Genetica **92**: 23-32, 1993.

Molecular evolution of the *Rh3* gene in *Drosophila*

Francisco José Ayala, Belinda S. W. Chang & Daniel L. Hartl*
Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, USA
** Author for correspondence*

Received 2 May 1993 Accepted 23 August 1993

Abstract

Previous investigations into the evolution of the *Drosophila* opsin gene family are extended by inter- and intraspecific DNA sequence comparisons of the *Rh3* locus in the *melanogaster* subgroup and *D. pseudoobscura*. Two separate statistical tests of the neutral-mutation hypothesis suggest that random genetic drift is responsible for virtually all of the observed amino acid replacement substitutions within the *melanogaster* subgroup. Analyses incorporating the *D. pseudoobscura* sequences are enigmatic due to the accumulation of multiple substitutions, because the McDonald-Kreitman test is not applicable to species comparisons that approach mutational saturation. However, the data from *D. pseudoobscura* are not inconsistent with selective neutrality. The ratio of amino acid polymorphisms within species to fixed differences between species imply that there are approximately 31 possible neutral single-step amino-acid-replacement substitutions at this locus. Synonymous substitutions are unevenly distributed among the structural domains of the *Rh3* gene. Patterns of synonymous polymorphism are analyzed with respect to GC content and codon bias, and are compared to other loci from the same species.

Introduction

Opsin genes in *Drosophila* have served as an excellent model for the study of both sensory transduction in the nervous system and molecular evolution in a multigene family. Opsin proteins, together with a covalently-bound chromophore, usually 11-*cis* retinal, are light-absorbing visual pigments found universally throughout the animal kingdom (Martin *et al.*, 1986). The opsins are integral disc membrane photoreceptor neurons and are members of the larger family of signal-transducing proteins with seven-transmembrane-domains that interact with G-proteins. Photons of light isomerize the chromophore from 11-*cis* to all-*trans* retinal, causing a conformational change in the protein moiety, triggering the phototransduction cascade (Nathans, 1992).

In *Drosophila*, four paralogous members of the opsin gene family have been identified, designated *Rh1* through *Rh4*. Although each have similar structures and functions, they have evolved different absorption spectra and cell-specific developmental patterns. The adult *Drosophila* visual system is composed of three simple eyes, the ocelli, in addition to the compound eyes, which are made up of approximately 800 units known as ommatidia. There are 20 cells in each ommatidium, including eight photoreceptor neurons designated R1-R8. The *Rh1* gene expresses a blue/green-absorbing opsin in cells R1-R6, which mediate optomotor behavior. The *Rh2* gene expresses a blue-absorbing opsin found in the ocelli. *Rh3* and *Rh4* encode ultraviolet-absorbing opsins, expressed alternately in R7 cells, which mediate phototactic responses. The opsin gene expressed in R8 has not been identified. (For a review of the *Drosophila* visual system see Smith, Stamnes & Zuker, 1991.)

A wide range of loci have been sequenced and analyzed from both *D. melanogaster* and *D. pseudoobscura*, and the patterns of nucleotide polymorphism reveal significant differences in the rates of both replacement and synonymous substitution between the different loci (Riley, 1989). Carulli and Hartl (1992) have analyzed the *Rh1* through *Rh4* genes in both species and found significant

differences in the rates of nucleotide substitution between the four members of this gene family. Furthermore, there is no correlation between the rates of synonymous and replacement substitutions among the four genes. These results suggest complex and varying levels of selective pressures acting upon these loci.

In this paper we present both inter- and intraspecific DNA sequence comparisons of *Rh3* in an attempt to further elucidate the factors influencing the evolution of this locus. Analyses of intraspecific sequence polymorphism can reveal patterns of evolution obscured by interspecific comparisons, and can also indicate whether those patterns are due to the adaptive fixation of advantageous mutations or to selectively neutral processes (Hudson, Kreitman & Aguadé, 1987; McDonald & Kreitman, 1991a; Sawyer & Hartl, 1992). We present DNA sequence from five alleles of *Rh3* from each of four species of the *D. melanogaster* subgroup, as well as three alleles from *D. pseudoobscura*.

Materials and methods

Drosophila *strains*

Several isofemale lines established from various *Drosophila* collections were used for this analysis. One isofemale line of *D. melanogaster* was studied from each of five collections representing Brazzaville (Congo), Cotonou (Benin), Harwich (United Kingdom), Lausanne (Switzerland), and St. Louis (Missouri); one isofemale line of *D. simulans* was studied from each of five collections representing Capetown (South Africa), Brazzaville (Congo), Le Cap (Haiti), Morrow Bay (California), and St. Louis (Missouri); five isofemale lines of *D. teissieri* were studied from collections made in Brazzaville (Congo); and five isofemale lines of *D. yakuba* were studied from collections made in the Ivory Coast. The isofemale lines from Africa and Haiti were kindly provided by Jean David and Pierre Capy. *D. pseudoobscura* strains PSU 434 from California Toiyabe National Forest and PSU 231 from Arizona Kaibab National Forest (kindly provided by Dr. Stephen Schaeffer), and the *D. pseudoobscura* sequence from Carulli and Hartl (1992) were also used. The *Rh3* gene was amplified and sequenced from one individual chosen at random from each of the isofemale lines. DNA extraction was performed according to the protocol of Gloor and Engels (1991).

Amplification of Rh3 *DNA*

PCR and sequencing primers were designed based on the published sequences (Zuker *et al.*, 1987; Carulli & Hartl, 1992) using the OLIGO 4.0 program for the Macintosh (Rychlik, 1992). The following PCR primers were used.

