Molecular Drift of the *Bride of Sevenless* (*boss*) Gene in *Drosophila*[1]

Francisco José Ayala and Daniel L. Hartl
Department of Organismic and Evolutionary Biology, Harvard University

DNA sequences were determined for three to five alleles of the *bride-of-sevenless* (*boss*) gene in each of four species of *Drosophila*. The product of *boss* is a transmembrane receptor for a ligand coded by the *sevenless* gene that triggers differentiation of the R7 photoreceptor cell in the compound eye. Population parameters affecting the rate and pattern of molecular evolution of *boss* were estimated from the multinomial configurations of nucleotide polymorphisms of synonymous codons. The time of divergence between *D. melanogaster* and *D. simulans* was estimated as ~1 Myr, that between *D. teissieri* and *D. yakuba* as ~0.75 Myr, and that between the two pairs of sibling species as ~2 Myr. (The *boss* genes themselves have estimated divergence times ~50% greater than the species divergence times.) The effective size of the species was estimated as $\sim 5 \times 10^6$, and the average mutation rate was estimated as $1\text{–}2 \times 10^{-9}$/nucleotide/generation. The ratio of amino acid polymorphisms within species to fixed differences between species suggests that ~25% of all possible single-step amino acid replacements in the *boss* gene product may be selectively neutral or nearly neutral. The data also imply that random genetic drift has been responsible for virtually all of the observed differences in the portion of the *boss* gene analyzed among the four species.

Introduction

Variation in the nucleotide sequences and relative frequencies of alleles reflects the combined expression of evolutionary forces in natural populations. Variation in the nucleotide sequence results from processes such as mutation, recombination, and gene conversion; variation in allele frequency results from such processes as migration, random genetic drift, and natural selection. In principle, it should be possible to infer the major evolutionary processes and to estimate the relevant parameters on the basis of allele-frequency data from natural populations. With assays based on electrophoretic variation in proteins—for a long time the primary source of data about molecular and allele-frequency variation in natural populations—the problem proved to be insurmountable because of empirical insufficiency (Lewontin 1974, pp. 189–271, 1991), in part owing to the inability to resolve molecular differences among alleles with the same electrophoretic mobility and in part owing to virtually all electrophoretic studies being "snapshots" of the genetic variation present in a particular population at a particular time and therefore lacking a vital time dimension. Nevertheless, protein electrophoresis contributed a great deal of information about the levels of genetic variation in natural populations, as well as about population structure and genetic differentiation within and among species (Ayala 1976; Nei 1987, pp. 176–253). During

1. Key words: molecular drift, *bride of sevenless, Drosophila,* polymorphisms, divergence times.

Address for correspondence and reprints: Daniel L. Hartl, Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138.

Mol. Biol. Evol. 10(5):1030–1040. 1993.

0737-4038/93/1005-0007$02.00

the same period, major theoretical advances were made in developing the neutral theory of molecular evolution (Ewens 1979, pp. 251–281; Kimura 1983, pp. 194–304; Nei 1987, pp. 327–403), including a sampling theory for interpreting genetic variation found in natural populations (Ewens 1972).

The advent of DNA sequencing, particularly techniques based on the polymerase chain reaction and therefore applicable to multiple alleles, made the identification of alleles unambiguous. DNA sequence data are also rich in information, because, when there is linkage equilibrium, each nucleotide site can be regarded as an independent realization of the stochastic evolutionary process governing the entire gene, with due regard for functional differences among the sites. For example, the multinomial configurations of twofold or fourfold synonymous nucleotide sites are realizations of a process of mutation and random genetic drift in which the differences between synonymous substitutions may be regarded as selectively neutral or nearly neutral (or at least subject to much weaker selective constraints than are amino acid replacements). Likewise, the multinomial configurations of amino acid replacements are realizations of a process of mutation and random genetic drift with potentially important contributions from natural selection. Some natural populations are sufficiently polymorphic to enable a thoroughgoing analysis and estimation of key parameters. For example, synonymous and replacement nucleotide polymorphisms in the *6-phosphogluconate dehydrogenase* gene in *Escherichia coli* permitted estimates of mutation rates and effective population size and also suggested that most of the amino acid polymorphisms are slightly detrimental to fitness (Sawyer et al. 1987; Hartl and Sawyer 1991).

