Nucleotide Sequence of the *Xdh* Region in *Drosophila pseudoobscura* and an Analysis of the Evolution of Synonymous Codons[1]

Margaret A. Riley
Museum of Comparative Zoology, Harvard University

The nucleotide sequence of the *Xdh* region of *Drosophila pseudoobscura* is presented. The *Xdh* gene structure and organization are compared with the homologous region in *D. melanogaster*. This locus is shown to have similar organization in the two species, although an additional intron and three insertion/deletion events are described for the *D. pseudoobscura* coding region. The encoded proteins are predicted to have very similar charges and hydrophobic/hydrophilic domains even though 11% of the amino acids are different. A gene 5′ to *Xdh*, putative *l(3)s12*, is suggested from sequence similarity between the species. Synonymous differences at the *Xdh* locus between the two species are analyzed using a new method described in the preceding paper by Lewontin. This analysis shows that synonymous positions within the *Xdh* locus are evolving at very different rates, being dependent on level of codon redundancy. A comparison of synonymous divergence between *D. melanogaster* and *D. pseudoobscura* in five additional genes reveals variation in the level of synonymous substitution.

Introduction

The rate of synonymous nucleotide change in the coding regions of genes has become a crucial measurement in the study of molecular evolution. The reasons are threefold: precise alignment of coding regions is possible, these positions evolve at relatively high rates, and their rates of evolution, it has been argued, approach the selectively neutral rate. Species-level comparisons of DNA sequences will soon represent a large sample of genetic loci in *Drosophila*. In addition to addressing questions of synonymous position evolution, these data are providing details on the rates of nucleotide substitution experienced at different genetic loci within the same species, on the level of selection experienced at different nucleotide positions within a locus, and on the different forms of selection operating to produce the observed levels of substitution, e.g., selection at the protein sequence level, at the codon level (e.g., codon bias), and at the nucleotide level (e.g., transition/transversion bias).

There has been increasing acceptance of the hypothesis that molecular divergence is linear with time (e.g., see Miyata et al. 1980; Hayashida and Miyata 1983; Nei 1987), although not without some resistance (Gillespie 1986). This concept is of particular importance to the application of certain models of molecular evolution. It is necessary both to establish the existence of clocklike behavior in nucleotide substitutions and to provide the appropriate model (e.g., pseudogenes, silent sites, or intron positions) with which to measure the underlying rate of substitution for a chromosomal region. The

1. Key words: *Xdh, Drosophila,* synonymous substitutions.

Address for correspondence and reprints: Dr. Margaret A. Riley, Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts 02138.

Mol. Biol. Evol. 6(1):33–52. 1989.

0737-4038/89/0601-0003$02.00

assumption is, then, that this rate represents the best approximation of the neutral mutation rate. Positions that deviate from this rate can be examined to establish which forces, e.g., natural selection or drift, are responsible for the observed deviation.

The alcohol dehydrogenase locus, *Adh,* has been the subject of numerous species-level sequence comparisons (Bodmer and Ashburner 1984; Cohn 1985; Coyne and Kreitman 1986; Schaeffer and Aquadro 1987). These data have substantiated previous reports (e.g., Kafatos et al. 1977; Jukes and King 1979) that rates of evolution for functionally distinct nucleotide positions (e.g., amino acid replacement sites, synonymous sites, and intron positions) are very different. Patterns of substitution across the *Adh* exons reveal that all replacement sites and all synonymous sites do not experience the same levels of mutation and/or substitution, i.e., the observed changes are clustered in certain exons. Finally, comparison of *Adh* and a closely linked, uncharacterized gene 3′ to *Adh* shows that the rates of substitution at replacement, synonymous, and intron positions vary at different loci within the same chromosomal region (Schaeffer and Aquadro 1987). It is necessary to determine the generality of the results obtained at *Adh,* where the majority of *Drosophila* species comparisons have been made.

In the present paper a nucleotide sequence of the xanthine dehydrogenase locus (*Xdh*) in *D. pseudoobscura* is presented. Features of *Xdh* gene organization in *D. pseudoobscura* are described, and a detailed comparison with the *Xdh* nucleotide sequence from *D. melanogaster* (Keith et al. 1987; Lee et al. 1987) is made. The large coding region, 4,002 homologous base pairs between *D. melanogaster* and *D. pseudoobscura,* permits more powerful statistical tests of the nucleotide substitution data than are possible for *Adh.*

The nucleotide differences observed between *D. melanogaster* and *D. pseudoobscura* at *Xdh* and five additional loci are analyzed using a new method of comparative sequence analysis. The details of this method are presented in the preceding paper (Lewontin 1988). Basically, the analysis provides a codon-by-codon description of the nucleotide differences observed and inferred between two sequences. The synonymous differences are then parceled into groups based on the degeneracy of the amino acid code [two-, three-, four-, and sixfold-degeneracy classes of Li et al. (1985)].

Material and Methods

The strain of *Drosophila pseudoobscura* chosen for sequence analysis was kindly provided by T. P. Keith. The fly line was obtained and the second chromosome was made homozygous as described by Keith (1983). This line corresponds to the electromorph designation 1.00/1.02/1.01/1.02/1.03 in the XDH protein electrophoretic survey of Keith et al. (1985).

Total genomic DNA was isolated by following standard procedures. *Sau*3A partially cleaved genomic DNA was ligated into the *Bam*HI site of lambda vector EMBL4, packaged, and plated out onto host bacterial strain Q359. Fifty thousand recombinant plaques were screened according to the method of Maniatis et al. (1982) for sequences homologous to the *D. melanogaster* 4.6-kb *Eco*RI *Xdh* fragment described in Keith et al. (1987). Eight hybridizing phage were isolated. One, Jr436p2, contains the entire *Xdh* locus in a 16-kb insert. *Xdh* was localized on the phage restriction map by blot hybridization with the same probe on nylon filters (Gene Screen).