For *D. melanogaster:*
5′-CACTGCACTAACCTTCAGATG-3′
and 5′-CACATTGCGTTGAGTATGACC-3’.

For *D. simulans*, *D. teissieri*, and *D. yakuba:*
5′-CTGGGCTGGCTGTGATTTCGTCTT-3′
and 5′-GGCGTTTGGTATTACGAGTTC-3’.

For *D. pseudoobscura:*
5′-CCAAATGATGTATCAGCAGAA TG-3′
and 5′-TTGTATTACGAGTTCGTTTCC-3’.

The optimal annealing temperatures for all primers were estimated with the OLIGO 4.0 program. Following 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 ddH_2O, followed by phenol extraction, precipitation in the presence of 2.5 M NH_4OAc and one volume 100% ethanol, and washing with 70% ethanol (Sambrook, Fritsch & Maniatis, 1989).

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, Du & Wilson, 1993, and recommendations of the manufacturer). Organic contaminants were removed with the phenol-chloroform extraction protocol provided with the sequencing kit. Both strands of the entire *Rh3* coding region from all the *melanogaster* subgroup strains were sequenced. For the *D. pseudoobscura* strains both strands of all but the final 50 bp at the 3′ end of the coding region were sequenced; also, the first 21 bp at the 5′ end were ignored because of ambiguous sequence alignment.

Sequence analysis

DNA sequences were aligned with the MacVector program for the Macintosh (IBI, 1991). Nucleotide

substitutions from the aligned DNA sequences were classified into fixed differences between species and polymorphisms within species with reference to their phylogenetic relationships, which 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*. Statistical analyses were performed with the Statview program for the Macintosh (Abacus Concepts, 1992).

Results

The variable nucleotide positions found in the *Rh3* gene are listed in Figure 1. In Table 1 the nucleotide substitutions are classified according to their effect on amino acid sequence (replacement or synonymous) and status in the population (fixed between species or polymorphic within species). A subset of the data which includes only the sequences from the *melanogaster* subgroup (*D. melanogaster*, *D. simulans*, *D. teissieri*, and *D. yakuba*) was analyzed separately, in addition to the entire data set including all sequences examined. A G-test of independence with the Williams correction for continuity (Sokal & Rohlf, 1981) was used for all data sets to test the null hypothesis that the ratios of replacement to synonymous fixed substitutions is the same as the ratio of replacement to synonymous polymorphic substitutions. This null hypothesis represents the expected outcome if all alleles are evolving by selectively neutral processes; deviations from that expectation would be indicative of adaptive evolution in response to natural selection (McDonald & Kreitman, 1991a). The G-values and their probabilities are given in Table 1. Only the *melanogaster* subgroup data appear consistent with the neutral-mutation hypothesis.

Table 1. Number of replacement and synonymous nucleotide substitutions in the Rh3 gene that are fixed between species or polymorphic within species.

	Melanogaster subgroup[a]		*Melanogaster* subgroup and *D. pseudoobscura*	
	Fixed	Polymorphic	Fixed	Polymorphic
Replacement	3	1	29	1
Synonymous	49	47	177	52
G Value	0.82		8.14	
Probability	0.43		0.005	

[a] Includes *D. melanogaster, D. simulans, D. teissieri* and *D. yakuba*.

Table 2. Number of replacement and synonymous nucleotide substitutions in the Rh3 gene that are fixed between species or polymorphic within species, classified by structural domain.

	Melanogaster subgroup[a]					
	Transmembrane		Extracellular		Intracellular	
	Fixed	Poly.	Fixed	Poly.	Fixed	Poly
Replacement	1	0	1	1	1	0
Synonymous	14	23	22	14	13	10

	Melanogaster subgroup and *D. pseudoobscura*					
	Transmembrane		Extracellular		Intracellular	
	Fixed	Poly.	Fixed	Poly.	Fixed	Poly.
Replacement	13	0	10	1	6	0
Synonymous	72	28	67	16	38	3

[a] Includes *D. melanogaster, D. simulans, D. teissieri* and *D. yakuba*.

Another statistical test of the neutral-mutation hypothesis, known as the HKA test (Hudson, Kreitman & Aguadé, 1987), examines the correspondence between intraspecific polymorphism and interspecific divergence by comparing nucleotide sequence data from at least two loci in each of two species (the HKA test does not distinguish between replacement and synonymous differences). Variable nucleotide positions from *D. melanogaster* and *D. simulans* at the *Rh3* locus (Table 1) were compared with those of the *bride of sevenless* locus (*boss*; Ayala & Hartl, 1993). These data also appear consistent with the neutral-mutation hypothesis ($X^2 = 0.79$, 2 df, $P = 0.70$).

In Table 2, the nucleotide substitutions are subdivided according to structural domains within the protein. The Rh3 protein is predicted to contain