Most eukaryotic populations are not sufficiently polymorphic to enable strong inferences from "snapshot" data, without additional information concerning how the alleles and their frequencies change with time. Longitudinal studies of sufficient duration are often impractical, and studies of geographically isolated subpopulations are often difficult to interpret because the rates of migration are unknown. An innovative alternative to study multiple individuals in closely related species was pioneered by McDonald and Kreitman (1991*a*), with the *alcohol dehydrogenase* (*Adh*) gene in *Drosophila.* The studies were carried out to ascertain whether natural selection could be implicated in amino acid replacements between closely related species. Any nucleotide difference may be polymorphic within species or a fixed difference between species (i.e., monomorphic within species). The ratio of polymorphisms to fixed differences calculated for synonymous nucleotide sites provides a standard for comparison with the same ratio calculated for amino acid replacements. In the *Adh* data, the number of fixed differences among amino acid replacements was too large, relative to the number of amino acid polymorphisms, and hence natural selection was inferred. The multinomial configurations of polymorphic nucleotides within and among species can also be analyzed to yield estimates of a number of key parameters of population structure, including nucleotide mutation rates, effective population number, species divergence time, and the number of single-step amino acid replacements that are likely to be either beneficial or else selectively nearly neutral (Sawyer and Hartl 1992).

In this paper, we estimate rates of nucleotide polymorphism and divergence in the *bride-of-sevenless* (*boss*) gene in *D. melanogaster* and three closely related species. The *boss* gene was chosen to contrast with *Adh:* whereas *Adh* codes for an enzyme in intermediary metabolism, *boss* is a developmental control gene. In particular, the *boss* gene product is a membrane-bound receptor whose only known function is to interact specifically with the product of the *sevenless* gene to trigger differentiation of the R7 photoreceptor cell in the compound eye (Krämer et al. 1991; Rubin 1991; Van Vactor

et al. 1991). In the case of *boss,* the population data suggest a recent evolutionary history dominated by the effects of mutation and random genetic drift.

Material and Methods

Drosophila Strains

We studied isofemale lines established from various collections. One isofemale line of *Drosophila melanogaster* was studied from each of five collections representing Brazzaville (Congo), Cotonou (Benin), Harwich (United Kingdom), Lausanne (Switzerland), and St. Louis; one isofemale line of *D. simulans* was studied from each of five collections representing Capetown, Brazzaville (Congo), Le Cap (Haiti), Morrow Bay (California), and St. Louis; three isofemale lines of *D. teissieri* were studied from collections made in Brazzaville (Congo); and four isofemale lines of *D. yakuba* were studied from collections made in the Ivory Coast. The isofemale lines from Africa and Haiti were provided by Jean David and Pierre Capy. The *boss* gene was amplified and sequenced from one individual chosen at random from each of the isofemale lines. The published sequence of *boss* from *D. melanogaster* (Hart et al. 1990) was also included in the data analysis.

Amplification of *boss* DNA

PCR and sequencing primers were initially designed according to the published sequence of *boss* from *D. melanogaster* (Hart et al. 1990). DNA was extracted according to the protocol of Gloor and Engels (1992). The primers used for amplification were 5′-TCCGCGTTGTTCGTAATCAA-3′ and 5′-GCGGGATTCACGCACTCGTA-3′, which correspond, respectively, to nucleotides 632–651 and 2631–2612 in the published sequence (Hart et al. 1990). Subsequent sequencing was carried out with a primer-walking approach based on knowledge of the nucleotide sequences from all the strains. After amplification (Saiki et al. 1988), the PCR products were purified by diluting the 30-μl reaction volume to a final volume of 100 μl with H_2O, followed by phenol extraction, precipitation in the presence of 2.5 M NH_4OAc and 1 volume of 100% ethanol, and washing with 70% ethanol (Sambrook et al. 1989, pp. E.10–E.13).

DNA Sequencing

Sequencing of all the PCR products was performed with an Applied Biosystems model 373A DNA sequencing system and the *Taq* DyeDeoxy™ terminator cycle-sequencing kit (Halloran et al. 1993; recommendations of the manufacturer). Organic contaminants were removed with the phenol-chloroform extraction protocol provided with the sequencing kit. Nucleotide sites 760–2325 in *boss,* numbered as in Hart et al. (1990), were sequenced in both strands in the PCR product from single flies, one from each of the 17 isofemale lines.

Sequence Analysis

Sequences were analyzed with the Lineup, Plotsimilarity, and Codonfrequency programs from the Genetics Computer Group (GCG) sequence analysis software package (Devereux et al. 1984) and also with software written by Stanley Sawyer that implements the estimation procedures outlined by Sawyer and Hartl (1992). Nucleotide substitutions from the aligned DNA sequences were tabulated as described by McDonald and Kreitman (1991*a*), with the following exception: a site was considered to have undergone at least as many mutations as the number of species found

to be polymorphic at the site, because each of the within-species groupings has a single common ancestor. The phylogenetic relationships among the four *Drosophila* species examined are well established and consistent across a variety of characters (Lachaise et al. 1988). The species define two phylads, one including *D. melanogaster* and *D. simulans*, the other including *D. teissieri* and *D. yakuba*. Because of the topology of the species tree, a site that is a fixed difference between *D. melanogaster* and *D. simulans* and also between *D. teissieri* and *D. yakuba* must have undergone at least two mutations in the between-species branches, if it is assumed that the locus was not polymorphic at the time of the common ancestor of all four taxa.

