied to supplement this result.

DATA ANALYSIS

DNA sequences were obtained from the genetic data bases maintained at the National Institute of Genetics, which include GenBank, DDBJ (DNA Database of Japan), and EMBL. The *Adh* sequences used are listed in Table 1. They include five sequences of Hawaiian *Drosophila* (7), 11 sequences of *D. melanogaster* (8, 9), 3 sequences of *D. simulans* (refs. 10–12), a sequence of *D. yakuba* (from the data base), 16 sequences of *D. pseudoobscura* (13), and a sequence of *D. subobscura* (14). As for the mammalian data, one sequence of each of the four species in Table 1 was used (15–18). The ODEN sequence analysis package developed by Yasuo Ina (National Institute of Genetics) was used to obtain and analyze sequences from the data base.

The numbers of amino acid replacement and synonymous substitutions were counted by hand and by the ODEN package using aligned sequences. The within-species comparisons were done by hand. Using the method of Nei and Gojobori (19), which is incorporated into the ODEN package, the numbers of nonsynonymous and synonymous substitutions for the between-species comparisons were estimated. This method divides the nucleotide substitutions into synonymous and nonsynonymous classes, and then corrections for multiple hits are estimated under the assumption of random mutability among the four kinds of bases (20, 21). For the present analysis, the divergence among sequences is not high, and this method is considered to give satisfactory estimates. By choosing a representative sequence from a species, phylogenetic trees were constructed by the neighbor-joining method (22), also incorporated in the ODEN package, from the corrected distance matrix. The total numbers of nonsynonymous and synonymous substitutions were estimated by summing all branch lengths of the tree.

RESULTS AND DISCUSSION

The number of replacement and synonymous substitutions is given in Table 2 together with the ratio of replacement to synonymous substitutions. The number in parentheses is the value corrected for multiple substitutions. It can immediately be noted from Table 2 that the ratio of replacement to synonymous substitutions is low for polymorphism as compared with that for between-species divergence in accordance with the previous proposal (2). However, note that the values for the *melanogaster* group are different from those of McDonald and Kreitman (2), because some of the sequences they used were not available in the genetic data bases, and different samples were used here. Also note that the sample size is not large, and hence the within-species values may not be quite reliable. The purpose of counting them is simply to show that the ratio of nonsynonymous to synonymous substitutions is lower for the within-species comparisons than for the between-species ones.

Another remarkable fact revealed by the results in Table 2 is that the ratio for the between-species divergence differs among the three groups of *Drosophila*. Let us examine this finding in more detail. I have performed ordinary χ^2 analyses. The estimated number of substitutions for the multiple hits correction is used for the test. Table 3 gives the contingency

---

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Table 1. *Adh* sequences used in this study

Species	Sequence	Source
Hawaiian *Drosophila*		Ref. 7
Drosophila difference	DROADHDD	
Drosophila heteroneura	DROADHDH	
Drosophila picticornis	DROADHDPA	
Drosophila plantibia	DROADHDP	
Drosophila silvestris	DROADHDS	
melanogaster group		Refs. 8–12; *
Drosophila melanogaster	DROADHC(A, B, C, D, E, F, G, H, I, J, K)	
Drosophila simulans	DROADH(AAM, GX, AX)	
Drosophila yakuba	DROADK	
obscura group		Refs. 13, 14
Drosophila pseudoobscura	DROADHAH (6, 10, 12, 13, 14, 15, 16, 54, 69) DROADHBO (31, 32, 34, 35, 36, 39, GE)	
Drosophila subobscura	DROADHBB	
Mammals		
Homo sapiens	HUMADH2BA	Ref. 15
Papio hamadryas	BABADH	Ref. 16
Mus caroli	MUSADHIA	Ref. 17
Rattus norvegicus	RATADH	Ref. 18

*Unpublished result of M. Ashburner.

table, and the χ^2 value is 7.89 with 2 df, thus the probability is $0.01 < P < 0.05$: the difference of the ratio among the three groups is significant. Several additional χ^2 tests were carried out. As for the two groupings (2 × 2 test), χ^2 values are as follows: Hawaiian vs. others, and replacement vs. synonymous, $\chi^2 = 6.47$ with $0.01 < P < 0.05$; *obscura* vs. others, and replacement vs. synonymous, $\chi^2 = 6.03$ with $0.01 < P < 0.05$. However, if the uncorrected number of substitutions is used, the value of χ^2 becomes slightly smaller than the above and is statistically significant only for the test of Hawaiian vs. others.