A 9-kb portion of the *Xdh* region was subcloned into M13MP18 or MP19 vectors, and single-stranded DNA was prepared from infected bacteria (Bankier and Barrel 1983). A sequential series of overlapping clones was generated using the procedure of Dale et al. (1985). DNA was sequenced by the dideoxy chain-termination method

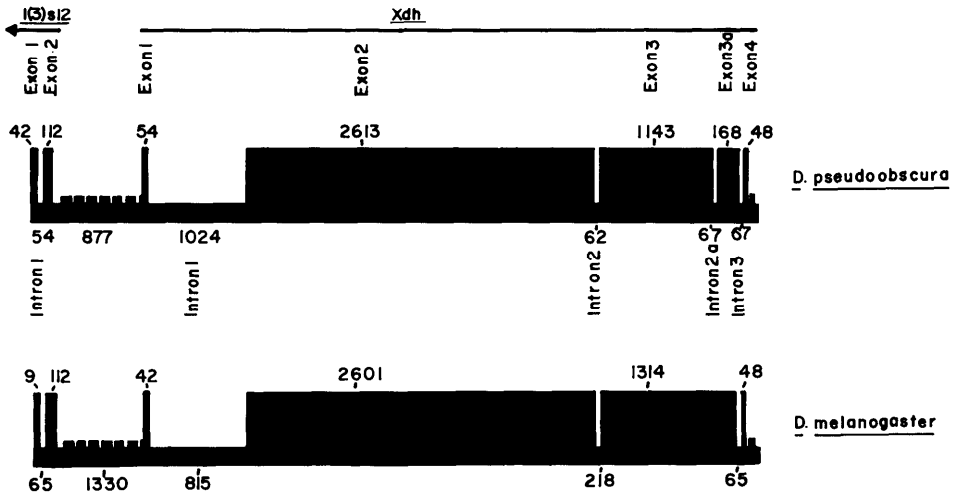

FIG. 1.—Diagram of the *Xdh* region in *Drosophila pseudoobscura* and *D. melanogaster*. Dark boxes indicate exons, dashed lines indicate intergenic regions, and connecting lines indicate introns. Lengths are given (in base pairs) above each figure for exons and below each figure for introns and intergenic regions.

of Sanger et al. (1977) by using a ^{35}S-labeled nucleotide and buffer-gradient polyacrylamide gels (Biggins et al. 1983). A 6,423-bp portion of the *Xdh* region was sequenced. A diagram of this region and the corresponding region in *D. melanogaster* is given in figure 1. The complete *D. pseudoobscura* nucleotide sequence, translation of coding sequence, and aligned base pairs in *D. melanogaster* are given in figure 2.

The DNA sequence was compiled using the programs of Staden (1982, 1984) and was analyzed using the DNA sequence analysis package of Pustell and Kafatos (1984). The XDH protein sequence was analyzed using the hydropathy algorithms of Hopp and Woods (1981).

The *D. pseudoobscura* sequence was aligned with the corresponding region from *D. melanogaster* (Keith et al. 1987; Lee et al. 1987). The alignment was accomplished by dividing the entire *D. pseudoobscura* sequence into blocks of 100 bp and searching for $\geq$50% similarity with search windows of 50, 10, and 4 nucleotides. Positions were assigned based on best overall sequence similarity for each region. Insertions and deletions were positioned according to this criterion. This method of alignment is intended to give the highest level of similarity between the two sequences.

The numbering of the *D. melanogaster* sequence begins at an internal *Eco*RI site. The numbering of the *D. pseudoobscura* sequence begins at the 5′ end of the 6,423-bp region sequenced. Uncertainties regarding the site of transcription initiation preclude the use of a standard sequence numbering scheme. As a point of reference, the AUG codon in *D. melanogaster* (bp −1407) corresponds to bp 1086 in the *D. pseudoobscura* sequence presented in figure 2.

The nucleotide differences observed between the two coding sequences were analyzed using the programs of Lewontin (1988). This analysis distinguishes between observed and inferred nonsynonymous (replacement changes) and synonymous differences. Inferred differences are scored when multiple hits in a single codon require that previous substitutions occur. The most conservative path, i.e., that path involving the least number of replacement changes, is chosen when multiple paths are possible. The program divides synonymous codons into four groups (two-, three-, four-, and