Site	Pub	*mel* abcde	*sim* fghij	*tei* klmno	*yak* pqrst	Type	*pse* uvw	Type
11	G	.....	CCCCC	CCCCC	CCCCC	FR		
28	C	.....	.....	.....	.....		GGG	FR
30	G	.....	AAAAA	.....	.....	FS	TTT	FS
31	T	.....	.....	.....	.....		CCC	FR
35	G	.....	.....	..A..	.....	PR	...	
39	C	.....	.....	.....	.....		TTT	FS
47	C	.....	.....	.....	.....		GGG	FR
48	C	.....	.....	.....	.....		TTT	FS
50	C	.....	.....	.....	.....		TTT	FR
52	C	.....	.....	.....	.....		TTT	FS
57	G	.....	.....	.....	.....		TTT	FS
63	G	.....	.....	.....	.....		TTT	FR
69	G	.....	.....	.....	.....		AAA	FS
72	G	CCCCC	CCCCC	CCCCC	CCCCC	PS	CCC	
78	C	.....	.....	.....	.....		AAA	FS
82	A	.....	.....	.....	.....		TTT	FR
87	T	.....	.....	AAAAA	AAAAA	FS	...	
90	G	.....	.....	.....	.....		CCC	FS
93	G	.....	.....	.....	AAAAA	FS	...	
114	G	.....	.....	.....	.....		TTT	FR
117	G	.....	.....	.....	.....		AAA	FS
120	G	.....	.....	.....	.....		AAA	FS
123	G	.....	TTTTT	TTTTT	TTTTT	FS	TTT	
129	T	.....	.....	CCCCC	.....	FS	AAA	FS
132	C	.....	TT...	.....	.....	PS	...	
135	G	.....	.....	.....	.....		AAA	FS
142	C	.....	..TTT	.....	.....	PS	...	
144	A	.....	.....	.....	.....		GGG	FS
146	C	.....	.....	.....	.....		TTT	FR
147	T	.....	.....	.....	.....		AAA	FS
150	C	.....	.....	.YTTT	TTTTT	PS,FS	TT.	PS
156	G	.....	AAAAA	.....	.....	FS	...	
162	C	.....	T....	.....	.....	PS	GGG	FS
165	A	.....	.....	GGGGG	GGGGG	FS	...	
177	C	.....	.....	.....	.....		TTT	FS
186	C	.....	.....	.....	.....		TTT	FS
192	C	.....	AT...	GGGGG	GGGGG	PS,PS,FS	GGT	PS
198	C	.....	.A...	.....	.....	PS	...	
201	C	.....	....T	.....	.....	PS	...	
204	C	.....	.....	.....	.....		TTT	FS
207	C	.....	T....	.....	.....	PS	GGG	FS
208	C	.....	.....	.....	.....		GGG	FR
210	G	.....	.....	.....	.....		CCC	FS
213	G	.....	.....	.....	.....		AAA	FR
220	C	.....	.....	.....	.....		AAA	FR
222	G	.....	.....	.....	.....		CCC	FS
225	C	.....	.....	.....	.....		GGG	FS
231	A	.....	.....	.....	.....		CCC	FS
234	G	.....	.....	.....	.....		CCC	FS
237	G	.....	.....	.....	.....		AAA	FS
240	T	.....	C....	CCCCC	CCCCC	PS,FS	GGG	FR
246	C	.....	...TT	.....	.....	PS	...	
252	C	.....	GGGGG	GGGGG	GGGGG	FS	...	FR
255	A	.....	.....	...G.	.....	PS	GGG	FS
261	A	.....	.....	.....	.....		GGG	FS
267	G	.....	.....	.....	.....		AAA	FS
268	C	.....	.....	.....	.....		AAA	FS
270	A	.....	.....	.....	CCCCC	CCCCC FS	GGG	FS
276	C	.....	.....	.....	.....		GGG	FS
282	T	.....	.....	.....	.....		CCC	FS
285	A	.....	.....	.....	.....		CCC	FS
288	G	.....	.....	.....	.....		TTT	FS
291	C	.....	.....	.....	.....		GGG	FS
315	C	.....	.....	.....	TTTTY	PS	TTT	FS
325	G	.....	.....	.....	.....		AAA	FR
333	T	.....	.....	.....	.....		GGG	FS
336	G	.....	.....	AAAAA	CCCCC	FS,FS	CCC	
339	A	.....	.....	.....	.....		CCC	FS
351	T	.....	.C.C.	.....	.....	PS	CCC	FS
354	C	.....	.TTTT	TTTTT	TTTTT	PS,FS	...	
360	C	.....	.....	.....	.....		TTT	FS
366	A	.....	.....	.....	.....		CCC	FS
369	T	.....	CCCCC	CCCCC	CCCCC	FS	...	FS
372	G	.....	..T..	AAAAA	AAAAA	PS,FS	CCC	FS
373	C	.....	.....	.....	.....		TTT	FS
374	T	.....	.....	.....	.....		CCC	FR
375	G	.....	...A.	.....	.....	PS	...	
378	T	.....	.....	.....	.....		AAA	FS
381	T	.....	.....	.....	.....		CCC	FS
382	C	.....	.....	.....	.....		TTT	FS
387	A	.....	.....	.....	.....		CCC	FS
390	C	.....	.....	.....	.....		TTT	FS
399	T	.....	.....	.....	.....		CCC	FS
402	A	.....	GGGGG	GGGGG	GGGGG	FS	GGG	

Fig. 1. The first two columns contain the position numbers and corresponding nucleotides from the coding region of the published *D. melanogaster Rh3* sequence (Zuker *et al.*, 1987). Strains are identified with species names (*mel* = *D. melanogaster*; *sim* = *D. simulans*; *yak* = *D. yakuba*; *tei* = *D. teissieri*; *pse* = *D. pseudoobscura*) and collection sites (a, p, q, r, s, t = Brazzaville; b = Cotonou; c = Harwich; d = Lausanne; e, j = St. Louis; f = Capetown; g = Congo; h = Le Cap; i = Morrow Bay; k, l, m, n, o = Ivory Coast; u = published sequence of Carulli & Hartl (1992); v = California; w = Arizona). Nucleotides identical to the published sequence are shown as periods (.). Heterozygous sites are represented as: Y = C or T; K = G or T; S = G or C; W = A or T. The last column lists the types of substitutions observed at each site: FR = fixed replacement; FS = fixed synonymous; PR = polymorphic replacement; PS = polymorphic synonymous. Boxed regions indicate transmembrane domains.