At any rate, the ratio is highest for the Hawaiian group and lowest for the *obscura* group. Remember that the Hawaiian species of this study are thought to have experienced repeated bottlenecks (4), whereas the *obscura* group apparently maintained large population sizes (8). Thus, the present result suggests that amino acid substitutions are more rapid in small populations than in large ones.

To supplement the above finding, I examined some mammalian *Adh* sequences. It should be noted that the mammalian *Adh* is of separate origin from the insect *Adh* (23), and therefore there is no guarantee that the selective constraints are the same in the two phyla. Nevertheless, the two types of *Adh* show some structural similarities reflecting the common occurrence of similar building units (23), and examination of mammalian *Adh* provides some useful information. The result of the analysis is included in Table 2. The ratio of the replacement to the synonymous substitutions is higher for the human–baboon comparison than for the mouse–rat value, although the χ^2 is 2.3 and the difference is not statistically significant. Nevertheless, the result is compatible with the idea that amino acid substitutions are facilitated by small population sizes, because primates are thought to have had small population sizes compared with rodents (24). This tendency of mammalian gene evolution has been reported for other loci by Li *et al.* (25) and Gillespie (ref. 26, pp. 117–119). These authors examined 20 loci and found that the generation-time effect is more pronounced in synonymous substitutions than in nonsynonymous substitutions—i.e., synonymous evolution is more rapid in rodents than in primates, whereas the difference between the orders of mammals is not so obvious for nonsynonymous changes. As a result, the ratio of nonsynonymous to synonymous substitutions is often higher in primates than in rodents. Further analysis along this line using more loci is necessary.

Table 2. Number of replacement and synonymous substitutions among sampled sequences, between closely related species, and within species (polymorphism)

Species	Replacement	Synonymous	Ratio*
Hawaiian *Drosophila*			
Between species	17 (18.2)	27 (30.3)	0.630 (0.601)
melanogaster group			
Between species	11 (11.1)	30 (32.7)	0.367 (0.339)
Polymorphism	1	18	0.056
obscura group			
Between species	13 (13.0)	55 (69.3)	0.236 (0.188)
Polymorphism	1	8	0.125
Mammalian *Adh*			
Human–baboon	14 (15.0)	16 (16.9)	0.875 (0.888)
Mouse–rat	41 (41.8)	69 (85.3)	0.594 (0.490)

The number in parentheses has been corrected for multiple substitutions.

*Ratio = replacement/synonymous.

What does such a relationship between amino acid substitution and population size imply? As pointed out before, this relationship is consistent with the slightly deleterious mutation or the nearly neutral theory of molecular evolution (5, 6), because in small populations, slightly deleterious replacement mutations are effectively neutral and can accumulate, whereas in large populations, they are selected against.

The nearly neutral theory has been modified to include very slightly advantageous mutations (27–29) and provides a reasonable explanation to the present finding. In other words, *Adh* evolves within the allowed range of functional constraints. The range is bigger in small populations than in large ones, and the fitness of the locus fluctuates in conjunction with substitutions. Let us examine, in this regard, what types

Table 3. A 2 × 3 contingency table for testing independence of substitution type and species group

Substitution	Species group			Total
	Hawaiian	*melanogaster*	*obscura*	
Replacement	18.2 (11.75)	11.1 (10.61)	13.0 (19.94)	42.3
Synonymous	30.3 (36.75)	32.7 (33.19)	69.3 (62.36)	132.3
Total	48.5	43.8	82.3	174.6

$\chi^2 = 7.89$, with $0.01 < P < 0.05$. The number in parentheses is the value under the assumption of independence.

of amino acid substitutions have occurred in Hawaiian *Drosophila*. Table 4 shows the types of amino acid substitutions. It is remarkable to find that more than half of the amino acid substitutions are conservative, such as between valine and isoleucine. Note here that only about one-fourth of single step mutations result in conservative substitutions.

Note that the threonine/lysine polymorphism of *D. melanogaster Adh* is nonconservative. As reported previously (31–33), this polymorphism is likely to be maintained by some kind of balancing selection. I would like to point out that the balancing selection becomes effective in a large population like *D. melanogaster*, but it may be nearly neutral in a small population.