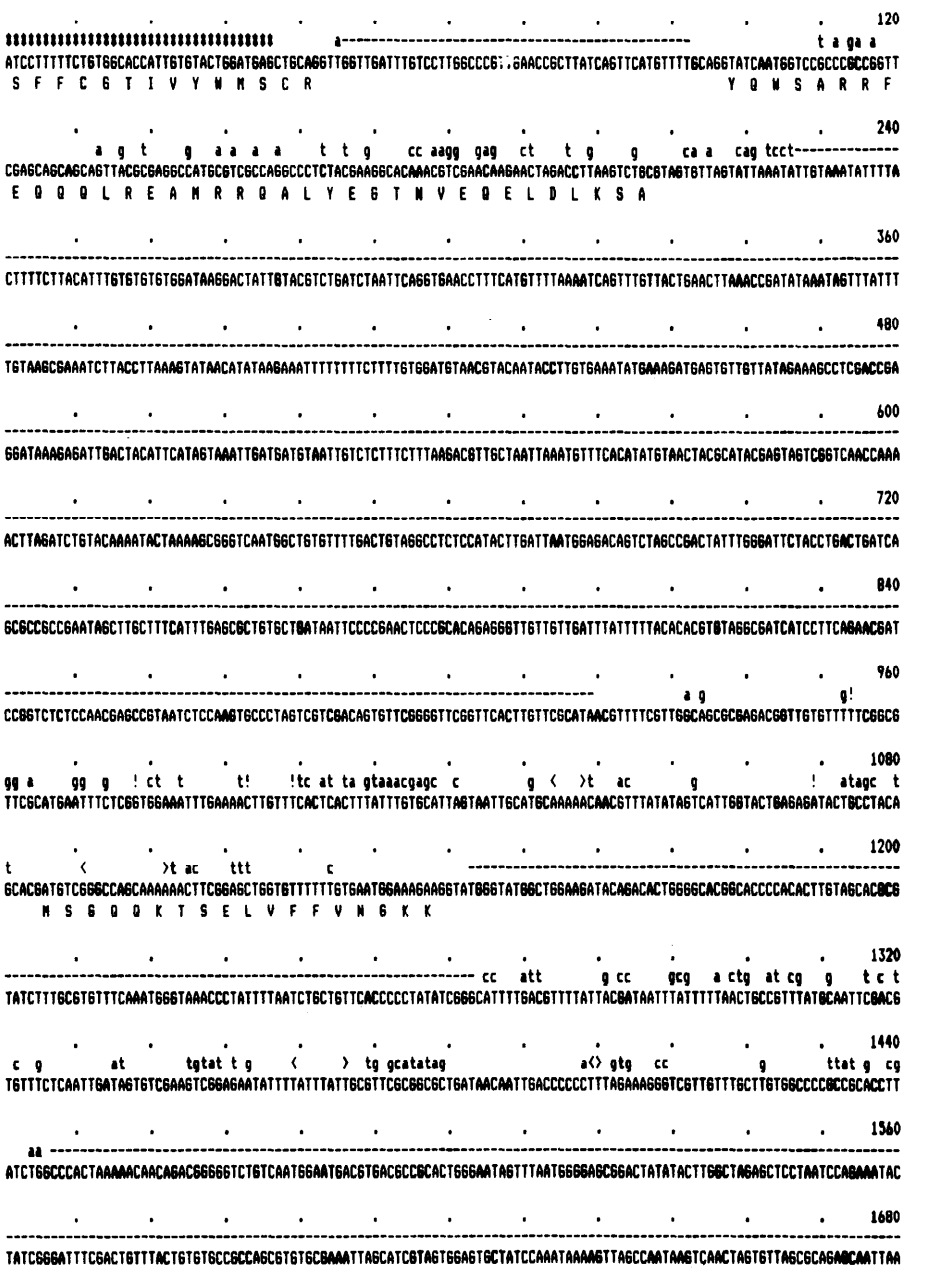

FIG. 2.—DNA sequence of the *Xdh* region in *Drosophila pseudoobscura* and predicted protein sequence. The asterisks (*; on left side of top row) indicate regions not compared with *D. melanogaster*. Dashes (-) indicate regions too divergent for confident *D. melanogaster* sequence alignment. Letters a, t, c, and g indicate nucleotide substitutions in *D. melanogaster* sequence. Angle brackets indicate deletions in *D. melanogaster* sequence relative to *D. pseudoobscura*. Exclamation marks (!) indicate insertions in *D. melanogaster* sequence relative to *D. pseudoobscura*. The first codon of each *Xdh* exon is boxed and labeled, and the donor splice junction of each intron is underlined. The putative *l(3)s12* region extends from bp 1 through bp 209.

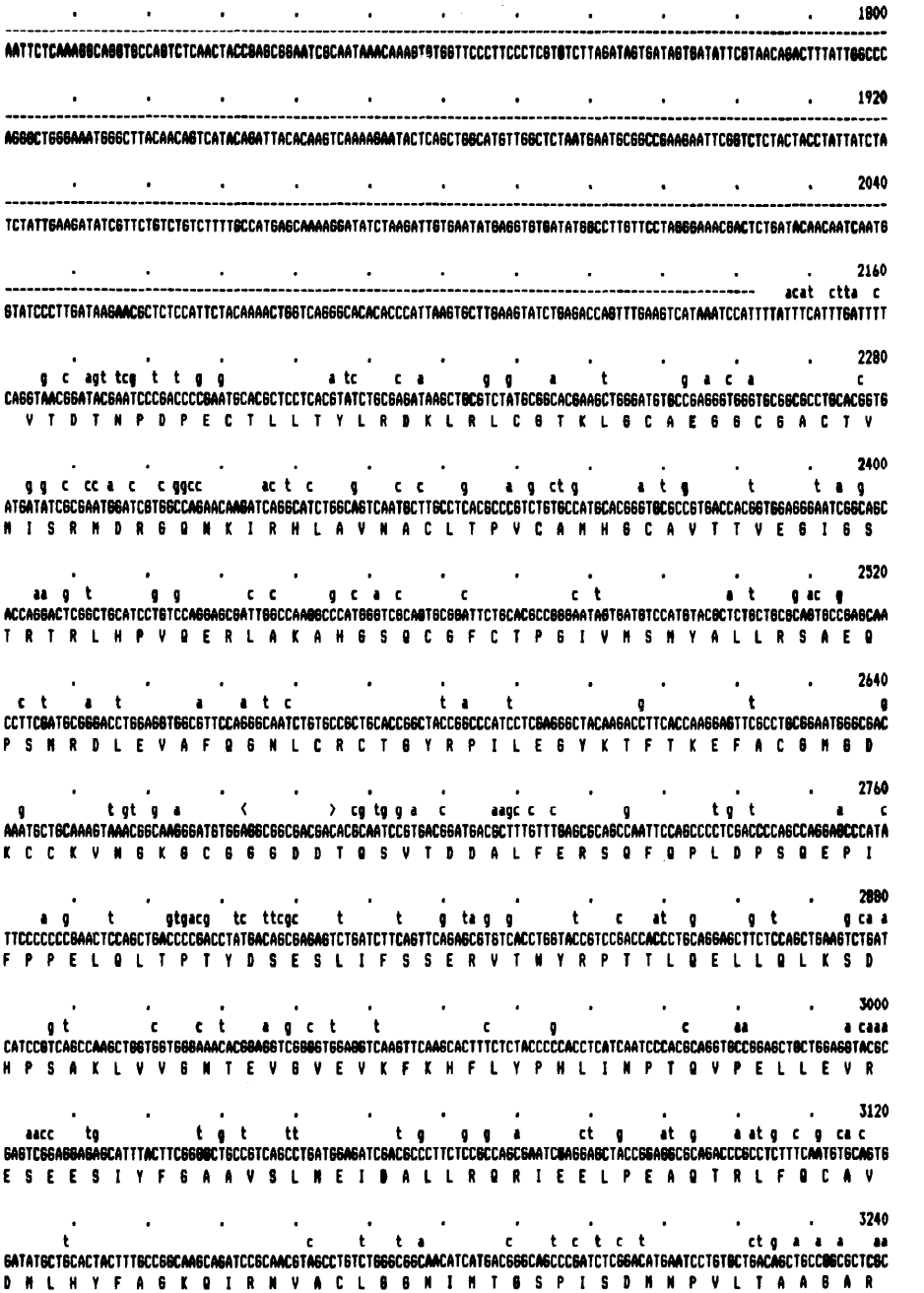

FIG. 2 (*Continued*)

sixfold degenerate) and scores the number of transition and transversion events per synonymous codon group. An estimate for the mean number of evolutionary events per codon within each group of synonymous codons is calculated, and a 95% confidence interval for this estimate is provided.

Five additional loci, sequenced in both *D. melanogaster* and *D. pseudoobscura*,