seven transmembrane domains connected by inter- and extracellular loops (Zuker *et al.*, 1987). G-tests of goodness of fit with the Williams correction (Sokal & Rohlf, 1981) were performed to test whether the nucleotide substitutions were randomly distributed among the three main structural domains of *Rh3*. Within the *melanogaster* species subgroup, synonymous substitutions were found to be uniformly distributed (G = 3.04, 2 df, P = 0.29). Within the entire data set, comprising both the *melanogaster* subgroup species and *D. pseudoobscura*, the synonymous substitutions were significantly non randomly distributed, with fewer substitutions than expected occurring in the intracellular region and more than expected in the extracellular domain (G = 7.07, 2 df, P = 0.03). With respect to amino acid replacements, the numbers are insufficient to test for a non random distribu-

Site	Pub	*mel* abcde	*sim* fghij	*tei* klmno	*yak* pqrst	Type	*pse* uvw	Type
403	A	.....	.....	.....	.....		GGG	FR
405	C	.....	TT...	.....	.....	PS	...	
417	T	.....	C	.....	.....	PS	CCC	FS
423	A	.....	.....	.....	.....		GGG	FS
426	C	.....	.....	.....	.....		TTT	FS
429	T	.....	.....	.....	.....		GGG	FS
432	C	.....	.....	.....	.....		GGG	FS
441	C	.....	.....	GGGGG	GGGGG	FS	GGG	
444	T	.....	.....	.....	.....		CCC	FS
447	G	.....	.....	.....	.....		CCC	FS
450	T	.....	.....	.....	.....		CCC	FS
453	A	.....	.....	.....	.....		CCC	FS
462	T	.....	.....	.....	.....		CCC	FS
465	A	.....	.....	.....	.....		CCC	FS
467	T	.....	.....	.....	.....		AAA	FR
477	C	.....	.....	.....	.....		AAA	FS
480	T	.....	GGGGG	GGGGG	GGGGG	FS	GGG	
483	A	.....	.T.T.	.....	.....	PS	...	
504	G	.....	.A...	.....	.....	PS	...	
519	T	.....	.....	.....	.....		AAA	FS
522	C	.....	.A..T	.....	.....	PS,PS	...	
528	C	.....	.....	.....	.....		AA.	PS
531	A	.....	.....	.....	.....		TTT	FS
540	C	.....	T....	.....	.....	PS	TTT	FS
541	A	.....	.....	.....	.....		CCC	FR
546	C	.....	.....	.....	.....		TTT	FS
552	T	.....	.....	.....	.....		GGG	FS
555	A	.....	.....	TTTTT	TTTTT	FS	CCC	FS
564	T	.....	....A	AAAAA	AAAAA	PS,FS	GGG	FS
579	G	.....	.....	.....	.....		AAA	FS
580	A	.....	.....	.....	.....		TTT	FR
582	T	.....	.....	.....	.....		CCC	FS
588	C	.....	A.AA.	TTTTT	TTTTT	PS,FS	...	
591	T	.....	.C...	.....	.....	PS	...	
594	T	.....	.....	.....	.....		CCC	FS
597	G	.....	.....	.....	.....		CCC	FS
600	G	.....	.....	.....	.....		CCC	FS
603	G	.....	.....	.....	.....		AAA	FS
606	A	.....	.....	TYCCC	TTTTT	PS,FS	...	
612	G	.....	.....	.....	.....		CCC	FS
615	A	.....	GGGGG	CCCCC	CCCCC	FS,FS	GGG	
618	C	.....	.....	.....	.....		TTG	FS,PS

Site	Pub	*mel* abcde	*sim* fghij	*tei* klmno	*yak* pqrst	Type	*pse* uvw	Type
627	T	.....	.....	CCCCC	CCCCC	FS	...	
636	C	.....	.....	AAAAA	AAAAA	FS	GGG	FS
639	C	.....	.....	.....	.....		GGG	FS
642	T	.....	C	CCCCC	CCCCC	PS,FS	...	
648	C	.....	T....	TTTTT	TTTTT	PS,FS	TTT	
657	A	.....	CCCCC	CCCCC	CCCCC	FS	TTT	FS
663	T	.....	.....	.....	.....		CCC	FS
668	C	.....	GGGGG	GGGGG	GGGGG	FR	...	FR
672	C	.....	.....	.....	.....		.TT	PS
690	C	.....	TTTTT	.....	.....	FS	...	
693	G	.....	.....	.....	.....		CCC	FS
696	T	.....	CCCCC	CCCCC	CCCCC	FS	CCC	
702	C	.....	.S...	.....	.....	PS	...	
705	G	.....	.....	.....	.....		CCC	FS
717	C	.....	.YYT.	.....	.....	PS	...	
723	C	.....	.YYT.	.....	.....	PS	...	
726	C	.....	.....	.....	.....		GGG	FS
732	T	.....	.....	.....	AAAAA	FS	...	
744	G	CCCCC	CCCCC	CCCCC	CCCCC	PS	CC.	PS
747	T	.....	.....	CCCCC	CCCCC	FS	...	
756	G	.....	.....	..AAA	AAAAA	PS,FS	...	
759	A	.....	.....	GGGGG	GGGGG	FS	GGG	
762	A	.....	.....	.....	.....		GGG	FS
766	C	.....	.....	AAAAA	AAAAA	FS	...	
768	G	.....	.....	AAAAA	AAAAA	FS	TTT	FS
777	C	.....	T....	TTTTT	TTTTT	PS,FS	...	
789	C	.....	.....	.....	.....		TTT	FS
792	G	.....	.....	.....	.....		CCC	FS
795	A	.....	.....	.....	.....		TTT	FR
798	G	.....	.....	TTTTT	TTTTT	FS	...	
804	T	.....	CCCCC	CCCCC	CCCCC	FS	CCC	
807	G	.....	..KT.	.....	.....	PS	AAA	FS
819	A	.....	.....	.....	.....		GGG	FS
821	A	.....	GGGGG	GGGGG	GGGGG	FR	GGG	
829	A	.....	.....	.....	.....		GGG	FR
831	G	.....	.....	.....	.....		TTT	FS
834	G	.....	.....	.....	.....		AAA	FS
840	C	.....	.....	.....	.....		TTT	FS
843	G	.....	T...A	.....	.....	PS,PS	CCC	FS
846	A	.....	.....	.....	.....		TTT	FS
852	A	.....	.....	.....	.....		GGG	FS
861	C	.....	.....	.....	TTTTT	FS	...	