Results

The sequenced region of *boss* comprises a large portion of the protein-coding region, including ~60% of the extracellular domain and ~75% of the transmembrane domain. However, the sequenced region does not include any of the cytoplasmic domain of the *boss* protein. The PCR product also includes an intron of ~100 bp, which was not sequenced except for regions around the borders.

The variable nucleotide positions from the sequenced region of the *boss* gene are presented in figure 1. The number of nucleotide differences, classified according to their effect on amino acid sequence (replacement vs. silent) and according to their status in the population (fixed difference between species vs. polymorphic difference within species) are presented under the columns labeled "Overall" in table 1. A *G*-test of independence, with the Williams correction for continuity (Sokal and Rohlf 1981, pp. 737–738), indicates no significant deviation from the expectations of the neutral theory ($G = 0.031$, $P = 0.89$). In other words, the ratio of polymorphisms to fixed differences was found to be virtually the same for both replacement and synonymous nucleotide substitutions. Subdivision of the *boss* sequence into the extracellular and transmembrane regions also shows nonsignificance (table 1). Subsets of the data with one or more alleles removed are also nonsignificant (data not shown).

Under the assumption of selective neutrality, the ratio of observed replacement to observed synonymous substitutions, for both fixed differences and polymorphisms, should equal the ratio of total possible neutral replacement substitutions to total possible neutral synonymous substitutions (McDonald and Kreitman 1991*a*; Sawyer and Hartl 1992). For the sequenced region of the *boss* gene, the ratio is ~(8 + 13):(71 + 106) = 1:8.43 (table 1). The total number of possible synonymous substitutions in the sequenced region of *boss* is 1,043. If it is assumed that all synonymous substitutions are selectively neutral, this calculation implies that in the sequenced region of *boss*, which codes for 522 amino acids, there are ~1,043/8.43 = 124 possible amino acid replacements that are selectively neutral or nearly neutral. The binomial standard error on this estimate is ~24.

Comparison of the ratio of synonymous substitutions that are fixed to those that are polymorphic reveals similar patterns of synonymous-site evolution for both the *boss* gene and *Adh* (table 2). (In this comparison, *Drosophila teissieri* was excluded because corresponding data from *Adh* are unavailable.) The average "scaled χ^2 values," which measure codon-usage bias (Shields et al. 1988), for *D. melanogaster, D. simulans, D. teissieri,* and *D. yakuba* are 0.32, 0.35, 0.55, and 0.45, respectively. These values are roughly in the lower-middle range (0.05–1.05) of scaled χ^2 calculated for a wide variety of *Drosophila* genes (Shields et al. 1988).

Analysis of the joint multinomial configurations of nucleotides at silent sites for the *boss* locus provides estimates of species divergence times (t_{div}), gene divergence

Site	mel abcde	sim fghij	yak klmn	tei opq	type
762 T	.....	.....	....	..A	PR
778 G	.....	A.A..	....	...	PS
781 G	.....	....A	....	...	PS
784 C	.....	....T	....	...	PS
802 T	.....	CG.GG	GGGG	GGG	PS,PS,FS
830 T	.....	.....	....	CCC	FR
850 A	.....	.....	....	..G	PS
862 G	.....	.....	..S.	...	PR
873 G	...C.	.....	....	...	PR
877 G	.....	.....	TTTT	...	FS
886 C	TT...	.....	TTTT	TTT	PS,FS
889 C	..AAA	.....	....	...	PS
898 A	.....	.....	GGGG	GGG	FS
919 A	.....	.....	....	.G.	PS
934 A	.....	.....	....	G.G	PS
967 G	.....	A.A..	....	...	PS
997 C	.....	.T...	....	...	PS
1033 A	.....	GGGGG	GGGG	GGG	FS
1042 T	.....	GGGGG	GGGG	GGG	FR
1066 T	.....	.....	CCCC	CCC	FS
1072 A	.....	...G.	GGGG	CGG	PS,FS,PS
1096 C	.....	.....	S...	...	PS
1100 C	.....	.....	..T.	...	PS
1102 G	.....	.....	.C..	...	PS
1120 T	.....	.....	....	CCC	FS
1126 G	.....	.....	AAAA	AAA	FS
1135 A	.....	.T...	....	...	PR
1138 T	.....	.....	....	CCC	FS
1148 T	.....	.....	CCCC	CCC	FS
1162 T	.....	.....	..C.	C.C	PS,PS
1168 C	.....	T.T..	....	...	PS
1169 C	.....	T.T..	....	...	PS
1171 A	.....	GGGGG	GGGG	GGG	FS
1186 T	.....	.C...	CCCC	.CC	PS,FS,PS
1189 C	.....	.....	....	T..	PS
1192 G	.....	A.A.A	....	...	PS
1198 C	.....	TTTTT	TTTT	...	FS,FS
1201 C	.....	....T	....	...	PS
1203 T	.....	.....	....	.C.	PR
1204 T	.....	.....	GGGG	GGG	FS
1228 T	.....	.....	.G..	C..	PS,PS
1229 T	.C...	CCCCC	CCCC	CCC	PS
1234 C	.....	T.T..	....	...	PS
1246 C	.....	.....	.TT.	.T.	PS,PS
1255 T	.....	.....	....	.C.	PS
1259 T	.....	.....	GAGG	GGG	PR,FR
1261 A	.....	.....	GGGG	GGG	FS
1270 A	.....	.....	....	.C.	PS
1276 G	.....	.....	CCCC	CCC	FS
1282 T	.....	.....	CCCC	CCC	FS
1285 C	.....	..GG.	....	...	PS
1294 A	.....	.....	....	CCC	FS
1297 C	.....	A....	....	...	PS
1318 C	.....	.....	....	.G.	PS
1324 C	.....	....T	....	...	PS