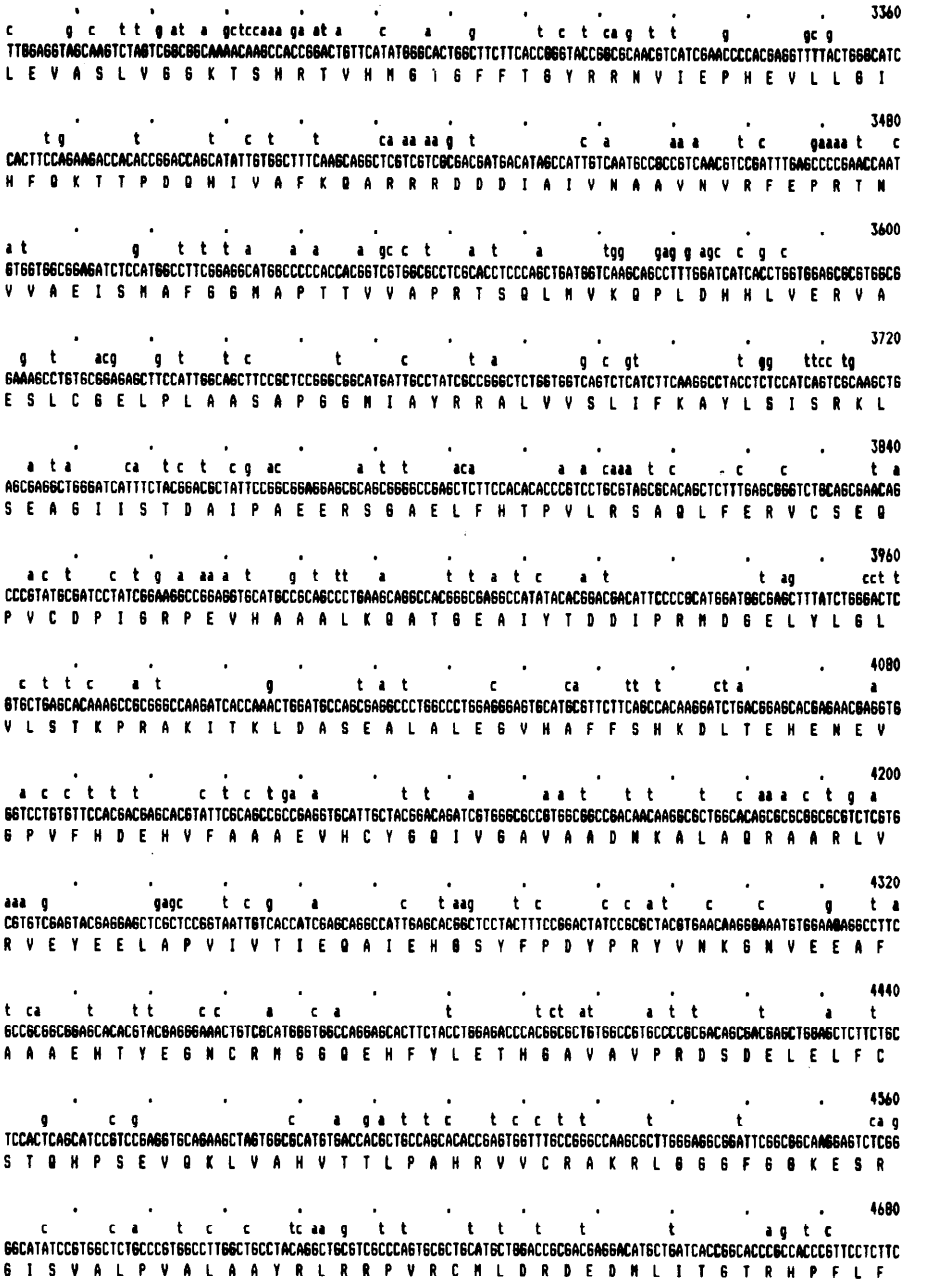

FIG. 2 (*Continued*)

were analyzed using this program. They are *Adh* (Schaeffer and Aquadro 1987), *Hsp*82 partial exon 2 (Blackman and Meselson 1986), the 5′ exon of *Ubx* (Wilde and Akam 1987), the *Gart* locus (Henikoff and Eghtedarzadeh 1987), and the putative *l(3)s12* locus described in the present paper.

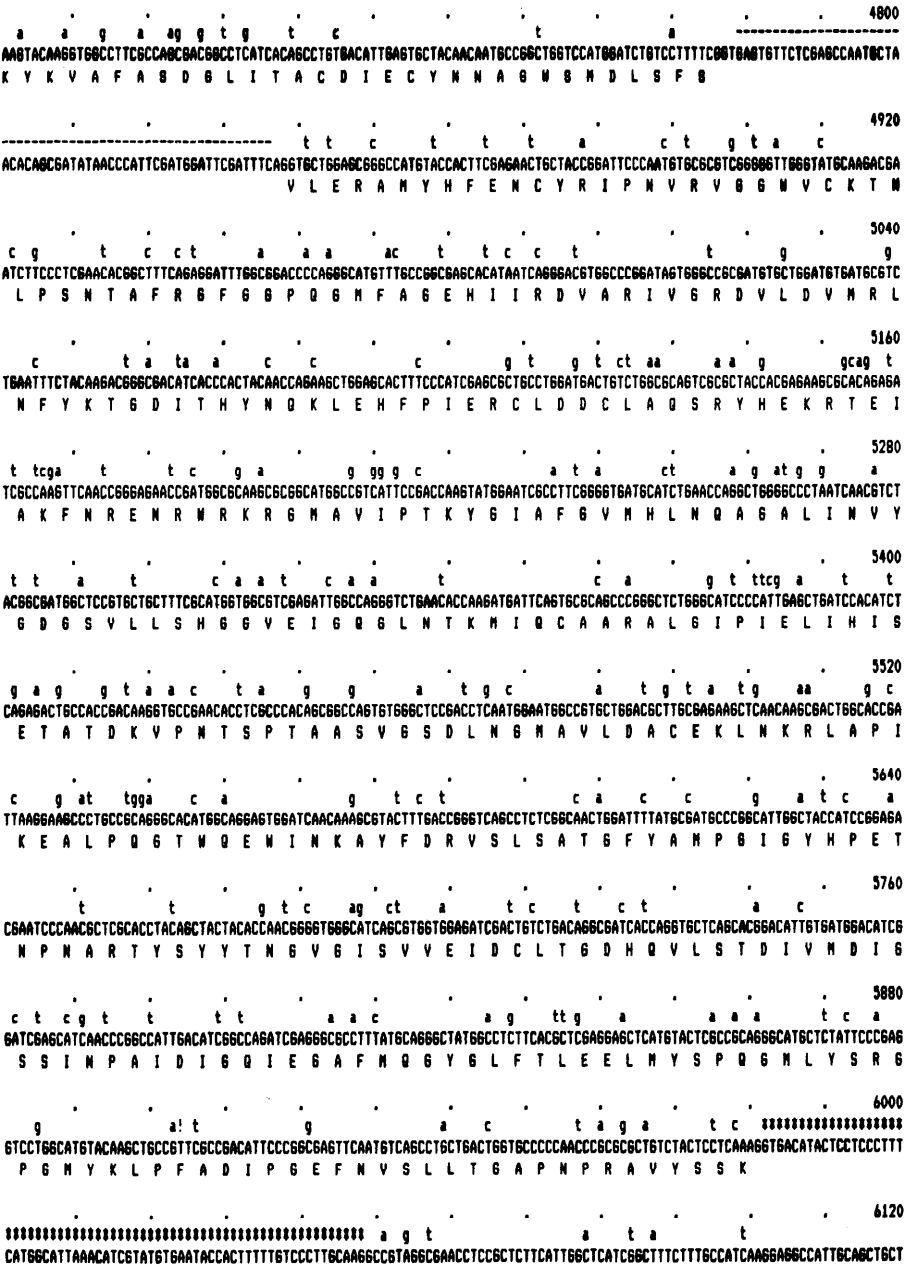

FIG. 2 (*Continued*)

Results

A. Putative *l(3)s12* Comparisons and 5′ Intergenic Region

Molecular and genetic mapping studies in *Drosophila melanogaster* have placed the lethal complementation group, *l(3)s12* (Hilliker et al. 1980), in a position 5′ to the *Xdh* locus and within an 8.1-kb *Sal*I fragment surrounding the *Xdh* region (Clark and Chovnick 1986). In an attempt to further localize and characterize this locus, the