Site	Pub	*mel* abcde	*sim* fghij	*tei* klmno	*yak* pqrst	Type	*pse* uvw	Type
867	A	.....	CCCCC	CCCCC	CCCCC	FS	TTT	FS
870	C	.....	.....	.....	.....		TTT	FS
876	G	A....	.....	.....	.....	PS	...	
883	T	.....	.....	.....	.....		GGG	FR
884	G	.....	.....	.....	.....		CCC	FR
885	C	.....	.....	.....	.....		AGG	PS,FS
888	G	.....	.....	.....	.....		CCC	FS
894	G	.....	.....	.....	.....		AAA	FS
897	G	.....	.....	.....	.....		AAA	FS
903	A	.....	.....	.....	.....		CCC	FS
906	T	.....	.....	..W..	.....	PS	GGG	FS
912	G	.....	.....	.....	.....		CCC	FS
915	G	.....	.....	.....	.....		CCC	FS
921	C	.....	.....	.....	.....		AAA	FS
924	C	.....	.....	.....	.....		AAA	FS
933	T	.....	.....	.....	.....		CCC	FS
939	C	.....	.....	.....	.....		TTT	FS
940	C	.....	.....	.....	.....		TTT	FS
942	T	.....	.....	.....	.....		GGG	FS
943	T	.....	.....	.....	CCCCC	FS	CCC	FS
945	G	.....	.....	.....	.....		CCC	FS
948	G	.....	.....	.....	.....		TTT	FS
954	A	.....	.....	.....	.....		CCC	FS
966	T	.....	.....	.....	.....		CCC	FS

Site	Pub	*mel* abcde	*sim* fghij	*tei* klmno	*yak* pqrst	Type	*pse* uvw	Type
975	T	.....	.....	.....	.....		CCC	FS
976	G	.....	.....	.....	.....		AAA	FR
984	A	.....	.....	...G.	.....	PS	...	
999	C	T....	.....	.....	.....	PS	AAA	FS
1005	G	.....	.....	.....	.....		CCC	FS
1014	C	.....	.....	.....	.....		TTT	FS
1020	A	.....	.....	.....	.....		CCC	FS
1023	T	.....	...C.	.....	.....	PS	...	
1026	T	.....	CCCCC	CCCCC	CCCCC	FS	...	FS
1035	C	.....	.....	.....	.....		TTT	FS
1050	G	.....	.....	AAAAA	AAAAA	FS	...	
1056	A	.....	.....	.....	.....		CCC	FS
1059	T	.....	.....	.....	.....		CCC	FS
1071	G	.....	.....	.....	.....		CCC	FS
1072	C	.....	.....	.....	.....		AAA	FR
1076	A	.....	.....	.....	.....		GGG	FR
1080	A	.....	.....	.....	.....		GGG	FS
1086	G	.....	.....	.....	.....		CCC	FS
1089	G	.....	.....	.....	.....		CCC	FS
1107	C	.....	.....	.....	.TYY.	PS		
1131	A	.....	.....	GGGGG	GGGGG	FS		
1137	G	.....	.....	AAAAA	AAAAA	FS		
1140	C	.YY..	.....	.....	.....	PS		

tion. Three-dimensional log-linear modeling analyses (Sokal & Rohlf, 1981) were also performed on the data sets of Table 2 to test for significant differences in the ratios of replacement to synonymous substitutions, both between fixed and polymorphic substitutions and between the three main structural domains. The tests revealed no differences in the association between status in the population and effect on amino acid sequence for the different structural domains (after a single iteration, *melanogaster* subgroup: G = 2.37, 2 df, P > 0.1; entire data set: G = 2.51, 2 df, P > 0.1).

Under the assumption of selective neutrality, the observed ratio of replacement to synonymous substitutions should be equal to the ratio of total possible neutral replacement substitutions to total possible synonymous substitutions (McDonald & Kreitman, 1991a; Sawyer & Hartl, 1992). For *Rh3* in the *melanogaster* subgroup, this ratio is (3 + 1): (49 + 47) = 1:24, and the total number of possible synonymous substitutions is 739. The total number of possible neutral amino-acid-replacement substitutions in the *melanogaster* subgroup at this locus is therefore approximately 739/24 = 31. The binomial standard error on this estimate is approximately 11.

The average percent G + DC content at four fold degenerate sites (GC_4) within the *melanogaster* subgroup species is 80.8%, 80.7%, 81.9% and 81.3% for *D. melanogaster*, *D. simulans*, *D. teissieri*, and *D. yakuba*, respectively. These values are among the highest known for any *D. melangoaster* locus (E. Moriyama, unpublished observations). The average 'scaled χ^2 value' which provides a measure of codon bias (Shields *et al.*, 1988) for the four species is 0.51, 0.47, 0.51 and 0.49, given in the same order as above. These values are intermediate in range for *D. melanogaster* loci (Shields *et al.*, 1988). The average GC_4 of the *D. pseudoobscura RH3* alleles is 81.8%, and the average scaled χ^2 value is 0.41.