Site	mel abcde	sim fghij	yak klmn	tei opq	type
1339 T	.....	.....	....	C..	PS
1342 G	.....	A....	....	...	PS
1345 T	.....	.....	AAAA	GAA	FS,PS
1348 T	.....	.....	C...	CCC	PS,FS
1355 C	.....	.....	....	AAA	FR
1369 G	.....	.C..C	....	...	PS
1372 G	.....	.C..C	....	...	PS
1381 T	C..CC	CCCCC	CCCC	CCC	PS
1384 C	...TT	.....	....	...	PS
1387 C	T....	.....	....	...	PS
1393 G	.....	C..C.	....	...	PS
1399 C	.....	.....	T...	...	PS
1414 A	G....	GGGGG	GGGG	GGG	PS
1417 C	.....	.AA.A	AAAA	AAA	PS,FS
1426 C	T....	.....	....	...	PS
1432 A	.....	.....	....	R..	PS
1441 T	.....	.....	....	GGG	FR
1453 C	.....	.....	....	.A.	PS
1456 G	.....	.A.AA	....	...	PS
1466 C	.....	T....	....	...	PS
1468 G	.....	.....	....	.A.	PS
1478 T	.....	.....	CCCC	CCC	FS
1482 G	.....	.....	....	..C	PR
1506 A	.....	.....	....	.T.	PR
1516 A	.....	.....	....	GGG	FS
1520 T	.....	CCCC.	CCCC	CCC	PS,FS
1527 T	...A.	.....	....	...	PR
1540 A	.R...	GGGGG	GGGG	GGG	PS
1549 G	.....	.....	AAAA	...	FS
1552 T	.....	.....	....	CCC	FS
1558 A	.....	.....	....	..G	PS
1564 C	.....	.....	TTTT	TTT	FS
1609 T	.....	.....	.C..	YCC	PS,PS
1612 C	.....	.....	.T..	...	PS
1615 T	.....	.....	CCCC	CCC	FS
1621 G	.....	.....	.C.C	...	PS
1636 T	.....	C.CCC	....	C.C	PS,PS
1637 T	.....	.....	..C.	...	PS
1651 C	.....	.....	AG.G	...	PS,PS
1657 T	.....	CCCCC	C.C.	CCC	FS,PS
1666 G	.....	.....	T...	...	PS
1678 G	AR...	.....	....	...	PS
1687 T	.....	.....	CCCC	CCC	FS
1690 T	.....	.....	....	K..	PS
1700 A	.....	CCCCC	CCCC	CCC	FR
1702 A	.....	GGGGG	GGGG	GGG	FS
1708 A	.....	GG..G	....	...	PS
1714 T	.G...	GGGGG	GGGG	GGG	PS
1723 T	.....	CCCCC	CCCC	CCC	FS
1726 A	.....	...G.	....	GGG	PS,FS
1732 T	GG...	GGGGG	GGGG	GGG	PS
1769 G	AAAAA	AAAAA	AAAA	AAA	PR
1771 T	.....	.....	....	KGG	PS
1801 T	.....	....A	....	...	PS
1810 C	.....	.GGG.	GGGG	GGG	FS,PS

times (t_{gene}), mutation rates to each nucleotide (α_j; j = A, T, G, and C), and aggregate mutation rate at silent sites (μ_s), all scaled according to the haploid effective population size (N_e) (Sawyer and Hartl 1992). These estimates are shown in table 3. In table 4, the estimates of t_{div} and t_{gene} have been converted to expressions scaled in millions of years by the following method: On the basis of comparisons of *Adh* between *D. simulans*