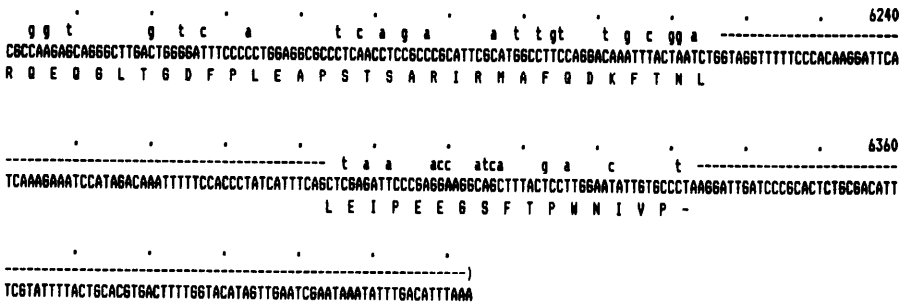

FIG. 2 (*Continued*)

sequence 5′ to the *Xdh* AUG codon was examined for potential open reading frames (ORF). Very strong nonrandom use of frame-two codons in the region of nucleotide positions 1–181 was found with a significant loss of codon bias between positions 61 and 93 (not shown).

A 42-bp ORF extends from position 1 in *D. pseudoobscura* to position 42, where a putative intron donor splice junction is encountered. The 54-bp intron terminates at an acceptor splice junction located at position 96. A second ORF picks up in frame with the first and extends 112 bp to a termination codon (TAG). These putative intron-exon boundaries correspond precisely to the regions of nonrandom codon use described above, and intron/exon splice junction sequences agree well with the consensus sequences derived by Mount (1982) and Keller and Noon (1985). The aligned *D. melanogaster* sequence displays very strong (75%) base-pair conservation in the proposed exons; five nonsynonymous and 22 synonymous nucleotide differences are observed (fig. 2).

The location of the described exons, the strong nonrandom codon usage corresponding precisely to the predicted exon-intron boundaries, and strong coding region and donor/acceptor splice sequence conservation between the species suggest that nucleotide positions 1–208 in the *D. pseudoobscura* sequence contain the 3′ portion of the *l(3)s12* locus. The region extending 3′ from the putative *l(3)s12* coding region to the *Xdh* AUG codon contains no additional long ORFs.

The 877 bp of noncoding sequence between the *Xdh* AUG codon and the putative *l(3)s12* sequence was compared with the corresponding 1,330 bp in the *D. melanogaster* sequence. The high level of nucleotide divergence and insertion/deletion variation precludes precise sequence alignment except for the 166-bp sequence immediately 5′ to the *Xdh* AUG codon (fig. 2), which exhibits 75% sequence conservation (excluding insertion/deletion) in this maximally aligned stretch. Blocks of similarity allow the inference of insertion/deletion events.

B. *Xdh* Intron Comparisons

Intron positions and splice junction sequences in the *D. pseudoobscura Xdh* coding region were inferred from comparison with the *D. melanogaster* sequence. Intron 1 consists of 1,024 bp in *D. pseudoobscura* and 815 bp in *D. melanogaster.* There are several blocks of striking sequence conservation in the 5′ half of the intron. Employing a search window of 150 bp with 50% sequence identity produces the alignment of the *D. pseudoobscura* intron sequence from bp 1264 through 1446 with the *D. melanogaster*

sequence from bp −1297 through −1127, an alignment displayed in figure 2. This high conservation results from blocks of perfectly conserved sequences interspersed with highly divergent sequences. An alignment of 62 nucleotides 3′ to the intron donor sequence is given for comparison. This very low level of sequence conservation, precluding meaningful sequence alignment, is representative of the entire intron, except for the conserved blocks indicated above.

In contrast to the 5′ portion, the remainder of the intron is saturated with nucleotide differences and cannot be aligned. The presence of the conserved blocks of sequence in the 5′ half of the intron does, however, reveal that at least four separate insertion/deletion events are responsible for the intron length variation observed.

Intron 2 sequences are aligned with one *D. melanogaster* insertion event inferred. The resulting 62 aligned base pairs are 63% divergent. An additional intron, 2A, is inferred in the *D. pseudoobscura* sequence because a 66-bp insertion, containing numerous stop codons, interrupts the coding sequence. Consensus intron donor and acceptor splice sequences are present at both ends of the insertion, and perfect sequence alignment is observed 5′ and 3′ to the putative intron. Intron 3 alignment results in a 62% nucleotide divergence estimate. Taken together, intron 2 and 3 levels of divergence predict that, on average, intron positions have experienced a mean number of 1.34 substitutions/site (Jukes and Cantor 1969). Substitution saturation of intron positions would occur at 75% observed divergence or at a mean number of substitutions per site approaching 3.0. The value of 1.34 should be taken to illustrate the least number of hits per intron position, since the alignment procedure was designed to produce optimal sequence similarity. Because of the very high levels of divergence reported for these introns and because of the very different intron sizes between the two species, only the alignment for the 5′ half of intron 1 is included in figure 2.

C. XDH Amino Acid Sequence Comparisons

There are 143 differences observed in the 1,334 amino acids compared between *D. melanogaster* and *D. pseudoobscura.* These differences are not distributed uniformly across the amino acid sequence when analyzed as six equal blocks of sequence ($\chi^2 = 15.2$, df = 5, $P < 0.01$). There are more replacement differences in the region between *D. pseudoobscura* amino acids 445 and 667 and fewer differences in the region from position 1112 to the end of the sequence. Of the amino acid changes deduced from the *D. pseudoobscura* and *D. melanogaster Xdh* sequences, 51.4% are charge conservative, 28.8% are charge changes, and 19.9% are polar changes. The *D. pseudoobscura* XDH protein has one fewer acidic residue and two additional basic residues relative to XDH in *D. melanogaster.* The pI of the *D. pseudoobscura* XDH polypeptide, 6.39, is very similar to that found in *D. melanogaster,* 6.48.

The *Xdh* amino acid differences were examined with reference to their occurrence in the hydrophobic and hydrophilic domains of the protein, as predicted from the hydropathy algorithms of Hopp and Woods (1981) (not shown). The most dramatic result is that the overall hydrophobicity and the specific locations of the hydrophobic regions of the protein in *D. pseudoobscura* do not change, even though 11% of the protein sequence is different from that in *D. melanogaster.* The hydrophilic regions experienced the most obvious modifications, involving primarily increases or decreases in the strength of hydrophilic domains.