In Table 3, the patterns of synonymous substitution at the*Rh3* locus in the *melanogaster* subgroup are compared with those from two other loci, *alcohol dehydrogenase* (*Adh*; McDonald & Kreitman, 1991a) and *boss* (Ayala & Hartl, 1993) (*D. teissieri* data were excluded from this comparison because corresponding data from *Adh* are unavailable). A 2 × 3 G-test of independence with the Williams correction (Sokal & Rohlf, 1981) indicates a significant difference in the patterns of synonymous-site evolution between *Rh3* and the other two loci (G = 7.42, 2 df, P = 0.025). The *Rh3* gene has either a deficit of polymorphic synonymous substitutions or an excess of fixed synonymous substitutions.

The average number of nucleotide differences per site between two randomly chosen alleles of Rh3, a measure of nucleotide diversity independent of both sequence length and sample size (Nei, 1987), was calculated for all five species. These values are 0.0007, 0.0120, 0.0021, 0.0002, and 0.0043 for *D. melanogaster*, *D. simulans*, *D. teissieri*, *D. yakuba*, and *D. pseudoobscura*, respectively. The nucleotide diversities of *Rh3* in *D. melanogaster* and *D. simulans* are among the lowest and highest, respectively, of any known locus from these species.

Table 3. Number of synonymous nucleotide substitutions in the *Adh*, *boss* and *Rh3* genes of the *D. melanogaster* subgroup[a] that are fixed between species or polymorphic within species.

	Fixed		Polymorphic
Adh	20		44
boss	60		75
Rh3	47		41
G Value		7.42	
Probability		0.025	

[a] Includes *D. melanogaster*, *D. simulans* and *D. yakuba*.

Discussion

Ever since Darwin first proposed the theory of evolution by natural selection, the special problem of the origin and evolution of visual systems has intrigued evolutionary biologists. The *Drosophila* visual system is extraordinarily complex, and genetic studies have elucidated a wide range of loci, including developmental, functional, and structural genes, within this system. The advent of molecular techniques has also made possible very fine-scale analyses of evolutionary processes at the molecular level. Carulli and Hartl (1992) have analyzed the DNA sequences of the four known opsin genes, *Rh1* through *Rh4*, in both *D. melanogaster* and *D. pseudoobscura*. We have extended this investigation into the evolution of the opsin loci by analyzing both inter- and intra-specific DNA sequence comparisons of one of the opsin loci, *Rh3*.

Genetic variation in natural populations is the result of a combination of adaptive processes which influence allelic frequencies through natural selection, and selectively-neutral processes, including mutation, genetic drift and intra-allelic recombination. McDonald and Kreitman (1991a) have pointed out that, in the absence of adaptive evolution, the ratio of replacement to synonymous fixed nucleotide substitutions should be the same as the ratio of replacement to synonymous polymorphic substitutions. Hence, analyses of both inter- and intra-specific nucleotide diversity can reveal whether a given locus is undergoing adaptive changes. The patterns of nucleotide substitution at the *Rh3* locus in the *melanogaster* subgroup are consistent with the neutral expectation (Table 1). Although considerable controversy has arisen over the validity of this test of the neutral-mutation hypothesis (e.g., Graur & Li, 1991; Whittam & Nei, 1991; McDonald & Kreitman, 1991b), a detailed theoretical foundation for the test has been provided by Sawyer and Hartl (1992). The HKA test, in which patterns of polymorphism and divergence are compared between different loci, again suggests that neutral processes are responsible for virtually all of the observed nucleotide substitutions in *Rh3* in the *melanogaster* subgroup.

When the patterns of *Rh3* nucleotide substitution from the *melanogaster* subgroup are combined with those of *D. pseudoobscura*, the interpretations of the McDonald-Kreitman test are not so straightforward, owing to the possibility of multiple nucleotide substitutions at many sites (an effect of mutational saturation). Specifically, a large number of fixed replacement substitutions suggests adaptive fixation of advantageous mutations, but a correction must be made for multiple substitutions. Although the divergence of the *melanogaster* subgroup has been estimated as occurring between 1.6 and 6.1 Myr (million years) ago (Caccone, Amato & Powell, 1988; Ayala & Hartl, 1993), the separation of *D. pseudoobscura* from the *melanogaster* lineage occurred approximately 30 to 40 Myr ago (Beverly & Wilson, 1984). Because comparisons between species that have evolved independently for so long must take into account the possibility of accumulation of multiple substitutions, it is difficult to obtain reliable estimates of the true number of substitutions (Lewontin, 1989). Replacement and synonymous substitutions will saturate at different rates, because the proportion of transversions is far higher among replacement substitutions than synonymous substitutions, and transitions and transversions occur at different rates (Hartl, 1988). The synonymous substitutions themselves fall into different degeneracy classes, which are also known to saturate at different rates (Riley, 1989). Because the time separating *D. pseudoobscura* from the *melanogaster* species subgroup is so large compared to the time since the radiation of the individual species, the differences in saturation rates will affect the number of fixed substitutions to a greater extent than polymorphic substitutions. If replacement substitutions have a lower number of neutral alternative states, and also have a lower saturation rate owing to the higher proportion of transversions, then the ratio of replacement to synonymous substitutions would be higher for fixations than polymorphisms. The discrepancy between the two ratios in the *melanogaster* subgroup/*D. pseudoobscura* comparison of Table 1 could therefore plausibly be attributed to different rates of saturation.