Site	mel abcde	sim fghij	yak klmn	tei opq	type
1838	A	.....	G...	...	PR
1849	C	A....	....	...	PS
1852	T CC.C.	CCCCC	CCCC	CCC	PS
1855	C	..A..	....	...	PS
1858	C .Y.T.	.....	....	...	PS
1864	A	.....	CCCC	CCC	FS
1870	G	CCCCC	....	...	FS
1871	A	GGGGG	GGGG	GGG	FR
1888	T	.....	CCCC	CCC	FS
1891	A	.....	....	GGG	FS
1900	C	.A...	....	...	PS
1903	C	....T	....	...	PS
1904	T	.....	CCCC	CCC	FS
1906	A	GGGGG	GGGG	..G	FS,PS
1909	G	CCCCC	CCCC	CCC	FS
1912	T	..C..	CCCC	CCC	PS,FS
1929	A	TTTTT	TTTT	TTT	FR
1960	T	.....	....	Y..	PS
1990	T	.....	....	Y..	PS
1999	C	.....	T...	...	PS
2005	C	.....	TTTT	TTT	FS
2011	G	.....	TTTT	TTT	FS
2029	T	.....	CCCC	CCC	FS
2038	C	.....	TTTT	TTT	FS
2047	T	.....	GGGG	GGG	FS
2050	T	.....	....	C..	PS
2053	C	.....	..A.	...	PS
2059	C	..TT.	....	...	PS
2062	C	AAAAA	....	...	FS
2068	C	...T.	....	...	PS

Site	mel abcde	sim fghij	yak klmn	tei opq	type
2078	T	.....	CCCC	...	FS
2086	A	TTTTT	GGGG	GGG	FS,FS
2092	T	CCCCC	CCCC	CCC	FS
2110	A	.....	....	GGG	FS
2116	T	.....	CCCC	CCC	FS
2125	A	.....	CCCC	CCC	FS
2128	T	.....	....	YCC	PS
2131	T	CCCCC	CCCC	CCC	FS
2137	A	.....	GGGG	GGG	FS
2138	G	.....	C...	...	PR
2146	C	.....	TTTT	TTT	FS
2152	T	.....	CCCC	CCC	FS
2158	T	.....	....	G..	PS
2167	A	G....	....	...	PS
2194	G	.....	CCCC	...	FS
2218	T	.....	.CCC	CCC	PS,FS
2227	T	.....	CCCC	CCC	FS
2233	T	.....	....	..C	PS
2236	T	.....	....	CCC	FS
2245	T	.....	CCCC	CCC	FS
2248	A GG...	.....	TTTT	TTT	PS,FS
2254	G	...AA	....	...	PS
2260	A	CCCCC	CCCC	CCC	FS
2263	G	.....	AAAA	AA.	FS,PS
2280	C	....T	....	...	PR
2281	T	.....	GGGG	GGG	FS
2284	A	.....	GGGG	GGG	FS
2285	C	.....	TT.T	...	PS
2317	A	.....	....	..C	PS
2320	C	.....	TT..	...	PS

FIG. 1.—*Drosophila melanogaster boss* gene. The first column contains the position numbers and corresponding nucleotides from the *D. melanogaster boss* locus reported by Hart et al. (1990). Strains are identified with species names (*mel* = *D. melanogaster; sim* = *D. simulans; yak* = *D. yakuba;* and *tei* = *D. teissieri*) and collection sites (a, o, p, and q = Brazzaville; b = Cotonou; c = Harwich; d = Lausanne; e and j = St. Louis; f = Capetown; g = Congo; h = Le Cap; i = Morrow Bay; and k, l, m, and n = Ivory Coast). Nucleotides identical to the published sequence are shown as periods (.). Heterozygous sites are represented as follows: R = A or G; Y = C or T; K = G or T; and S = G or C. The last column lists the types of substitutions observed at each site: FR = fixed replacement; FS = fixed synonymous; PR = polymorphic replacement; and PS = polymorphic synonymous.

and *D. yakuba,* Sawyer and Hartl (1992) estimated that N_e generations is equivalent to 0.645 Myr, using the values that Rowan and Hunt (1991) calculated for species divergence times among Hawaiian *Drosophila,* which were calibrated according to the times of formation of the islands on which the species are found. The mutation rate per million years at silent sites (μ) at the *boss* locus was estimated by dividing μ_s for *boss* from the *D. simulans*–versus–*D. yakuba* comparison by the number of regular third-position sites and then dividing again by 0.645 Myr: hence, $\mu = (5.32/395)/0.645 = 0.02$. [A regular third-position site is defined as any twofold- or fourfold-degenerate third-position site in a codon that codes for a monomorphic amino acid other than leucine or arginine; threefold-degenerate sites are considered to be regular (twofold degenerate) if the variant codons do not include ATA (Sawyer and Hartl 1992).] For the remaining species comparisons, the multiple of N_e generations equivalent to 1 Myr is estimated as μ_s divided by the number of regular third-position sites divided by μ. Divergence times, in millions of years, were then estimated by multiplying t_{div} and t_{gene} by the multiple of N_e generations per million years. Similarly, the mutation

