Copyright © 1996 by the Genetics Society of America

Nucleotide Variation at the *Gpdh* Locus in the Genus *Drosophila*

R. Spencer Wells

Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts 02138

Manuscript received May 27, 1995
Accepted for publication February 8, 1996

ABSTRACT

The *Gpdh* locus was sequenced in a broad range of *Drosophila* species. In contrast to the extreme evolutionary constraint seen at the amino acid level, the synonymous sites evolve at rates comparable to those of other genes. *Gpdh* nucleotide sequences were used to infer a phylogenetic tree, and the relationships among the species of the *obscura* group were examined in detail. A survey of nucleotide polymorphism within *D. pseudoobscura* revealed no amino acid variation in this species. Applying a modified McDonald-Kreitman test, the amino acid divergence between species in the *obscura* group does not appear to be excessive, implying that drift is adequate to explain the patterns of amino acid change at this locus. In addition, the level of polymorphism at the *Gpdh* locus in *D. pseudoobscura* is comparable to that found at other loci, as determined by a Hudson-Kreitman-Aguadé test. Thus, the pattern of nucleotide variation within and between species at the *Gpdh* locus is consistent with a neutral model.

THE study of nucleotide variation in natural populations has revolutionized experimental evolutionary genetics. By comparing DNA sequences within and between closely related species, we are able to detect the results of purifying selection (KREITMAN 1983), balancing selection (HUDSON *et al.* 1987) and adaptive evolution (McDONALD and KREITMAN 1991). The dual nature of a DNA coding sequence allows these inferences to be made: each sequence of nucleotides consists of those that are likely to be subject to strong selective forces (nonsynonymous sites) and those that are largely neutral (synonymous sites). If nucleotide variation is partitioned into synonymous and nonsynonymous classes, the action of selection is detected as a significant deviation from the expected values in either class. Because of the decoupling of synonymous and nonsynonymous variation, DNA sequences are much more information-rich than amino acid sequences.

While evolutionary geneticists have historically studied those genes showing a moderate-to-high level of protein variation within species (KEITH 1983; KREITMAN 1983; RILEY *et al.* 1992), the phenomenon of interspecific protein variation for genes that are monomorphic within species has not been adequately investigated. For example, the *Adh* locus, encoding the enzyme alcohol dehydrogenase (E.C. 1.1.1.99), has been sequenced in many species of the genus *Drosophila.* As it shows no electrophoretic variation in many of these species, it is often taken as the paradigm of monomorphic locus evolution (SCHAEFFER and MILLER 1992a). However, when levels of electrophoretic variation across many species are compared, *Adh* falls in the middle of the distribution of average heterozygosities (POWELL 1976; WELLS 1995). The *Gpdh* locus (encoding glycerol-3-phosphate dehydrogenase, E.C. 1.1.1.8), in contrast, is monomorphic in almost all *Drosophila* species. Of the 205 species that have been assayed for electrophoretic variation at this locus, only two exhibit variant alleles at frequencies high enough to be classified as polymorphic: *D. melanogaster* and *D. subarctica* (LAKOVAARA *et al.* 1977; DE STORDEUR and PASTEUR, 1978; COYNE *et al.* 1979; LAKOVAARA and KERÄNEN 1980). The *Gpdh* locus then is an excellent model system for studies on the molecular evolution of monomorphic proteins.

I have undertaken a study of the nucleotide sequence variation and divergence at this locus in a broad sample of *Drosophila* species: *D. pseudoobscura, D. miranda, D. affinis, D. azteca, D. ambigua, D. bifasciata, D. subobscura* (all members of the *obscura* species group and with *D. melanogaster,* part of the subgenus *Sophophora*), *D. melanogaster, D. busckii* (subgenus *Dorsilopha*), *D. virilis* (subgenus *Drosophila*) and *D. lebanonensis* (subgenus *Scaptodrosophila*). The species were chosen to provide a range of comparisons among both closely and distantly related taxa, and population variation was assayed within a species (*D. pseudoobscura*) where good comparative data exists from other loci. Two questions have been adressed in the present study: (1) Is the amino acid divergence between species at this locus the result of drift or adaptive evolution? (2) What insights can this locus provide into the phylogenetic relationships among the species studied?

MATERIALS AND METHODS

Fly stocks: Stocks of *D. affinis, D. azteca, D. ambigua, D. bifasciata, D. busckii,* and *D. lebanonensis* were obtained from the National Drosophila Species Resource Center (Bowling Green, OH). *D. subobscura* genomic DNA was obtained from

Corresponding author: R. Spencer Wells, Department of Structural Biology, Fairchild Center, Stanford University School of Medicine, Stanford, CA 94305. E-mail: wells@lotka.stanford.edu

TABLE 1

Number of interspecific synonymous (K_s, above diagonal) and nonsynonymous (K_a, below diagonal) substitutions per site in the *Gpdh* coding region using the method of LI *et al.* (1985)

	aff	amb	azt	bif	bus	leb	mel	mir	pse	sub*	vir
aff	—	0.2295	0.0569	0.2576	1.0276	1.0176	0.9090	0.1721	0.1932	0.2877	0.7195
amb	0.0035	—	0.2288	0.1182	0.9850	1.0350	0.8561	0.1518	0.1677	0.1859	0.7859
azt	0.0012	0.0024	—	0.2632	1.0285	1.0701	0.8367	0.1766	0.1978	0.2926	0.7344
bif	0.0047	0.0047	0.0059	—	1.0015	0.9945	0.9437	0.1533	0.1695	0.1762	0.8259
bus	0.0073	0.0109	0.0085	0.0121	—	0.9623	1.0680	1.0236	1.0744	1.1113	0.6922
leb	0.0143	0.0107	0.0131	0.0131	0.0149	—	1.3121	0.9905	1.0436	1.1199	1.0029
mel	0.0053	0.0062	0.0065	0.0062	0.0074	0.0093	—	0.9109	0.9191	0.9563	1.1210
mir	0.0024	0.0012	0.0012	0.0047	0.0097	0.0119	0.0077	—	0.0260	0.2256	0.8052
pse	0.0024	0.0012	0.0012	0.0047	0.0097	0.0119	0.0077	0.0000	—	0.2441	0.8595
sub*	0.0037	0.0037	0.0024	0.0049	0.0101	0.0111	0.0068	0.0024	0.0024	—	0.8807
vir	0.0121	0.0109	0.0109	0.0122	0.0094	0.0125	0.0099	0.0097	0.0097	0.0089	—

Abbreviations: aff, *D. affinis;* amb. *D. ambigua;* azt, *D. azteca;* bif, *D. bifasciata;* bus, *D. busckii;* leb, *D. lebanonensis;* mel, *D. melanogaster;* mir, *D. miranda;* pse, *D. pseudoobscura;* sub, *D. subobscura;* vir, *D. virilis.* Mean number of synonymous sites compared = 235.3.