Table 1
Summary of Differences Observed between *Drosophila melanogaster* and *D. pseudoobscura* for the Genes *Xdh, Adh, Hsp*82, *Ubx, Gart,* and *l(3)s12*

Locus	No. of Codons Compared	No. (%) of AA Diff.	No. of Syn Codons[a]	No. of Syn Sites	No. (%) of Syn Diff.
Xdh	1334	143 (10.7)	1,191	1,371	568 (41.4)
Adh	254	21 (8.7)	233	261	76 (29.1)
*Hsp*82	375	13 (3.5)	362	401	97 (24.2)
Ubx	248	23 (9.3)	225	236	75 (31.8)
Gart	1,353	196 (14.5)	1,157	1,367	583 (42.7)
l(3)s12	39	5 (12.8)	34	44	21 (47.7)

NOTE.—AA = amino acid; Diff. = differences; Syn = synonymous.
[a] Codons with at least one Syn site.

D. *Xdh* Nucleotide Substitution Analysis

1. *Distribution of Nucleotide Differences*

The two protein coding sequences, excluding two insertions of 12 bp each in *D. pseudoobscura* and one insertion of 3 bp in *D. melanogaster,* were examined for differences in the number of nucleotide substitutions, according to the method of Lewontin (1988).

The results of this analysis for *Xdh* are presented in tables 1 and 2, along with the results for five additional loci that have been sequenced in both *D. melanogaster* and *D. pseudoobscura.* No significance tests have been performed on the *l(3)s12* data because of the small sample size.

Since the divergence of the species, *Xdh* has accumulated 711 nucleotide differences in the 4,005 bp of homologous coding region. The distribution of these differences, analyzed in six equal blocks of sequence, is not significantly heterogeneous ($\chi^2 = 9.0$, df = 5, $P > 0.1$). Of the synonymous positions, 41.4% differ, again showing no significant heterogeneity ($\chi^2 = 2.6$, df = 5, $P < 0.7$). The breakdown of synonymous differences into two-, three-, four-, and sixfold-degeneracy groups is presented in table 2.

2. *Transition and Transversion Synonymous Events*

The expected number of transition and transversion events for each synonymous codon group in *Xdh* is calculated from the number of differences observed within a degeneracy group and from the predicted ratio of transition-to-transversion events per degeneracy group, a prediction based on the organization of the genetic code and corrected for *Drosophila* codon usage bias. For example, degeneracy group 3 experienced 18 transition and 10 transversion events. Based simply on the number of synonymous substitutions possible for the threefold-degeneracy class, a 2:4 ratio of transition-to-transversion substitutions is predicted. However, when actual codon frequencies are examined, a 2.9:3.1 ratio of events is predicted (see table 3). This new ratio is used in generating the expected number of transition-to-transversion events for degeneracy group 3.

For groups 2, 3, or 4, there is no bias in the number of transition and transversion events observed that cannot be explained by the organization of synonymous codons in the genetic code (group 3: $\chi^2 = 3.3$, df = 1, $P > 0.05$; group 4: $\chi^2 = 0.4$, df = 1, P

Table 2
Details of Synonymous Differences between *Drosophila melanogaster* and *D. pseudoobscura* for the Genes *Xdh*, *Adh*, *Hsp*82, *Ubx*, *Gart*, and *l(3)s12*

Group and Locus	No. of Syn Sites	No. of Syn Diff.[a]	No. of Transition Diff.[a]	No. of Transversion Diff.[a]	Hits[b]	95% Confidence Interval
2:						
Xdh.....	438	154 (17)	154 (17)		0.6	0.48–0.81
Adh.....	77	15 (1)	15 (1)		0.3	0.12–0.43
*Hsp*82....	178	23 (1)	23 (1)		0.2	0.08–0.22
Ubx.....	80	25 (1)	25 (1)		0.5	0.28–0.92
Gart	388	135 (18)	135 (18)		0.6	0.50–0.80
l(3)s12....	15	6 (0)	6 (0)		0.9	0.20–5.05
3:						
Xdh.....	65	28 (3)	18 (0)	10 (3)	0.8	0.47–2.11
Adh.....	21	2 (1)	2 (1)		0.1	0.00–0.30
*Hsp*82....	28	11 (0)	11 (0)		0.9	0.28–0.99
Ubx.....	6	3 (0)	2 (0)	1 (0)	1.0	0.11–0.99
Gart	55	25 (2)	11 (2)	14 (0)	0.7	0.45–1.10
l(3)s12....	0	0 (0)	0 (0)			
4:						
Xdh.....	469	258 (37)	81 (11)	177 (26)	1.0	0.85–1.22
Adh.....	101	47 (3)	21 (0)	26 (3)	0.8	0.52–1.11
*Hsp*82....	107	47 (2)	19 (2)	28 (0)	0.7	0.47–0.97
Ubx.....	119	41 (9)	13 (0)	28 (9)	0.5	0.32–0.66
Gart	481	274 (61)	89 (18)	185 (43)	1.1	0.95–1.30
l(3)s12....	7	3 (1)	0 (0)	3 (1)	0.7	0.10–5.05
6:						
Xdh.....	360	128 (19)	46 (3)	82 (16)	1.1	0.88–1.55
Adh.....	56	12 (5)	7 (1)	5 (4)	0.6	0.26–1.57
*Hsp*82....	78	16 (5)	8 (0)	8 (5)	0.6	0.32–1.16
Ubx.....	22	6 (6)	1 (1)	5 (5)	0.5	0.08–1.46
Gart	420	149 (17)	73 (7)	76 (10)	1.3	1.00–1.65
l(3)s12....	20	12 (1)	5 (0)	7 (1)	5.0	2.75–5.05

NOTE.—Abbreviations are as in table 1.

[a] Numbers in parentheses are number of inferred synonymous substitutions.

[b] Number of evolutionary events per codon, calculated using the mean of the codon usage between *D. melanogaster* and *D. pseudoobscura* (see Lewontin 1988).

> 0.5). Group 6 codons did exhibit barely significant transversion bias at arginine codons but not at leucine codons (Leu: $\chi^2 = 2.7$, df = 1, $P > 0.05$; Arg: $\chi^2 = 3.0$, df = 1, $P < 0.05$).