The present findings also have implications for the concept of the molecular clock. Zuckerkandl and Pauling (34) were the first to propose the molecular clock concept based on amino acid changes of a couple of proteins, including hemoglobin and cytochrome *c*. Since then, it has been quite controversial (see ref. 35). The neutral theory (1) provides a simple explanation for the molecular clock, such that the neutral mutation rate is equal to the substitution rate in evolution. In particular, it is assumed that the selective constraint remains more or less constant for those loci that have been established a long time ago, and therefore the proportion of neutral mutations remains mostly constant. One would then expect that amino acid-altering and synonymous substitutions obey the molecular clock pattern. If amino acid substitutions are facilitated by small population sizes, the molecular clock concept needs some revision.

The nearly neutral theory is more complicated in that the rate depends on population size and generation length as well as on the intensity of selection (27–29). It proposes that, for the nearly neutral class, the effect of generation time and that of population size cancel each other, resulting in a more uniform evolutionary rate per year than the strictly neutral theory predicts, provided that mutation depends on generation length. If the amino acid substitution rate becomes rapid relative to the synonymous one while population size is small as in the *Adh* locus, the above prediction of the nearly neutral theory would apply to the amino acid substitutions. DNA evolution of *Drosophila* is estimated to be several times more rapid than that of mammals (36, 37). In other words, the evolutionary rate of a large part of the DNA reflects more faithfully the true mutation rate that depends on generation length than the rate of protein evolution. However, one must be careful to generalize the above prediction, since *Adh* may be an exceptional locus. One needs to know the relative numbers of replacement and synonymous substitutions at more loci.

Table 4. Types of amino acid substitution between sequences of Hawaiian *Drosophila*

Type	No. of occurrences
Conservative*	
Val-Ile	5
Val-Leu	2
Met-Leu	1
Ala-Thr	1
Nonconservative	
Asn-Ser†	1
Asn-His†	1
Asn-Lys†	1
Asn-Thr	1
Ala-Asn	1
Ala-Val	1
Thr-Met	1
Lys-Glu	1

The three-letter code is used for denoting residues.

*Pairs of residues belonging to one of the following groups: Ser, Thr, Pro, Ala, and Gly; Asn, Asp, Gln, and Glu; His, Arg, and Lys; Met, Ile, Leu, and Val; Phe, Tyr, and Trp (30).

†Substitution between these pairs of residues is accepted almost as readily as that between conservative pairs in evolution (30).

The present finding is consistent with the previous results on molecular evolution and polymorphism in relation to the nearly neutral theory. I mention here just a couple of examples. Aquadro *et al.* (38) found that the level of within-species nucleotide diversity differs between species of *Drosophila* even if the level of protein polymorphism is almost the same. They posit that the nucleotide diversity reflects the species population size, whereas the protein diversity is limited by slightly deleterious mutations. Li and Sadler (39) reported that the level of nucleotide diversity of the human population is much lower than that of *Drosophila* population, in spite of the fact that the level of protein diversity is only slightly lower in human. All of these facts consistently suggest that the ratio of the amino acid replacements to synonymous substitutions is higher in small populations. However, it would be difficult to discriminate, from such results, between the nearly neutral model and the episodic model of Gillespie (26).

A final comment is concerned with the report of Ward and Skibinski (40), who show that a major proportion of allozyme variation can be explained by the strictly neutral theory in terms of the quantitative relationship between heterozygosity and genetic distance. Their statistics are based on pooling data from many loci over many species, and the strictly neutral and the nearly neutral models become indistinguishable for such statistics (28). As emphasized by the authors, however, there is no reason to suppose that the relationship between heterozygosity and genetic distance fits the selection theory.

I thank Drs. H. Tachida and F. Tajima for their stimulating discussions. Thanks are also due to them and to Drs. M. Kimura, C. J. Basten, J. B. Walsh, and D. O. F. Skibinski for their valuable comments and to Drs. C. Aquadro and D. Begun for their constructive and helpful criticisms. This is contribution no. 1946 from the National Institute of Genetics. This work was supported by a Grant-in-Aid from the Ministry of Education, Science and Culture of Japan.