* Comparisons to *D. subobscura* do not include exons 7 and 8.

J. ROZAS, University of Barcelona (extracted from an isofemale line caught in Barcelona, Spain). A *D. miranda* genomic library in λEMBL4 was provided by R. NORMAN, Arizona State University. One *D. pseudoobscura* sequence (line JR45) has been published previously (WELLS 1995). The other *D. pseudoobscura* sequences used in intraspecific sequence comparisons were obtained from genomic DNA provided by S. SCHAEFFER, Pennsylvania State University (lines PS281, PS297, PS298, PS299, PS314, PS315 collected at Kaibab National Forest, Arizona). All *D. pseudoobscura* lines were isochromosomal for chromosome *4*, which contains both *Adh* and *Gpdh.*

Genomic DNA preparation and cloning: DNA was extracted from adult flies by standard methods (AUSUBEL *et al.* 1987; ASHBURNER 1989) and cut with restriction endonucleases *Eco*RI, *Bam*HI, *Hin*dIII, *Pst*I, *Sac*I and *Xho*I. After separation on a 0.7% agarose gel and capillary transfer to a nylon membrane, the blot was probed using a mixture of the partial *D. melanogaster* cDNA probes Gpd-411 (VON KALM *et al.* 1989) and pNB-1 (provided by R. MACINTYRE). A single band was seen for at least one restriction enzyme digest in each species, and these enzymes were used for library construction. The enzymes used were as follows: *D. affinis, Eco*RI; *D. ambigua, Bam*HI; *D. azteca, Bam*HI; *D. bifasciata, Hin*dIII; *D. busckii, Eco*RI; *D. lebanonensis, Bam*HI; *D. subobscura, Hin*dIII.

Libraries were constructed in λDASH (Stratagene). Approximately 5 μg of genomic DNA were cut to completion for each library with the appropriate restriction enzyme, extracted with phenol/chloroform and ethanol precipitated. λDASH was cut with the corresponding enzyme and *Xho*I to prevent self-ligation. Ligation of genomic DNA and vector was performed overnight at 4°. Packaging reactions were performed according to manufacturer's directions (Gigapack II Gold, Stratagene), and phage were titered on P2392 cells. Approximately 50,000 recombinant phage were screened, and two positives were picked for each species. Recombinant phage (100,000) were screened for the *D. miranda* library, yielding two positive clones.

λ DNA was extracted using standard procedures (AUSUBEL *et al.* 1987) and cut with the same enzyme used to construct the library (the *D. miranda Gpdh* region was subcloned using *Eco*RI). For each species, a fragment corresponding to the correct size was subcloned into pEMBL19 (+), and the double-stranded plasmid DNA was used as a sequencing template. In some cases, the restriction enzyme used in constructing

TABLE 2

Amino acid variation at the *Gpdh* locus in the genus *Drosophila*

	Amino acid position																
Species	3	28	45	47	72	81	171	172	178	194	196	288	315	337	351	353	362
D. melanogaster-F	D	A	L	D	P	V	T	D	V	D	A	E	G	N	D	S	N
D. affinis	·	S	M	·	·	·	·	·	·	·	·	·	K	·	·	·	·
D. azteca	·	S	M	·	·	·	·	·	·	·	·	·	K	K	·	·	·
D. ambigua	·	S	·	·	·	·	·	·	·	·	·	·	K	K	·	·	T
D. bifasciata	·	·	V	·	·	·	·	E	·	·	·	·	K	·	·	·	T
D. subobscura	·	·	M	·	·	A	·	·	·	·	·	·	K	K	*	*	*
D. pseudoobscura	·	S	M	·	·	·	·	·	·	·	·	·	K	K	·	·	T
D. miranda	·	S	M	·	·	·	·	·	·	·	·	·	K	K	·	·	T
D. virilis	E	·	M	·	T	·	·	·	·	E	·	D	·	K	·	F	T
D. busckii	E	S	M	·	·	·	K	·	·	·	S	·	·	·	·	F	·
D. lebanonensis	·	·	·	E	·	·	N	·	I	G	·	·	·	K	E	·	T

·, identity with the *D. melanogaster-F* sequence.

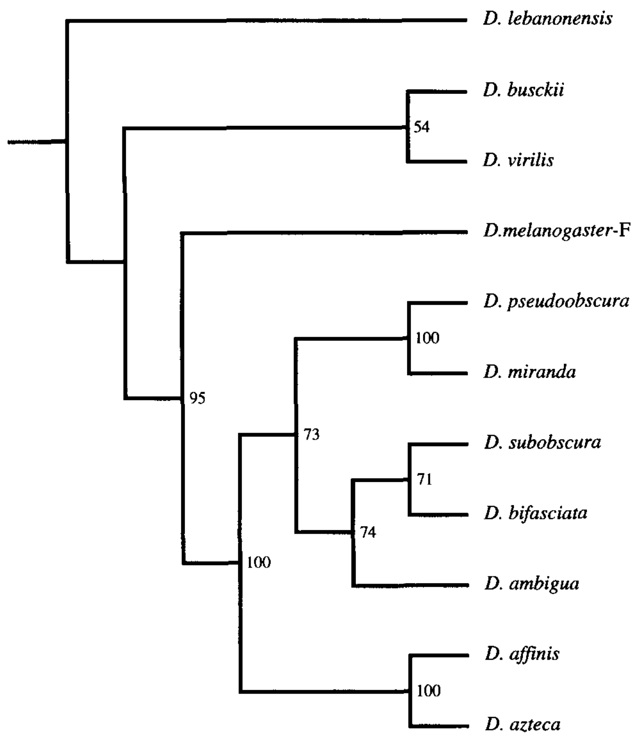

FIGURE 1.—Maximum parsimony tree of *Drosophila Gpdh* data. The complete nucleotide sequence of *Gpdh* exons 1–6 was used to generate the tree in PAUP (SWOFFORD 1991). The branch-and-bound search algorithm was used, and bootstrapping was performed with the same search algorithm. Bootstrap values (number of times supported out of 100 replicates) are shown inside the nodes.

the library resulted in a truncated *Gpdh* clone. PCR amplification from genomic DNA and direct sequencing (see below for methods) and/or recloning from the same library were then used to obtain the missing regions. *D. subobscura* exons 7 and 8 could not be cloned or PCR amplified, and are not included in the present analysis.

PCR: The *Gpdh* genomic region was PCR amplified from *D. pseudoobscura* genomic DNA using two sets of primers (nucleotide positions in the sequence of WELLS 1995): 140–162 and 829–851 for exons 1 and 2; 3510–3532 and 6429–6451 for exons 3–8. After amplification with one phosphorylated primer, the double-stranded product was digested with λ-exonuclease to yield a single-stranded template for sequencing (HIGUCHI and OCHMAN 1989), following the procedure described by BERRY *et al.* (1991).