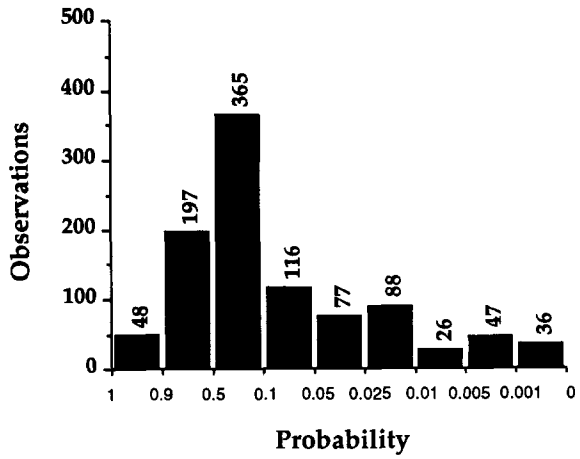

Fig. 2. Distributions of observed probability values of 2 × 2 contingency tables generated from computer simulations of forward coalescent processes. Out of 1000 replicates, 274 were significant at the $P < 0.05$ level.

Monte Carlo simulations were performed to test whether neutral processes could produce significant differences in these ratios. The simulation is a forward coalescent process with random mutation, in which two initially identical sequences diverge from each other and subsequently radiate in a random branching process. The simulated sequences

contain the same number of 2-fold degenerate replacement, and 2-, 3-, and 4-fold degenerate synonymous sites, as in the *D. melanogaster Rh3* gene (estimation of the number of neutral replacement sites as 31 is described in the results). Population parameters, including mutation rates, branch lengths, and relative number of neutral transition and transversion substitutions, were chosen to approximate actual values. A 2:1 ratio of transition to transversion mutations was assumed (Hartl, 1988). In addition, the simulations make the following assumptions: (1) all substitutions are selectively neutral, (2) replacement substitutions behave like 2-fold degenerate synonymous substitutions, i.e. there are no more than two possible neutral states at each position, and (3) all replacement substitutions are independent, i.e. there is no interaction between different positions. One thousand replicates were performed. At the end of each run, nucleotide substitutions were counted and classified as replacement or synonymous, and fixed or polymorphic, and a corresponding G-statistic was calculated. The resulting distribution of probability values is shown in Figure 2. A significant difference between the two ratios was observed at a frequency of 27.4%, indicating that the probability of a Type I statistical error (rejection of a true null-hypothesis) is high, and that differences in the ratios of replacement to synonymous substitutions that are fixed or polymorphic could be the result of differences in saturation rates at neutral sites.

The assumption that there are no more than two neutral states for any replacement site is supported by the observation that there are no sites in the *Rh3* data which contain more than one replacement substitution. The assumption of independence among neutral replacement positions is difficult to test. In *Rh3*, there is one instance of two replacement substitutions occurring within a single codon at nucleotide positions 883 and 884. Site-directed mutagenesis experiments (DuBose & Hartl, 1989), and comparisons across a wide phylogenetic range (Fitch & Markowitz, 1970), provide evidence that some replacement substitutions can affect the selective constraints operating on other positions within the same locus. It is unclear whether a violation of either of these two assumptions is likely to have a large effect on the ratios presented in Table 1. However, these simulations confirm the assertion of McDonald and Kreitman (1991a) that their test is not appropriate for species comparisons that approach saturation.

Increasing the number of sequences would be unlikely to resolve the ambiguities of the *melanogaster* subgroup/*D. pseudoobscura* comparison. In principle, the outcome of the McDonald-Kreitman test is independent of sample size (McDonald & Kreitman, 1991b; Sawyer & Hartl, 1992). Additional *D. pseudoobscura* sequences, for example, might increase the number of observed polymorphisms relative to fixations, but this effect would be the same for both replacement and synonymous differences, and would therefore not significantly alter the ratios of Tables 1 or 2.

Varying rates of synonymous-site evolution have been reported within genes (Lawrence, Hartl & Ochman, 1991), between codon degeneracy classes (Riley, 1989), between *Drosophila* lineages (Moriyama & Gojobori, 1992), and between loci (Moriyama & Gojobori, 1992). The *Drosophila* opsin gene family also exhibits a complex pattern of synonymous evolution. Carulli and Hartl (1992) found no consistent relationships between codon bias, base composition of third position sites, or synonymous substitution rates of the opsin genes between *D. melanogaster* and *D. pseudoobscura*. Furthermore, there was no consistent relationship between synonymous substitution rates and amino acid replacement rates among the four opsin genes. The *Rh3* and *Rh4* genes, for example, which share 75% amino acid identity and have similar functions and expression patterns, have very similar numbers of synonymous substitutions between the two species, but more than a two-fold difference in the number of replacement substitutions.

Has natural selection shaped the patterns of synonymous substitution in *Rh3*? Considerable evidence for selection of synonymous substitution in *Drosophila* has been reported (Shields *et al.*, 1988), and several mechanisms have been proposed (Sharp & Li, 1986). Within *Rh3*, we found synonymous substitutions to be evenly distributed among the three main structural domains in the *melanogaster* subgroup (Table 2). Between the *melanogaster* subgroup species and *D. pseudoobscura*, however, there is a significantly non random distribution of synonymous substitutions, with a large excess in the extracellular domain and a corresponding deficit in the intracellular region. This discrepancy is difficult to interpret. Several factors that may be

responsible for intragenic variation in synonymous substitution rate have been identified in bacteria, including stability of mRNA secondary structure (Lawrence, Hartl & Ochman, 1991), as well as selection for the use of non optimal codons at the beginning of highly-expressed genes, possibly to retard translation in order to prevent interference between the leading ribosome and RNA polymerase (Bulmer, 1988). It is unclear to what extent these factors may be operative in *Drosophila*.