1. Kimura, M. (1983) *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge, U.K.).
2. McDonald, J. H. & Kreitman, M. (1991) *Nature (London)* **351**, 652–654.
3. Carson, H. L. (1976) in *Population Biology and Evolution*, ed. Lewontin, R. C. (Syracuse Univ. Press, Syracuse, NY), pp. 123–137.
4. DeSalle, R. & Templeton, A. R. (1988) *Evolution* **42**, 1076–1084.
5. Ohta, T. (1972) *J. Mol. Evol.* **1**, 150–157.
6. Ohta, T. (1973) *Nature (London)* **246**, 96–98.
7. Rowan, R. G. & Hunt, J. A. (1991) *Mol. Biol. Evol.* **8**, 49–70.
8. Kreitman, M. (1983) *Nature (London)* **304**, 412–417.
9. Kreitman, M. & Aquade, M. (1986) *Genetics* **114**, 93–110.
10. Ashburner, M. & Bodmer, M. (1984) *Nature (London)* **309**, 425–430.
11. Cohn, V. H. & Moore, G. P. (1988) *Mol. Biol. Evol.* **5**, 154–166.
12. Laurie, C. C., Heath, E. M., Jacobson, J. W. & Thomson, M. S. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 9674–9678.
13. Schaeffer, S. W. & Miller, E. L. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 6097–6101.
14. Marfany, G. & Gonzalez-Duarte, R. (1992) *Mol. Biol. Evol.* **9**, 261–277.
15. Ikuta, T., Fujiyoshi, T., Kurachi, K. & Yoshida, A. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 2703–2707.
16. Trezise, A. E., Godfrey, E. A., Holmes, R. S. & Beacham, I. R. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 5454–5458.
17. Edenberg, H. J., Zhang, K., Fong, K., Borson, W. F. & Li, T.-K. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 2262–2266.
18. Crabb, D. W. & Edenberg, H. J. (1986) *Gene* **48**, 287–291.
19. Nei, M. & Gojobori, T. (1986) *Mol. Biol. Evol.* **3**, 418–426.

20. Jukes, T. H. & Cantor, C. R. (1969) in *Mammalian Protein Metabolism*, ed. Munro, H. N. (Academic, New York), Vol. 3, pp. 21–132.
21. Kimura, M. & Ohta, T. (1972) *J. Mol. Evol.* **2**, 87–90.
22. Saitou, N. & Nei, M. (1987) *Mol. Biol. Evol.* **4**, 406–425.
23. Jornvall, H. (1986) in *Molecular Evolution of Life*, eds. Baltscheffsky, H., Jornvall, H. & Rigler, R. (Cambridge Univ. Press, Cambridge, U.K.), pp. 231–236.
24. Ohta, T. & Kimura, M. (1971) *J. Mol. Evol.* **1**, 18–25.
25. Li, W.-H., Tanimura, M. & Sharp, P. M. (1987) *J. Mol. Evol.* **25**, 330–342.
26. Gillespie, J. H. (1991) *The Causes of Molecular Evolution* (Oxford Univ. Press, Oxford).
27. Ohta, T. & Tachida, H. (1990) *Genetics* **126**, 219–229.
28. Tachida, H. (1991) *Genetics* **128**, 183–192.
29. Ohta, T. (1992) *Genetics* **130**, 917–923.
30. Dayhoff, M. O., Schwartz, R. M. & Orcutt, B. C. (1978) in *Atlas of Protein Sequence and Structure*, ed. Dayhoff, M. O. (Natl. Biomed. Res. Found., Silver Spring, MD), Vol. 5, Suppl. 3, 345–352.
31. Lewontin, R. C. (1985) *Annu. Rev. Genet.* **19**, 81–102.
32. Kreitman, M. & Hudson, R. R. (1991) *Genetics* **125**, 565–582.
33. Hudson, R. R. (1990) *Oxford Surv. Evol. Biol.* **7**, 1–44.
34. Zuckerkandl, E. & Pauling, L. (1965) in *Evolving Genes and Proteins*, eds. Bryson, V. & Vogel, H. J. (Academic, New York), pp. 97–166.
35. Jukes, T. H., ed. (1987) *J. Mol. Evol.* **26**.
36. Moriyama, E. N. (1987) *Jpn. J. Genet.* **62**, 139–147.
37. Caccone, A., Amato, G. D. & Powell, J. R. (1988) *Genetics* **118**, 671–683.
38. Aquadro, C. F., Lado, K. M. & Noon, W. A. (1988) *Genetics* **119**, 875–888.
39. Li, W.-H. & Sadler, L. A. (1991) *Genetics* **129**, 513–523.
40. Ward, R. D. & Skibinski, D. O. F. (1985) *Genet. Res.* **45**, 315–340.