Sequencing: Both strands were sequenced by the dideoxy chain termination method (Sequenase 2.0, US Biochemical) using oligonucleotides specific for the *Gpdh* genomic region. Electrophoresis was carried out on "Long Ranger" gels (AT Biochem). Sequence was obtained from exons 1–8, encompassing the entire translated portion of the *Gpdh* gene. In addition, introns 1, 3, 4 and 5 were sequenced in the *D. pseudoobscura* lines.

Analysis: Nucleotide sequences were aligned manually with the *Gpdh* sequences of *D. melanogaster* (BEWLEY *et al.* 1989), *D. virilis* (TOMINAGA *et al.* 1992) and *D. pseudoobscura* (WELLS 1995). Estimates of synonymous and nonsynonymous substitution were calculated using the program of LI *et al.* (1985). Phylogenetic analyses were performed with the programs MacClade (MADDISON and MADDISON 1992), PAUP (SWOFFORD 1991) and MEGA (KUMAR *et al.* 1993).

RESULTS

Species divergence: A total of 1098 bp of sequence was obtained for each species except *D. subobscura*, where exons 7 and 8 (encompassing positions 1054–1098 in the exon contig, where position 1 is the first nucleotide of the initiation methionine) were not sequenced because they could not be cloned or PCR-amplified using *obscura* group-specific primers. No insertion-deletion events were observed in the coding region. The overall level of amino acid replacement was low, a result consistent with the slow evolutionary rate of this protein (BEWLEY *et al.* 1989; WELLS 1995). The synonymous sites, however, evolved at a rate similar to that of other genes that have been compared between these species (WELLS 1995). The estimated number of synonymous and nonsynonymous replacements per site, K_s and K_a, are shown in Table 1.

TABLE 3

Test of rate constancy at the *Gpdh* locus using the method of TAJIMA (1993)

Species			No. of unambiguous substitutions			
1	*2*	*3*	m_1	m_2	χ^2	p
aff	amb	mel	25	17	1.52	0.22
aff	bif	mel	24	24	0	1.00
aff	pse	mel	18	18	0	1.00
aff	sub	mel	25	24	0.02	0.89
amb	bif	mel	9	17	2.46	0.12
amb	pse	mel	13	21	1.88	0.17
amb	sub	mel	13	20	1.48	0.22
bif	pse	mel	18	18	0	1.00
bif	sub	mel	17	16	0.03	0.86
pse	sub	mel	23	22	0.02	0.88
mel	pse	leb	69	55	1.58	0.21
bus	vir	leb	51	54	0.09	0.77

TABLE 4

Nucleotide variation in seven *D. pseudoobscura* lines at the *Gpdh* locus

	285	300	323	324	378	415	541	3619	3643	3907	3919	3928	3994	4015	4027	4105	4189	4201	4246	4271	4278	4279	4285	4295	4301	4313	4319	4332	4380	4489	4490	4494	4498
JR45	C	T	T	C	G	T	A	G	C	T	C	G	C	T	T	T	T	G	G	G	T	Δ_4	G	T	G	C	T	C	A	A	A	C	C
PS281	·	Δ_1	·	·	·	·	·	·	·	C	·	·	·	·	G	·	·	A	·	C	C	A	·	·	·	T	·	T	·	C	·	·	·
PS297	·	·	·	·	A	C	G	A	·	C	G	·	·	·	·	·	·	·	A	·	C	A	·	·	·	T	A	T	·	T	G	T	T
PS298	·	·	·	A	·	·	·	·	A	C	·	C	·	·	·	·	·	·	·	·	·	A	·	C	C	T	·	T	·	C	·	·	·
PS299	A	·	G	·	·	·	G	·	A	C	·	·	·	·	·	C	C	·	·	·	·	A	·	·	·	T	·	T	·	C	·	·	·
PS314	·	·	·	·	·	·	·	·	·	C	·	·	·	·	·	·	·	·	·	·	·	A	·	·	·	T	·	T	G	C	·	T	·
PS315	·	·	·	·	·	·	·	·	·	C	·	·	T	C	·	·	·	·	·	·	·	A	A	·	C	T	·	T	G	T	G	T	T

Numbers refer to the sequence of WELLS (1995); Δ_1, deletion of TGGCCCT repeat unit; Δ_4, deletion of AA.

The inferred amino acid sequences are extremely conserved between species (Table 2). There are only 17 variable amino acid positions among these species, and the largest number of pairwise differences is 12 (between *D. lebanonensis* and *D. busckii*). Six sites exhibit more than one amino acid replacement (sites 28, 45, 171, 194, 337 and 362). Interestingly, the majority of these multiple changes are parallelisms or reversals. The high level of homoplasious amino acid replacement among species at this locus is discussed elsewhere (WELLS 1996) and will not be presented in detail here.

Nucleotide sequences of the coding region of exons 1–6, a total of 1050 nucleotides, were used to construct a phylogenetic tree. *D. lebanonensis* was used as the outgroup; several recent studies have shown that the subgenus *Scaptodrosophila,* of which *D. lebanonesis* is a member, represents the most ancient bifurcation in the genus *Drosophila* (GRIMALDI 1990; VILLAROYA and JUAN 1991; DESALLE 1992; KWIATOWSKI *et al.* 1994). The single maximum parsimony (MP) result is shown in Figure 1. There were 175 phylogenetically informative sites, and the treelength is 373. Two approaches were used to assess the reliability of this tree. First, the dataset was bootstrapped (FELSENSTEIN 1985), and the resulting values are shown in Figure 1. Second, neighbor joining (NJ; SAITOU and NEI 1987) trees were constructed using both synonymous and nonsynonymous sites. The tree topology obtained using synonymous sites (correction of JUKES and CANTOR 1969) is identical to that of the MP tree shown in Figure 1, while the nonsynonymous sites produced a completely nonsensical tree, given our current knowledge of relationships in the genus *Drosophila.* For instance, *D. pseudoobscura, D. miranda, D. azteca* and *D. ambigua* form a monophyletic group, as do *D. melanogaster* and *D. busckii.* Because of the high level of amino acid homoplasy in the dataset (WELLS 1996), it seems reasonable to exclude nonsynonymous sites from the NJ analysis. Thus, the NJ tree is identical to the MP tree.

The "molecular clock" test of TAJIMA (1993) was applied to the species in the *obscura* group and to the subgenera to assess whether the rate of nucleotide substitution among these species is constant. Briefly, this test measures whether the number of unambiguous changes on the terminal branches of a tree is the same between two species with a known outgroup. The results are shown in Table 3. None of the comparisons are significant, implying that the rate of nucleotide variation at the *Gpdh* locus is uniform among these species.