3. *Comparison of Synonymous Differences and Mean Number of Evolutionary Events between Synonymous Codon Groups*

Equivalency of observed substitution levels between the four groups of synonymous codons at *Xdh* was tested. A goodness-of-fit test that corrects for both the different number of codons per synonymous codon group and the expected ratio of differences per group was employed. On the basis of the degeneracy of synonymous positions within each synonymous codon group and under the assumption that each synonymous codon experiences the same mutation and selection pressures, groups 2, 3, and 4 are expected to have a ratio of differences of 1:2:3. Group 6 consists of 54% group 2–like

Table 3
Organization of the Genetic Code by Synonymous Codon Group

Degeneracy Group	No. of AA Represented	Total No. of Codons[a]	Expected No. of Events[b]	
			Transition	Transversion
2	10	20 (0.34)	20	
3	1	3 (0.05)	2 (2.91)[c]	4 (3.1)
4	6	24 (0.41)	24	48
6	2 Leu	6 (0.10)	10 (9.7)	8 (8.3)
	Arg	6 (0.10)	6 (6.2)	12 (11.8)

[a] Numbers in parentheses are percentage of total synonymous codons represented by this group.
[b] Expected number of transition-to-transversion events, under the assumption of equal synonymous codon usage.
[c] Numbers in parentheses are the number, weighted by *Drosophila* codon bias, employed in transition-to-transversion bias χ^2 expectations.

and 46% group 4–like codon equivalents (i.e., certain sites are two- or fourfold degenerate). The group 6 synonymous positions were partitioned into 2-like and 4-like classes of synonymous codon equivalents, and these were treated separately in the test. This assumes no interaction between multiple substitution events within a group 6 codon. The test was significant at the $P < .001$ level ($G = 41.4$; $df = 4$). The group 2 differences and group 6 4-like differences contributed most to this highly significant result. Group 2 appears to have twice as many differences as expected. Group 6 4-like codons appear to have fewer differences than expected.

A more detailed comparison of the synonymous changes observed between the species at group 2 and group 4 codons was undertaken (table 4). Both groups are shown to contribute equally to the total composition of codons in the *D. melanogaster* sequence and to the amount of amino acid replacements observed between the species. There is, however, a high level of synonymous substitution observed at group 2 codons ending in T and A, compared with similar group 4 codons. In addition, group 2 codons ending in G exhibit very low levels of synonymous substitution. Overall, group 4 codons have very similar levels of synonymous substitution per T-, C-, G-, and A-ending codon, while group 2 codons have highly variable levels of substitution across the four types.

The estimates for the mean number of evolutionary events per synonymous codon group in *Xdh* range from 0.6 to 1.1 (table 2). The expected mean number of evolutionary events differs between codon groups, owing to the nature of synonymous

Table 4
Comparison of Group 2 and Group 4 Synonymous Codon Substitution at *Xdh*

	T	St	C	Sc	G	Sg	A	Sa	Total Codons	Total Syn	Total Replace
Group 2	0.26	0.47	0.34	0.25	0.30	0.06	0.10	0.21	0.38	0.27	0.36
Group 4	0.19	0.28	0.36	0.24	0.27	0.19	0.18	0.28	0.39	0.45	0.34

NOTE.—T, C, G, and A = % of all codons in *Drosophila melanogaster* that end in T, C, G, or A, respectively; St, Sc, Sg, and Sa = % of each subset of codons in *D. melanogaster* substituted in *D. pseudoobscura*. Total Codons = % of all codons in groups 2 or 4 in *D. melanogaster*. Total Syn = % of all synonymous codons in groups 2 or 4 in *D. melanogaster*. Total Replace = % of all replacement codons in groups 2 or 4 in *D. melanogaster*.

site distribution within each group. The estimated number of events per codon, corrected for multiple substitution events, indicates that all synonymous sites are below the level of saturation for nucleotide changes.

Goodness-of-fit tests were performed on the four additional loci described above. *Adh* and *Hsp*82 are shown to have similar levels of differences at all synonymous codon groups (*Adh:* $G = 5.6$, $df = 3$, $P > 0.10$; *Hsp*82: $G = 2.2$, $df = 3$, $P > 0.5$), while *Ubx* and *Gart* are observed to have significant heterogeneity in observed numbers of differences between synonymous codon groups (*Ubx:* $G = 14.5$, $df = 3$, $P < 0.01$; *Gart:* $G = 32.38$, $df = 3$, $P < 0.01$).

4. *Comparison of Synonymous-Substitution Levels between Different Genes*

It is clear from tables 1 and 2 that the number of synonymous differences is significantly higher in *Xdh* and *Gart* than in the three additional loci presented. However, heterogeneity in rates among genes is contributed to almost equally by the low level of *Hsp*82 synonymous differences ($G = 47.0$, $df = 4$, $P < 0.01$).

There is significant variation among the group 2 synonymous changes when all four loci are included in the analysis ($G = 32.3$, $df = 4$, $P < 0.01$). *Xdh, Hsp,* and *Gart* contribute most to the significance of this test. Group 3 shows nonsignificant heterogeneity across the four loci ($G = 7.8$, $df = 4$, $P > 0.05$). Groups 4 and 6 exhibit barely significant heterogeneity (Group 4: $G = 13.0$, $df = 4$, $P < 0.02$; Group 6: $G = 11.8$, $df = 4$, $P < 0.02$).

The group 4 codons of the four additional loci were examined for departures from expected transition-to-transversion bias. As was seen at *Xdh,* there is no significant bias in transition-to-transversion events observed (*Adh:* $\chi^2 = 2.7$, $df = 1$, $P > 0.1$; *Hsp*82: $\chi^2 = 1.0$, $df = 1$, $P > 0.3$; *Ubx:* $\chi^2 = 0.1$, $df = 1$, $P > 0.8$; *Gart:* $\chi^2 = 0.1$, $df = 1$, $P > 0.05$).

5. *Comparison of Codon Bias at the Xdh Locus*

The *D. melanogaster* and *D. pseudoobscura Xdh* codon biases (see Lewontin 1988, table 2) were compared in a goodness-of-fit test, by using the null hypothesis that both species have the same codon bias at *Xdh.* This test compares, between the two species, the number of occurrences of a specific codon with the sum of occurrences of all codons within an amino acid codon group. The G value obtained ($G = 106$, $df = 40$) was significant at the $P < 0.001$ level.

Discussion

A. *Xdh* Region Sequence Comparisons Reveal 5′ Gene, Additional *Xdh* Intron, and Putative Regulatory Sequences

The aligned sequences reveal a similar organization of the *Xdh* locus in the two species. However, an additional intron was identified in *Drosophila pseudoobscura,* and three separate insertion/deletion events were inferred in the coding sequence.

Conserved sequence 5′ to the *Xdh* coding region reveals an additional locus, tentatively assigned as *l(3)s12.* In addition, blocks of similar sequence immediately 5′ to the *Xdh* AUG codon are noted. This region presumably contains the transcription initiation site and leader sequence, which may account for the 75% sequence conservation seen in this maximally aligned stretch. However, numerous insertion and deletion events are predicted from this alignment. Insertion/deletion events in nontrans-

lated leader sequences have also been observed in *Adh* species comparisons (Schaeffer and Aquadro 1987).

Keith et al. (1987) note the absence of TATA or CAAT promoter elements in appropriate positions 5′ to the putative initiation region in *D. melanogaster.* This observation is now extended to the *D. pseudoobscura* sequence comparison. No consensus, or clear derivatives, of the commonly encountered promoter elements are located within 300 bp of the corresponding *D. pseudoobscura* transcription initiation region.