Table 1
Number of Replacement and Synonymous Nucleotide Substitutions in the *boss* Gene That Are Either Fixed between Species or Polymorphic within Species

	Overall		Extracellular		Transmembrane	
	Fixed	Polymorphic	Fixed	Polymorphic	Fixed	Polymorphic
Replacement ...	8	13	6	9	2	4
Synonymous ...	71	106	35	77	36	29
G-value	0.031		0.432		1.000	
P	0.89		0.52		0.40	

rates to individual nucleotides per generation (μ_j) were estimated by dividing α_j (table 3) by $2N_e$. (The value of N_e was derived from the estimate of N_e generations per million years, assuming 10 generations per year.) The estimates of species divergence times in table 4 add weight to the estimates of Lemeunier et al. (1986), who date the *D. melanogaster*–versus–*D. simulans* split as 0.8–3 Mya, and those of Caccone et al. (1988), who estimate the age of the common ancestor of all four species to be 1.6–6.1 Mya.

Discussion

The *Adh* gene has long been a model for studies of genetic variation in *Drosophila,* and much evidence already exists for positive natural selection acting on this gene (Lewontin 1985; Hudson et al. 1987; Vouidibio et al. 1989). The sequencing results that McDonald and Kreitman (1991*a*) had with *Adh,* supporting its predominantly adaptive evolution, prompted this application of the same statistical test to another locus, *bride-of-sevenless* (*boss*), to contrast with *Adh.* Because *boss* is a developmental gene expressed during the third larval instar (Hart et al. 1990), it is presumably less likely than *Adh* to evolve in response to substrates present in the external environment. Unlike the situation with *Adh,* the patterns of nucleotide variation in the sequenced region of the *boss* gene are consistent with the expectations of the neutral theory. This conclusion is based on a sequenced region that includes ~60% of the extracellular domain and ~75% of the transmembrane domain. The results do not preclude the possibility of adaptive evolution in unsequenced regions of these domains or in the cytoplasmic domain.

Our conclusions are based on a statistical test suggested by McDonald and Kreitman (1991*a*). The test examines the homogeneity of entries in a 2×2 contingency table, and it has elicited considerable controversy concerning its validity and interpretation (see Graur and Li 1991; McDonald and Kreitman 1991*b;* Whittam and Nei 1991). A detailed theoretical foundation for the use of such a 2×2 contingency table has recently been provided (Sawyer and Hartl 1992). The homogeneity test appears to be very robust in that many evolutionary perturbations not resulting from selection—e.g., fluctuations in population size, hitchhiking, or gene conversion—should affect both replacement and synonymous sites equally and therefore may alter the number of fixed and polymorphic differences but not their ratios. Therefore, departures from the expectations of the neutral theory provide strong evidence for adaptive evolution. That the results from *Adh* and *boss* are different confirms that the data are not artifacts attributable to some population phenomenon such as fluctuation in population size

Table 2
Comparison of *boss* and *Adh*, by the Number of Synonymous Nucleotides That Are Either Fixed between Species or Polymorphic within Species

	Fixed	Polymorphic
boss	60	75
Adh[a]	20	44
G-value	3.171	
P	0.08	

[a] Tabulated from data of McDonald and Kreitman (1991*a*).

or mutational saturation of silent sites, because these processes would be expected to affect both genes to the same extent.

Although *Adh* and *boss* are similar in the ratios of fixed to polymorphic substitutions at synonymous sites, any differences would not necessarily indicate selection at synonymous sites, because the ratios may be affected by hitchhiking or gene conversion occurring at only one of the two loci. Bias in codon usage among synonymous codons in *Drosophila* provides considerable evidence for natural selection among synonymous codons (Shields et al. 1988; Sharp and Li 1989; Moriyama and Hartl, accepted), but it is not clear whether the selection pressures for codon usage are strong enough—and whether statistical tests powerful enough—to yield a significant deviation from the neutral expectation in the ratio of fixed to polymorphic sites. In the particular case of *boss* versus *Adh,* differences in codon-usage bias among the four species are unlikely to have had a significant impact on the ratio of fixed to polymorphic synonymous substitutions, because codon-usage bias would proportionately affect the numbers of both classes.