Within-species variation: Thirty-three variable nucleotide positions were found within the 1461 nucleotides assayed for variation in *D. pseudoobscura* (Table 4), 16 in the amino acid coding region of the exons and 17 in the introns. All of the variable positions are silent (either synonymous or noncoding). Six of the seven lines sequenced are from a single population (Kaibab National Forest, Arizona), and the other line (JR45) is from a population 400 miles away (James Reserve, California). Previous nucleotide polymorphism studies have shown, however, that there is no detectable population substructure among North American *D. pseudoobscura* (RILEY *et al.* 1989; SCHAEFFER and MILLER 1992b), and samples from different geographic locations can be treated as a single population; this is in contrast to the high degree of interpopulation differentiation seen for chromosomal inversions in this species (LEWONTIN *et al.* 1981). The number of pairwise differences between *D. pseudoobscura Gpdh* haplotypes ranges from six to 17 (mean = 10.9), and JR45 is not excessively different from the other lines. This is confirmed in the NJ

PS298
PS299
JR45
PS281
PS314
PS297
PS315

FIGURE 2.—Neighbor joining tree of *D. pseudoobscura Gpdh* alleles (correction of JUKES and CANTOR 1969).

TABLE 5

McDonald-Kreitman analysis of *Gpdh* divergence within the *obscura* group, comparing to 16 synonymous segregating sites and 0 nonsynonymous segregating sites within *D. pseudobscura*

Species	Fixed differences		p
	Synonymous	Nonsynonymous	
D. affinis	36	2	0.49
D. azteca	38	1	0.71
D. ambigua	33	1	0.68
D. bifasciata	33	4	0.23
D. miranda	4	0	1.00
D. subobscura	44	2	0.55

Fisher's exact p values shown are for the 2 × 2 contingency table. *D. subobscura* comparison does not include exons 7 and 8.

tree (Figure 2) where JR45 clusters among the Kaibab lines. Thus, all lines were combined as a single population sample for the following analyses.

The test of McDonald and Kreitman (1991) was applied to the data from the *obscura* group species; comparisons among more distantly related species were not performed to avoid having to correct for multiple hits. For the purposes of the present analysis, population data from a single species was used (the seven *D. pseudoobscura* lines discussed above). This should not bias the test excessively, as the infinite alleles model (which holds approximately for most DNA polymorphism data) predicts that new segregating sites should be found as frequently within a sample from one species (population) as another, assuming the long-term effective population sizes of the two species (populations) are the same (Li 1977; Kimura 1983). The results are shown in Table 5. None are significant, implying that drift alone can explain the amino acid divergence between these species.

The test of Hudson, Kreitman and Aguadé (1987) was applied to the *Gpdh* polymorphism data using *Adh* as the comparison locus. For this test, I used the *Adh* polymorphism data of Schaeffer and Miller (1993) from seven lines of *D. pseudoobscura,* six of which (the Kaibab lines) were the same as those from which *Gpdh* was sequenced; the other line, MV21, was collected in Mesa Verdé, Colorado (Schaeffer and Miller 1992, 1993). *D. ambigua* was used as the comparison species for both loci. Only synonymous positions in the coding region were used, as these could be aligned unambiguously between the species and there is little worry of adaptive fixation at these sites elevating the divergence estimates. The results are shown in Table 6. The observed level of synonymous nucleotide polymorphism at the *Gpdh* locus is not significantly higher than that expected from the divergence data (P = 0.36). Polymorphism is not distributed uniformly along the sequence, as there is an excess in exon 4. Of the 16 segregating sites in the coding region, 10 are found in exon 4 (uniform expectation = 5.68; χ^2 = 5.09, 1 d.f., P = 0.024). This is demonstrated graphically in a sliding-window plot of the number of segregating sites (Figure 3).

TABLE 6

Hudson-Kreitman-Aguadé analysis of *Gpdh* synonymous sites

Locus	No. of segregating sites		No. of differences	
	Observed	Expected	Observed	Expected
Gpdh	16	12.316	36	39.684
Adh	11	14.684	51	47.316

χ^2 = 0.842, p = 0.36.

DISCUSSION

Pattern of nucleotide substitution: The number of synonymous substitutions per site, K_s, at the *Gpdh* locus is comparable to that observed for other genes compared between two reference species, *D. melanogaster* and *D. pseudoobscura* (Wells 1995). The level of nonsynonymous substitution, however, is the lowest yet found for a gene compared between these two species (Wells 1995). The ratio of these two values provides a measure of the strength of purifying selection acting on the amino acid sequence of the protein (Riley 1989). In Figure 4, the distribution of K_s/K_a values among all possible pairwise species comparisons is shown. The values at the lower end of the distribution show evidence of a higher level of amino acid substitution relative to their level of synonymous substitution, although not enough to result in a significant deviation using the test of McDonald and Kreitman (1991). Those values at the upper end of the distribution exhibit a dearth of amino acid substitutions for their level of synonymous divergence; the highest K_s/K_a value obtained is 171 in the comparison of *D. melanogaster* and *D. affinis,* which have a level of synonymous divergence comparable to that found in comprisons of *D. melanogaster* with other *obscura* group species but differ at only three amino acid positions (other *obscura* group species differ at four to five amino acid positions). The mean K_s/K_a value obtained from the distribution (94.6) is comparable to that obtained from four phylogenetically independent observations (*affinis vs. melanogaster, pseudoobscura vs. ambigua, subobscura vs. bifasciata,* and *busckii vs. virilis;* mean K_s/K_a = 105.1). It is interesting to note that the number of differences between *D. pseudoobscura* JR45 and *D. miranda* (six) is within the range for the number of differences between the *D. pseudoobscura* alleles (six to 17). These species are estimated to have diverged 1–4 million years ago (Schaeffer and Miller 1992a; Beckenbach *et al.* 1993); more *D. miranda* sequences should be obtained to confirm this observation.