The intraspecific analyses also reveal differences in the relative numbers of fixed and polymorphic synonymous substitutions between *Rh3* and two other loci from the same species comparisons, *Adh* and *boss* (Table 3). These differences in the evolutionary accumulation of synonymous mutations, however, could be attributed to non-selective processes, such as gene conversion or hitchhiking, and therefore do not necessarily indicate adaptive evolution.

Acknowledgements

We thank Jean David & Pierre Capy for providing *melanogaster* subgroup strains, and Steve Schaeffer for providing *D. pseudoobscura* stains.

References

Abacus Concepts, StatView, 1992. Abacus Concepts, Inc., Berkeley CA.

Ayala F.J. & D.L. Hartl, 1993. Molecular drift of the *bride of sevenless (boss)* gene in Drosophila. Mol. Biol. Evol. (In press).

Beverly, S.M. & 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.

Bulmer, M., 1988. Codon usage and intragenic position. J. Theor. Biol. 133: 67-71.

Caccone, A., G.D. Amato & J.R. Powell, 1988. Rates and patterns of scnDNA and mtDNA divergence within the *Drosophila melanogaster* subgroup. Genetics 118: 671-683.

Carulli, J.P. & D.L. Hartl, 1992. Variable rates of evolution among Drosophila opsin genes. Genetics 132: 193-204.

DuBose, R.F. & D.L. Hartl, 1989. An experimental approach to testing modular evolution: Directed replacement of α-helices in a bacterial protein. Proc. Natl. Acad. Sci. USA 86: 9966-9970.

Fitch, W.M. & E. Markowitz, 1970. An improved method for determining codon variability in a gene and its application to the rate of fixation of mutations in evolution. Biochem. Gen. 4: 579-593.

Gloor, G. & W. Engels, 1991. Single-fly DNA preps for PCR. DIS 1: 17-18.

Graur, D. & W.-H. Li, 1991. Scientific correspondence. Nature 354: 114-115.

Halloran, N., Z. Du & R. Wilson, 1993. Sequencing reactions for the Applied Biosystems 373A automated sequencer. In: Methods in Molecular Biology, Vol 10: DNA sequencing: Laboratory Protocols (H.G. Griffen & A.M. Griffin, Eds.). The Humana Press, Inc. Clifton, NJ (In press).

Hartl, D.L., 1988. A primer of population genetics. Sinauer, Sunderlan, MA.

Hudson, R.R., M. Kreitman & M. Aguadé, 1987. A test of neutral molecular evolution based on nucleotide data. Genetics 116: 153-159.

IBI, MacVector 3.5, 1991. IBI, New Haven, CT.

Lachaise, D., R.C. Cariou, J.R. David, F. Lemeunier, L. Tsacas & M. Ashburner, 1988. Historical biogeography of the *Drosophila melanogaster* species subgroup. Evol. Biol. 22: 159-227.

Lawrence, J.G., D.L. Hartl & H. Ochman, 1991. Molecular considerations in the evolution of bacterial genes. J. Mol. Evol. 33: 241-250.

Lewontin, R.C., 1989. Inferring the number of evolutionary events from DNA coding sequence differences. Mol. Biol. Evol. 6: 15-32.

Martin, R.L., C. Wood, W. Baehr & M.L. Applebury, 1986. Visual pigment homologies revealed by DNA hybridization. Science 232: 1266-1269.

McDonald, J.H. & M. Kreitman, 1991a. Adaptive protein evolution at the *Adh* locus in Drosophila. Nature 351: 652-654.

McDonald, J.H. & M. Kreitman, 1991b. Scientific correspondence. Nature 354: 116.

Moriyama, E.N. & T. Gojobori, 1992. Rates of synonymous substitution and base composition of nuclear genes in Drosophila. Genetics 130: 855-864.

Nathans, J., 1992. Rhodopsin: Structure, function, and genetics. Biochem. 31: 4923-4931.

Nei, M., 1987. Molecular Evolutionary Genetics. Columbia University Press, New York City.

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

Rychlik, W., 1992. OLIGO 4.0. National Biosciences, Inc., Plymouth, MN.

Saiki, R.K., D.H. Gelfand, S. Stoffel, S.J. Scharf, R.G. Higuchi, G.T. Horn, K.B. Mullis & H.A. Erlich, 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239: 487-491.

Sambrook, J., E.F. Fritsch & T. Maniatis, 1989. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Sawyer, S.A. & D.L. Hartl, 1992. Population genetics of polymorphism and divergence. Genetics 132: 1161-1176.

Sharp, P.M. & W.-H. Li, 1986. An evolutionary perspective on synonymous codon usage in unicellular organisms. J. Mol. Evol. 24: 28-38.

Shields, D.C., P.M. Sharp, D.G. Higgins & F. Wright, 1988. 'Silent' sites in Drosophila genes are not neutral: Evidence of selection among synonymous codons. Mol. Biol. Evol. 5: 704-716.

Smith, D.P., M.A. Stamnes & C.S. Zuker, 1991. Signal transduction in the visual system of Drosophila. Annu. Rev. Cell. Biol. 7: 161-190.

Sokal, R.R. & F.J. Rohlf, 1981. Biometry. W.H. Freeman, San Francisco.

Whittam, T.S. & M. Nei, 1991. Scientific correspondence. Nature 354: 115-116.

Zuker, C.S., C. Montell, K. Jones, T. Laverty & G.M. Rubin, 1987. A rhodopsin gene expressed in photoreceptor cell R7 of the Drosophila eye: Homologies with other signal-trandsducing molecules. J. Neuro. 7: 1550-1557.