The *boss* gene is required for induction of development of the photoreceptor neuron R7 (Cagan et al. 1992). Because the *boss* gene product functions only during retinal development, it is presumably less likely to be involved in adaptation to the external environment than is *Adh.* However, this difference does not necessarily explain why *Adh* appears to evolve adaptively by natural selection whereas *boss* does not. All genes within an organism must evolve in concert in response to both internal and external selective pressures. Changes in one gene may ultimately result in adaptive changes in many others. For example, the evolution of *boss* must be coordinated with that of *sevenless,* which codes for the ligand of the *boss* gene product (Cagan et al. 1992). Thus, the evolution of almost any gene may have manifold rippling effects. A sequence analysis of many different functional classes of genes, such as housekeeping genes, regulatory genes, or tissue-specific or developmental-stage-specific genes, should be of considerable interest in determining which classes of genes—and what proportion of genes—undergo selective modification. Population geneticists have also debated whether adaptive evolution is due mainly to many mutations with small effects or to few mutations with large effects (Nei 1987, pp. 414–416). Analyses of the nucleotide sequences of multiple alleles also provide estimates of selection coefficients and of the number of amino acids of a protein that are susceptible to favorable mutation (Sawyer and Hartl 1992), and such data may also contribute to the resolution of this issue.

Table 3

Pairwise Comparisons of t_{div}, t_{gene}, μ_s, Number of Regular Third-Position Sites (Reg), and α to Individual Nucleotides [α(T), α(C), α(A), and α(G)], All Scaled to N_e

Drosophila Species Compared	t_{div} (95% Confidence Interval[a])	t_{gene}	μ_s	Reg[b]	α(T)	α(C)	α(A)	α(G)
D. simulans vs.								
D. yakuba[c]	3.3 (2.2, 4.4)	4.2	5.3	395				
D. melanogaster vs.								
D. simulans	2.0 (1.2, 2.9)	2.9	4.0	392	0.010	0.016	0.006	0.014
D. teissieri vs. *D. yakuba*	1.2 (0.6, 1.7)	2.2	5.1	391	0.011	0.023	0.006	0.019
D. melanogaster/D. simulans vs.								
D. teissieri/D. yakuba	2.0 (1.4, 2.7)	3.0	8.3	382	0.020	0.035	0.012	0.031

[a] Estimated as by Sawyer and Hartl (1992).

[b] Defined as two-, three-, or fourfold-degenerate monomorphic silent sites, excluding ATA isoleucine codons (Sawyer and Hartl 1992).

[c] Entries for α's are absent because the *D. simulans*–versus–*D. yakuba* comparison was used for normalization with corresponding *Adh* data.

Table 4

Pairwise Comparisons of t_{div} and t_{gene}, Both in Millions of Years, N_e, and Mutation Rates to Individual Nucleotides per Generation (μ_T, μ_C, μ_G, and μ_A)

Drosophila Species Compared	t_{div} (95% Confidence Interval[a])	t_{gene}	N_e	μ_T	μ_C	μ_A	μ_G
D. melanogaster vs. *D. simulans* ...	0.99 (0.57, 1.40)	1.40	4.9×10^6	1.04×10^{-9}	1.61×10^{-9}	5.87×10^{-10}	1.41×10^{-9}
D. teissieri vs. *D. yakuba*	0.74 (0.38, 1.08)	1.40	6.2×10^6	9.07×10^{-10}	1.87×10^{-9}	5.13×10^{-10}	1.56×10^{-9}
D. melanogaster/D. simulans vs. *D. teissieri/D. yakuba*	2.12 (1.44, 2.80)	3.12	1.0×10^7	9.79×10^{-10}	1.70×10^{-9}	5.57×10^{-10}	1.48×10^{-9}

[a] Estimated as by Sawyer and Hartl (1992).

Acknowledgments

We thank Dan Krane, Stanley Sawyer, John Carulli, and Belinda Chang for comments and discussion, and we thank Etsuko Moriyama for analyzing the *boss* data by using her adaptation of the Sawyer software to the Macintosh. This work was supported by NIH grant GM40322.

LITERATURE CITED

AYALA, F. J., ed. 1976. Molecular evolution. Sinauer, Sunderland, Mass.

CACCONE, A., G. D. AMATO, and J. R. POWELL. 1988. Rates and patterns of scnDNA and mtDNA divergence within the *Drosophila melanogaster* subgroup. Genetics **118**:671–683.

CAGAN, R. L., H. KRÄMER, A. C. HART, and S. L. ZIPURSKY. 1992. The bride of sevenless and sevenless interaction: internalization of a transmembrane ligand. Cell **69**:393–399.

DEVEREUX, J., P. HAEBERLI, and O. SMITHIES. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–395.

EWENS, W. J. 1972. The sampling theory of selectively neutral alleles. Theor. Popul. Biol. **3**: 87–112.

———. 1979. Mathematical population genetics. Springer, New York.

GLOOR, G., and W. ENGELS. 1992. Single-fly DNA preps for PCR. Drosophila Inf. Serv. **71**: 148–149.