Divergence times: The values of K_s given in Table 1 can be used to infer the divergence times of the taxa

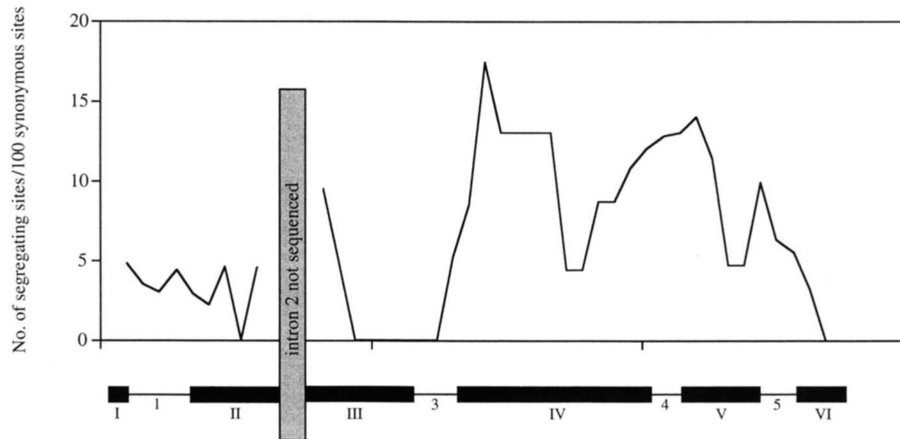

FIGURE 3.—Sliding window plot of the number of segregating sites per 100 synonymous sites at the *Gpdh* locus in *D. pseudoobscura*. Window size, 100 nucleotides; step, 30 nucleotides. Note that intron 2 (3 kb) was not sequenced, and therefore it is not included in the analysis shown here. A gray bar is shown where this intron would be found. Exons and introns are denoted by roman and arabic numerals, respectively.

in this study. Two extremes have been suggested for the divergence time between *D. melanogaster* and *D. pseudoobscura* based on nonnucleotide data: THROCKMORTON (1975) suggested, based on biogeographic information, that these species separated 30–35 million years ago, while BEVERLEY and WILSON (1984) modified this estimate using immunological distances, arriving at a value of ~45 million years. Using these two extremes and the relative levels of synonymous divergence per site in the other species estimated by two methods (LI *et al.* 1985; INA 1995), I have calculated the estimated divergence times between the species in this study (Table 7). These values are, on average, one-half of those obtained by BECKENBACH *et al.* (1993) in their analysis of nucleotide divergence in the mitochondrial cytochrome oxidase subunit II gene. BECKENBACH *et al.* used the number of transversions to calculate their divergence times; as the number of transversions was small in some of these comparisons, their estimates have a

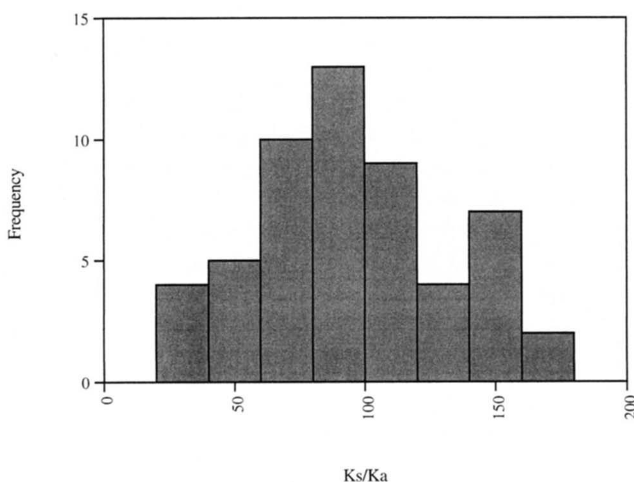

FIGURE 4.—Histogram of K_s/K_a ratios (calculated according to the method of LI *et al.* 1985) for *Gpdh;* all possible pairwise comparisons were made among the species in Table 1.

large variance, which might explain some of the differences between the values obtained by the two studies. Also, if the cytochrome oxidase subunit II gene has undergone adaptive amino acid fixations between any of the species in this study, we would expect those fixations to artificially inflate the divergence between the species involved (MCDONALD and KREITMAN 1991).

Phylogenetic analyses: How reliable is the gene tree obtained from the *Gpdh* data? The confirmation of the subgenus *Sophophora* (*melanogaster* + *obscura* groups) as a monophyletic group is not surprising; both molecular and morphological studies support monophyly for these species (STURTEVANT 1939, 1942; THROCKMORTON 1975; DESALLE 1992). The grouping of *D. busckii* and *D. virilis* is in agreement with recent work on the superoxide dismutase (*Sod*) gene by KWIATOWSKI *et al.* (1994). The relationships within the *obscura* species group are more contentious, however. Given the importance of two of the members of this group in genetic research (*D. pseudoobscura* and *D. subobscura*), it is worth examining their phylogenetic relationships in detail.

Early taxonomic work distinguished between two subgroups in the *obscura* group, the *affinis* subgroup (containing *D. affinis* and *D. azteca*) and the *obscura* subgroup (containing the other species in this study, both Nearctic and Paelarctic), on the basis of morphology (STURTEVANT 1942; BUZZATI-TRAVERSO and SCOSSIROLI 1955). More recent work has placed the *affinis* group as a sister clade to the nearctic species (*D. pseudoobscura* and *D. miranda*), with the palearctic species (*D. subobscura, D. bifasciata* and *D. ambigua*) separating first (LAKOVAARA and SAURA 1982; LATORRE *et al.* 1988; GODDARD *et al.* 1990; BECKENBACH *et al.* 1993). Only one of these studies (BECKENBACH *et al.* 1993) used DNA sequence data, however; the others used allozyme, DNA restriction map and DNA-DNA hybridization data for inferring the tree. The perils of allozyme analyses are well known, particularly the problem of homoplasy among different

TABLE 7

Estimates of divergence time based on *Gpdh* synonymous substitutions per site

Taxa compared	No. of pairwise comparisons	Mean K_s (LWL)	Divergence time (LWL)		Mean K_s (INA)	Divergence time (INA)	
D. melanogaster-obscura group	7	0.9045	30.0	45.0	0.7041	30.0	45.0
affinis subgroup-nearctic *obscura*	4	0.1850	6.1	9.2	0.1418	6.0	9.1
affinis subgroup-palearctic *obscura*	6	0.2599	8.6	12.9	0.2077	8.8	13.3
palearctic *obscura*-nearctic *obscura*	6	0.1853	6.1	9.2	0.1420	6.1	9.1
Sophophora-Dorsilopha	8	1.0400	34.5	51.7	0.8749	36.0	55.9
Sophophora-Drosophila	8	0.8415	27.9	41.9	0.6826	29.1	43.6
Sophophora-Scaptodrosophila	8	1.0729	35.6	53.4	0.8948	38.1	57.2
Dorsilopha-Drosophila	1	0.6922	23.0	34.4	0.5478	23.3	35.0

Estimates were determined using the methods of LI *et al.* (LWL; 1985) and INA (1995).

electromorphs (LEWONTIN 1991). The mtDNA restriction map survey of LATORRE *et al.* (using six-cutter restriction fragment length polymorphisms) is also plagued by the problem of homoplasy, as well as the fact that *D. azteca* was the only *affinis* subgroup species included. The DNA-DNA hybridization data of GODDARD *et al.*, which may be more representative of the entire genome than are single-gene trees, have not been analyzed using improved methods of tree reconstruction (*e.g.*, neighbor joining), and there are no estimates of reliability on the Fitch-Margoliash and UPGMA trees presented.

In their analysis, BECKENBACH *et al.* (1993) used both complete sequences and transversions at the mitochondrial cytochrome oxidase subunit II locus to infer trees of the *obscura* group species. While analyses based solely on transversions obviate the problem of biased transition/transversion ratios inherent in mitochondrial studies (TAMURA and NEI 1993), the number of informative sites is reduced. The bootstrap value that they obtain for the nearctic *obscura/affinis* subgroup clade is 62, lower than that obtained for the nearctic *obscura*/palearctic *obscura* clade in the *Gpdh* analysis, 73. However, in the MP analysis of the *Gpdh* nucleotide data, the treelength only increases from 373 to 377 if the *affinis/azteca* clade is placed as a sister group to the *pseudoobscura/miranda* clade, showing that the most likely relationship among these subgroups may be a trichotomy. The trichotomy hypothesis is supported by an examination of the level of synonymous substitution between the *obscura* group species (Table 1). *D. subobscura* shows a higher K_s relative to *pseudoobscura/miranda* than do the other two palearctic species, resulting in the palearctic species grouping further from *pseudoobscura/miranda* than do *affinis/azteca;* when *D. subobscura* is removed, the three groups appear to be roughly equidistant from each other. This agrees with BARRIO *et al.*'s (1994) recent phylogenetic analysis of a 2-kb region of the mitochondrial genome in the *obscura* group. Depending on the gene analyzed and the algorithm used, different topologies are obtained for the *obscura* and *affinis* group species. BARRIO *et al.* conclude that the Palearctic *obscura* group is polyphyletic (split into *subobscura* and *bifasciata/ambigua* lineages), and that the overall branching pattern is consistent with a rapid phyletic radiation of the subgroup species. While the *Gpdh* data do not support polyphyly for the Palearctic obscura group (although *D. subobscura* does seem to be the most divergent species), they are consistent with a rapid radiation of the subgroups.

STURTEVANT (1942) originally split the *obscura* group into the *affinis* and *obscura* subgroups on the basis of differences in the number of achrostichal hairs on the mesonatum, the number of teeth on the sex combs, and the shape of the carina and testes. Hybridization experiments show that *D. ambigua* is capable of hybridizing to both *D. pseudoobscura* and *D. persimilis* (KOSKE 1953; BUZZATI-TRAVERSO and SCOSSIROLI 1955), and suggest that *D. bifasciata* and *D. persimilis* can hybridize as well (KOSKE 1953). In contrast, no crosses between *affinis* subgroup species and either Palearctic or Nearctic *obscura* subgroup species have produced viable larvae (BUZZATI-TRAVERSO and SCOSSIROLI 1955). Thus, much of the nonmolecular data appears to agree with the present grouping of Nearctic and Paelarctic *obscura* subgroups as sister taxa. Molecular analysis of two "odd" species, *D. alpina* (thought to be the most basal branch in the Palearctic *obscura* subgoup; LAKOVAARA and SAURA 1982) and *D. helvetica* (the only Palearctic member of the *affinis* subgroup, it has some affinities to the Nearctic *obscura* subgroup species; LAKOVAARA and SAURA 1982) might shed some light on the relationships between the three subgroups.

MCDONALD-KREITMAN analysis: The test of MCDONALD and KREITMAN (1991) does not yield a significant result for these data, implying that adaptation has not played a major role in the amino acid divergence between species at the *Gpdh* locus. This test is incapable of detecting evidence for adaptation when the number of amino acid replacements between species is small, however. Single (or a few) amino acid replacements may compose a substantial portion of adaptations at the molecular level (PERUTZ 1984); if so, then some of the variation found to be evolving "neutrally" using this

test may in fact be adaptive. Unfortunately, although statistical approaches that detect the "footprint" of natural selection are at present better than functional approaches used to decipher the causes of variation, there is an inherent limit to the types of events that they can detect. In the case of the McDONALD-KREITMAN test, this limit, given a level of intraspecific variation with a synonymous:nonsynonymous polymorphism ratio of 16:0 (that of *Gpdh* in the present study), would be nine to 11 nonsynonymous replacements between the species shown in Table 5 (calculations not shown). For proteins that diverge very slowly between species, this level of nonsynonymous divergence is never realized. The question of whether amino acid variation in slowly evolving proteins is adaptive remains largely unanswerable with current statistical methods.

HKA analysis: The HUDSON-KREITMAN-AGUADÉ test shows that the level of nucleotide polymorphism within the *Gpdh* coding region does not deviate from neutral expectations, implying that the locus is not subject to balancing selection and that it has not undergone a recent slective sweep. Thus, as with the *Adh* locus in *D. pseudoobscura* (SCHAEFFER and MILLER 1992a), the electrophoretic monomorphism of the GPDH protein seems to result from strong purifying selection acting to remove amino acid variation. It is interesting, however, that there appears to be an excess of polymorphism in the exon 4-intron 4 region (Figure 3) when there is no amino acid variation. Perhaps there is a functionally important change at a synonymous site, as has been described recently by RICHARD and BECKMAN (1995).

The *D. pseudoobscura* polymorphism data may help to explain an anomalous result obtained by TAKANO *et al.* (1993) in their survey of nucleotide polymorphism at the *D. melanogaster Gpdh* locus. In *D. melanogaster* there is a fast/slow electrophoretic polymorphism caused by a T/A transversion at site 3338 in exon 6 of the gene (BEWLEY *et al.* 1989). This polymorphism has several features in common with the fast/slow polymorphism at *Adh* in this species, including a significant cline in allele frequencies on three continents (OAKESHOTT *et al.* 1982, 1984). In their investigation of polymorphism at the *Gpdh* locus, TAKANO *et al.* found an excess of polymorphism in the coding region relative to the introns and the 5′-flanking region, similar to the situation seen at *Adh* (HUDSON and KREITMAN 1991). Unfortunately, this excess is centered on exon 4, 500 bp upstream (5′) from the fast/slow site. Thus, the prediction of STROBECK (1983) with respect to balanced polymorphisms is not seen: that the excess of linked polymorphism seen at a site subject to balancing selection is centered around the selected site, and the excess arises as a consequence of divergence between the two allelic lineages within this region. TAKANO *et al.* were unable to explain this anomalous result. Interestingly, in *D. pseudoobscura,* we also see an excess of polymorphism in