GRAUR, D., and W.-H. LI. 1991. Scientific correspondence. Nature **354**:114–115.

HALLORAN, N., Z. DU, and R. K. WILSON. 1993. Sequencing reactions for the Applied Biosystems 373A automated sequencer. *In* H. G. GRIFFIN and A. M. GRIFFIN, eds. DNA sequencing: laboratory protocols. Vol. **23**. Humana, Totowa, N.J.

HART, A. C., H. KRÄMER, D. L. VAN VACTOR, JR., M. PAIDHUNGAT, and S. L. ZIPURSKY, 1990. Induction of cell fate in the *Drosophila* retina: the bride of sevenless protein is predicted to contain a large extracellular domain and seven transmembrane segments. Genes Dev. **4**: 1835–1847.

HARTL, D. L., and S. A. SAWYER. 1991. Inference of selection and recombination from nucleotide sequence data. J. Evol. Biol. **4**:519–532.

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

KIMURA, M. 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge.

KRÄMER, H., R. C. CAGEN, and S. L. ZIPURSKY. 1991. Interaction of bride of sevenless membrane-bound ligand and the sevenless tyrosine-kinase receptor. Nature **352**:207–212.

LACHAISE, D., M. CARIOU, J. R. DAVID, F. LEMEUNIER, L. TSACAS, and M. ASHBURNER. 1988. Historical biogeography of the *Drosophila melanogaster* species subgroup. Evol. Biol. **22**: 159–227.

LEMEUNIER, F., J. R. DAVID, L. TSACAS, and M. ASHBURNER. 1986. The melanogaster species group. Pp. 147–256 *in* M. ASHBURNER and H. L. CARSON, eds. The genetics and biology of *Drosophila*. Academic Press, New York.

LEWONTIN, R. C. 1974. The genetic basis of evolutionary change. Columbia University Press, New York.

———. 1985. Population genetics. Annu. Rev. Genet. **19**:81–102.

———. 1991. Electrophoresis in the development of evolutionary genetics: milestone or millstone? Genetics **128**:657–662.

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

———. 1991*b*. Scientific correspondence. Nature **354**:653.

MORIYAMA, E. N., and D. L. HARTL. Codon usage bias and base composition of nuclear genes in Drosophila. Genetics (accepted).

NEI, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

ROWAN, R. G., and J. A. HUNT. 1991. Rates of DNA change and phylogeny from the DNA sequences of the alcohol dehydrogenase gene for five closely related species of Hawaiian *Drosophila.* Mol. Biol. Evol. **8**:49–70.

RUBIN, G. M. 1991. Signal transduction and the fate of the R7 photoreceptor in *Drosophila.* Trends Genet. **7**:372–377.

SAIKI, R. K., D. H. GELFAND, S. STOFFEL, S. J. SCHARF, R. G. HIGUCHI, G. T. HORN, K. B. MULLIS, and H. A. ERLICH. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239**:487–491.

SAMBROOK, J., E. F. FRITSCH, and T. MANIATIS. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

SAWYER, S. A., D. E. DYKHUIZEN, and D. L. HARTL. 1987. A confidence interval for the number of selectively neutral amino acid polymorphisms. Proc. Natl. Acad. Sci. USA **84**: 6225–6228.

SAWYER, S. A., and D. L. HARTL. 1992. Population genetics of polymorphism and divergence. Genetics **132**:1161–1176.

SHARP, P. M., and W.-H. LI. 1989. On the rate of DNA sequence evolution in *Drosophila.* J. Mol. Evol. **28**:398–402.

SHIELDS, D. C., P. M. SHARP, D. G. HIGGINS, and F. WRIGHT. 1988. "Silent" sites in *Drosophila* genes are not neutral: evidence of selection among synonymous codons. Mol. Biol. Evol. **5**: 704–716.

SOKAL, R. R., and F. J. ROHLF. 1981. Biometry. W. H. Freeman, San Francisco.

VAN VACTOR, D. L. JR., R. L. CAGAN, H. KRÄMER, and S. L. ZIPURSKY. 1991. Induction in the developing compound eye of *Drosophila:* multiple mechanisms restrict R7 induction to a single retinal precursor cell. Cell **67**:1145–1155.

VOUIDIBIO, J., P. CAPY, D. DEFAYE, E. PLA, J. SANDRIN, A. CSINK, and J. R. DAVID. 1989. Short-range genetic structure of *Drosophila melanogaster* populations in an Afrotropical urban area and its significance. Proc. Natl. Acad. Sci. USA **86**:8442–8446.

WHITTAM, T. S., and M. NEI. 1991. Scientific correspondence. Nature **354**:115–116.

BARRY G. HALL, reviewing editor

Received February 4, 1993; revision received May 5, 1993

Accepted May 5, 1993