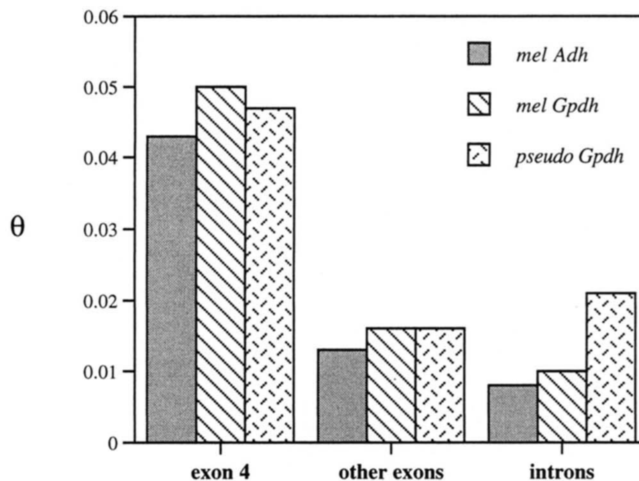

FIGURE 5.—Comparison of values of θ (silent sites) for different regions of *Adh* in *D. melanogaster* and *Gpdh* in *D. melanogaster* and *D. pseudoobscura.* The *Adh* fast/slow site in *D. melanogaster* is found in exon 4, while the *Gpdh* fast/slow site in *D. melanogaster* is found in exon 6. This figure is based on one shown in TAKANO *et al.* (1993).

exon 4 (Figure 5) relative to introns and other exons. That this excess is not sufficient to cause a significant deviation from the neutral expectation for the entire locus implies that in *D. melanogaster,* which has not been analyzed using an HKA test (TAKANO *et al.* 1993), we might also expect to find that the excess is not significant. This result, coupled with the phylogenetic relationships among the *D. melanogaster Gpdh* alleles (which show evidence for multiple mutations at the fast/slow site; TAKANO *et al.* 1993), argues against the selective maintenance of the fast/slow *Gpdh* polymorphism in *D. melanogaster.* Two questions are raised by the possibility that this polymorphism is not maintained by selection: (1) Why is there an excess of polymorphism in exon 4 in both species, when this exon does not exhibit a concomitant excess of interspecific divergence (WELLS 1995)? (2) Why is there a cline in fast/slow allele frequencies in *D. melanogaster*? Further sequencing surveys in both species should begin to address these questions and delineate the forces controlling variation at the *Gpdh* locus.

I thank A. BERRY, R. C. LEWONTIN, A. BARBADILLA and R. KLIMAN for advice and discussion. R. NORMAN provided the *D. miranda* genomic library, and J. ROZAS provided the *D. subobscura* DNA. Thanks to the National Drosophila Species Stock Center for providing fly stocks and to D. WEINREICH for sequencing assistance. This study was supported by National Institutes of Health grant GM-21179 to R. C. LEWONTIN. The author was a Howard Hughes Medical Institute Predoctoral Fellow during this study.

LITERATURE CITED

ASHBURNER, M., 1989 *Drosophila: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

AUSUBEL, F. M., R. BRENT, R. E. KINGSTON, D. D. MOORE, J. G. SEIDMAN *et al.* (Editors), 1987 *Current Protocols in Molecular Biology.* Wiley Interscience, New York.

BARRIO, E., A. LATORRE and A. MOYA, 1994 Phylogeny of the *Drosoph-*

ila obscura species group deduced from mitochondrial DNA sequence. J. Mol. Evol. **39:** 478–488.

BECKENBACH, A. T., Y. W. WEI and H. LIU, 1993 Relationships in the *Drosophila obscura* species group, inferred from mitochondrial cytochrome oxidase II sequences. Mol. Biol. Evol. **10:** 619–634.

BERRY, A. J., J. W. AJIOKA and M. KREITMAN, 1991 Lack of polymorphism on the *Drosophila* fourth chromosome resulting from selection. Genetics **129:** 1111–1117.

BEVERLEY, S. M., and A. C. WILSON, 1984 Molecular evolution in *Drosophila* and the higher Diptera II. A time scale for fly evolution. J. Mol. Evol. **21:** 1–13.

BEWLEY, G. C., J. L. COOK, S. KUSAKABE, T. MUKAI, D. L. RIGBY *et al.*, 1989 Sequence, structure and evolution of the gene coding for *sn*-glycerol-3-phosphate dehydrogenase in *Drosophila melanogaster.* Nucleic Acids Res. **17:** 8553–8567.

BUZZATI-TRAVERSO, A. A., and R. E. SCOSSIROLI, 1955 The "*obscura*" group of the genus *Drosophila.* Adv. Genet. **7:** 47–92.

COYNE, J. A., W. F. EANES, J. A. RAMSHAW and R. K. KOEHN, 1979 Electrophoretic heterogeneity of a-glycerophosphate dehydrogenase among many species of *Drosophila.* Syst. Zool. **28:**164–175.

DE STORDEUR, E., and S. PASTEUR, 1978 Sur l'evolution de l'enzyme a-*Gpdh* chez les Drosophilidés. C. R. Acad. Sc. Paris **287:** 93–95.

DESALLE, R., 1992 The phylogenetic relationships of flies in the family Drosophilidae deduced from mtDNA sequences. Mol. Phylogenet. Evol. **1:** 31–40.

FELSENSTEIN, J., 1985 Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39:** 783–791.

GODDARD, K., A. CACCONE and J. R. POWELL, 1990 Evolutionary implications of DNA divergence in the *Drosophila obscura* group. Evolution **44:** 1656–1670.

GRIMALDI, D. A., 1990 A phylogenetic, revised classification of genera in the Drosophilidae (Diptera). Bull. Am. Mus. Nat. Hist. **197:** 1–139.

HIGUCHI, R. G., and H. OCHMAN, 1989 Production of single-stranded DNA templates by exonuclease digestion following the polymerase chain reaction. Nucleic Acids Res. **17:** 5865.

HUDSON, R. R., M. KREITMAN and M. AGUADÉ, 1987 A test of neutral molecular evolution based on nucleotide data. Genetics **116:** 153–159.

INA, Y., 1995 New methods for estimating the number of synonymous and nonsynonymous substitutions. J. Mol. Evol. **40:** 190–226.

JUKES, T. H., and C. R. CANTOR, 1969 Evolution of protein molecules, pp. 21–132 in *Mammalian Protein Metabolism,* edited by H. N. MUNRO. Academic Press, New York.

KEITH, T. P. 1983 Frequency distribution of esterase-5 alleles in two populations of *Drosophila pseudoobscura.* Genetics **105:** 135–155.

KIMURA, M., 1983 *The Neutral Theory of Molecular Evolution.* Cambridge University Press, New York.

KOSKE, T., 1953 Artkreuzungsveruche in der *obscura*-Gruppe der Gattung *Drosophila.* Z. Indukt. Abstamm. Verebungsl. **85:** 373–381.

KREITMAN, M., 1983 Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster.* Nature **304:** 412–417.

KUMAR, S., K. TAMURA and M. NEI, 1993 *MEGA: Molecular Evolutionary Genetics Analysis,* version 1.0. The Pennsylvania State University, University Park, PA.

KWIATOWSKI, J., D. SKAECKY, K. BAILEY and J. AYALA, 1994 Phylogeny of *Drosophila* and related genera inferred from the nucleotide sequence of the Cu, Zn *Sod* gene. J. Mol. Evol. **38:** 443–454.

LAKOVAARA, S., and L. KERÄNEN, 1980 Variation at the a-*Gpdh* locus of Drosophilids. Hereditas **92:** 251–258.

LAKOVAARA, S., and A. SAURA, 1982 Evolution and speciation in the *Drosophila obscura* group, pp. 1–59 in *The Genetics and Biology of Drosophila,* Vol. 3b, edited by M. ASHBURNER, H. L. CARSON and J. N. THOMPSON. Academic Press, London.

LAKOVAARA, S., A. SAURA and P. LANKINEN, 1977 Evolution at the a-*Gpdh* locus in Drosophilidae. Evolution **31:** 319–330.

LATORRE, A., E. BARRIO, A. MOYA and F. J. AYALA, 1988 Mitochondrial DNA evolution in the *Drosophila obscura* group. Mol. Biol. Evol. **5:** 717–728.

LEWONTIN, R. C., 1991 Twenty-five years ago in *Genetics:* electrophoresis in the development of evolutionary genetics: milestone of millstone? Genetics **128:** 657–662.

LEWONTIN, R. C., J. A. MOORE, W. PROVINE and B. WALLACE, 1981 *Dobzhansky's Genetics of Natural Populations.* Columbia University Press, New York.

LI, W.-H., 1977 Distribution of nucleotide differences between two randomly chosen cistrons in a finite population. Genetics **85:** 331–337.

LI, W.-H., C.-I. WU and C.-C. LUO, 1985 A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon usage. Mol. Biol. Evol. **2:** 150–174.

MADDISON, W. P., and D. R. MADDISON, 1992 *MacClade: Analysis of Phylogeny and Character Evolution,* version 3.0. Sinauer Associates, Sunderland, MA.

MCDONALD, J. H., and M. KREITMAN, 1991 Adaptive evolution at the *Adh* locus in *Drosophila.* Nature **351:** 652–654.

OAKESHOTT, J. G., J. B. GIBSON, P. R. ANDERSON, W. R. KNIBB, D. G. ANDERSON *et al.*, 1982 Alcohol dehydrogenase and glycerol-3-phosphate dehydrogenase clines in *Drosophila melanogaster* on different continents. Evolution **36:** 86–96.

OAKESHOTT, J. G., S. W. MCKECHNIE and G. K. CHAMBERS, 1984 Population genetics of the metabolically related *Adh, Gpdh* and *Tpi* polymorphisms in *Drosophila melanogaster.* I. Geographic variation in *Gpdh* and *Tpi* allele frequencies in different continents. Genetica **63:** 21–29.

PERUTZ, M. F., 1984 Species adaptation in a protein molecule. Mol. Biol. Evol. **1:** 1–28.

POWELL, J. R., 1976 Protein variation in natural populations of animals. Evol. Biol. **8:** 79–119.

RICHARD, I., and J. S. BECKMAN, 1995 How neutral are synonymous codon mutations? Nature Genet. **10:** 259.

RILEY, M. A., 1989 Nucleotide sequence of the *Xdh* region in *Drosophila pseudoobscura* and an analysis of the evolution of synonomous codons. Mol. Biol. Evol. **6:** 33–52.

RILEY, M. A., M. E. HALLAS and R. C. LEWONTIN, 1989 Distinguishing the forces controlling genetic variation at the *Xdh* locus in *Drosophila pseudoobscura.* Genetics **123:** 359–369.

RILEY, M. A., S. R. KAPLAN and M. VEUILLE, 1992 Nucleotide polymorphism at the xanthine dehydrogenase locus in *Drosophila pseudoobscura.* Mol. Biol. Evol. **9:** 56–69.

SAITOU, N., and M. NEI, 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:** 406–425.

SCHAEFFER, S., and E. L. MILLER, 1992a Molecular population genetics of an electrophoretically monomorphic protein in the alcohol dehydrogenase region of *Drosophila pseudoobscura.* Genetics **132:** 163–178.

SCHAEFFER, S. W., and E. L. MILLER, 1992b Estimates of gene flow in *Drosophila pseudoobscura* determined from nucleotide sequence analysis of the alcohol dehydrogenase region. Genetics **132:** 471–480.

SCHAEFFER, S. W., and E. L. MILLER, 1993 Estimates of linkage disequilibrium and the recombination parameter determined from segregating nucleotide sites in the alcohol dehydrogenase region of *Drosophila pseudoobscura.* Genetics **135:** 541–552.

STROBECK, C., 1983 Expected linkage disequilibrium for a neutral locus linked to a chromosomal arrangement. Genetics **103:** 545–555.

STURTEVANT, A. H., 1939 On the subdivision of the genus *Drosophila.* Proc. Natl. Acad. Sci. USA **3:** 137–141.

STURTEVANT, A. H., 1942 The classification of the genus *Drosophila* with description of nine new species. Univ. Texas Publ. **4213:** 5–51.

SWOFFORD, D. L., 1991 *PAUP: Phylogenetic Analysis Using Parsimony,* version 3.0S. Illinois Natural History Survey, Champaign, IL.

TAJIMA, F., 1993 Simple methods for testing the molecular evolutionary clock hypothesis. Genetics **135:** 599–607.

TAKANO, T. S., S. KUSAKABE and T. MUKAI, 1993 DNA polymorphism and the origin of protein polymorphism at the *Gpdh* locus of *Drosophila melanogaster,* pp. 179–190 in *Mechanisms of Molecular Evolution,* edited by N. TAKAHATA and A. G. CLARK. Sinauer, Sunderland, MA.

TAMURA, K., and M. NEI, 1993 Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. **10:** 512–526.

THROCKMORTON, L. H., 1975 The phylogeny, ecology and geography of *Drosophila,* pp. 421–469 in *Handbook of Genetics,* vol. 3, edited by R. C. KING. Plenum, New York.

TOMINAGA, H., T. SHIBA and S. NARISE, 1992 Structure of *Drosophila virilis* glycerol-3-phosphate dehydrogenase gene and a comparison with the *Drosophila melanogaster* gene. Biochim. Biophys. Acta **1131:** 233–238.

VILLAROYA, A., and E. JUAN, 1991 ADH and phylogenetic relationships of *Drosophila lebanonensis*. J. Mol. Evol. **32:** 421–428.

VON KALM, L. J., J. WEAVER, J. DEMARCO, R. J. MACINTYRE and D. T. SULLIVAN, 1989 Structural characterization of the α-glycerol-3-phosphate dehydrogenase-encoding gene of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **86:** 5020–5024.

WELLS, R. S., 1995 Sequence and evolution of the *Drosophila pseudoobscura* glycerol-3-phosphate dehydrogenase locus. J. Mol. Evol. **41:** 886–893.

WELLS, R. S., 1996 Excessive homoplasy in an evolutionarily constrained protein. Proc. Roy. Soc. Lond., Ser. B (in press).

Communicating editor: A. G. CLARK