The 5′ half of intron I also contains several blocks of conserved sequences. A mutant allele of *Xdh* in *D. melanogaster* has reduced expression of *Xdh* in fat body of adults and larvae. This allele has an ~400-bp deletion in the 5′ region of the long first intron of *Xdh* (M. McCarron and A. Chovnick, personal communication). In addition, there is an *Xdh* overexpressor line which also exerts its effects largely in the fat body and has been genetically placed within the limits of the first intron (S. H. Clark and A. Chovnick, personal communication). These observations indicate that regulatory regions essential for proper qualitative tissue-specific expression of *Xdh* may lie within the large first intron. One possible candidate for the tissue-specific regulatory sequence may lie in the 182-bp highly conserved region.

Sequence comparisons between *D. melanogaster* and *D. pseudoobscura* are increasingly being used to localize sequences important in gene regulation. The high level of sequence divergence in noncoding regions allows clear identification of conserved sequences. Comparisons of this sort have resulted in the localization of previously uncharacterized genes (Cohn 1985; Schaeffer and Aquadro 1987) and in the description of sequences that may be involved in gene expression and regulation (Blackman and Meselson 1986; Wilde and Akam 1987).

The XDH protein has accumulated 143 differences between the two species. Even with 11% of the protein sequences different and with these amino acid differences clustered in the first half of the sequence, there is conservation of both the overall charge of the proteins and of the pattern of distribution of hydrophobic and hydrophilic domains.

Population surveys of electrophoretically distinguishable protein variation indicate that *Xdh* is one of the most highly polymorphic enzymes in *Drosophila* (Coyne 1976; Buchanon and Johnson 1983; Keith et al. 1985). In addition, it has been suggested (Lewontin 1985) that the polymorphism observed at *Xdh* may reflect variation essentially neutral to natural selection. The conservation of hydrophobic domains, the clustering of amino acid substitutions, and the conservation of overall charge between the *Xdh* sequences examined argue for stronger functional constraints on the protein structure than has been suggested in the past.

B. Overall Substitution Levels Are Different in Different Genes

The genes *l(3)s12, Xdh,* and *Gart* have accumulated significantly more nucleotide substitutions since the divergence of the species than has *Adh, Hsp*82, or *Ubx.* This higher level of nucleotide substitution is not confined to any particular position at either locus but, rather, appears as a corresponding increase in observed numbers of differences at replacement sites and at all types of synonymous sites. Possible explanations for this increased level of nucleotide substitution include higher mutation rates for certain regions of the chromosome, lower levels of functional constraint at all nucleotide positions in this region, and the result of a sampling artifact.

The first hypothesis, that of differential mutation rates across the *Drosophila*

genome, has been additionally implicated from the work on total single-copy genomic DNA (Zwiebel et al. 1982), the 68C glue gene cluster (Meyerowitz and Martin 1984), and *Adh* (Kreitman 1983; Schaeffer and Aquadro 1987).

The sampling hypothesis arises because *Xdh* has been shown to be a highly polymorphic locus in natural populations. Partial *Xdh* nucleotide sequences of several additional, electrophoretically distinguishable, alleles isolated in the survey of Keith et al. (1985) indicate that the high level of substitution observed in the species comparison is not the result of within-population differences (M. Riley, unpublished data). That is, all of the sequenced alleles would give the same high level of sequence divergence estimates in the species comparison.

The percent of synonymous substitution levels varies among the four genes and is loosely correlated with amino acid replacement levels. A similar correlation has previously been reported for mammals (Li et al. 1985). This suggests the possibility that regional differences in mutation rates influence rates of nucleotide evolution in addition to positional differences in levels of constraint. However, it is also possible that this correlation is spurious. *Adh,* for example, clearly shows nonrandom distribution of nonsynonymous and synonymous differences across the coding region (Bodmer and Ashburner 1984; Schaeffer and Aquadro 1987). Certain exons are deficient in both replacement and synonymous differences, implying different levels of sequence constraint experienced across the coding region for these two types of substitutions. *Xdh,* on the other hand, shows nonrandom distribution of replacement—but not of synonymous—differences. These results indicate that the level of mutation and/or the specific selection pressures experienced at replacement and synonymous sites within these two loci must be different.

C. Levels of Synonymous Substitution Are Different in Different Synonymous Codon Groups at *Xdh*

The method of nucleotide sequence analysis employed in the present study separates synonymous codons into groups and provides an estimate of mean number of evolutionary events for each codon group. The importance of this parceling process becomes apparent when within- and between-gene comparisons of synonymous substitution levels are made.

Xdh was examined for equivalency of substitution levels in the two-, three-, four-, and sixfold-degenerate codon groups. Significantly different levels of substitution are observed across the five groups; group 6 codons are divided into 2-like and 4-like codon equivalents. Higher levels of synonymous substitutions are observed at the twofold-degenerate sites. Groups 3, 4, and 6 (2-like) appear to be experiencing very similar levels of synonymous substitution, and group 6 (4-like) codons are observed to have too few substitutions. The expected number of group 6 substitutions employed in the test is only an approximation. Interactions between multiple synonymous substitutions within the same codon have not been considered, and it will certainly be true that the starting codon influences the paths available for synonymous substitution. What these differing levels of nucleotide substitution mean concerning the functional constraints at particular synonymous positions within the gene is unclear. One hypothesis, as yet untested, is that mRNA stability may exert a level of constraint in addition to codon usage bias. Certainly, unexpected complexity is indicated by these results.

D. Levels of Synonymous Substitutions Are Different in Different Genes

It has frequently been stated that synonymous substitution rates are nearly equal for different genes (King and Jukes 1969; Jukes 1980; Miyata et al. 1980; Hayashida and Miyata 1983; although see Kimura 1987) and that these rates may be close approximations of the neutral mutation rate (Kimura 1977; Nei 1987, p. 83). The analysis presented here shows that there are significantly different, gene-specific, levels of synonymous substitution. *Xdh* and *Gart* are observed to have a much higher level of synonymous substitution in all synonymous codon groups. Group 2 codons were substituted at much lower levels in *Hsp*82 and *Adh,* group 3 codons were substituted equally in all four genes, and group 6 codons were substituted at lower levels in *Adh* and *Hsp*82. This result suggests that there are unexpected selective pressures and/or mutational forces operating differently on different types of synonymous positions within and between different genes. It is clear that synonymous positions evolve more rapidly than replacement positions, but it is not as clear how close an approximation to the neutral rate synonymous positions provide.

E. Transition and Transversion Rates of Synonymous Substitution

It has been suggested that in coding regions, if mutations occur at random, then we should expect twice as many transversion as transition events (e.g., see Nei 1987). Some observations of synonymous substitution levels in *Drosophila* are consistent with this predicted transition-to-transversion ratio of 0.5 (Kreitman 1983); others are not (Ashburner et al. 1984; Schaeffer and Aquadro 1987). In general, however, it has been reported that there is a large apparent bias in favor of transitional events in different organisms (Fitch 1967; Vogel 1972; although see Jukes 1987) and that this bias is present even when coding constraints are relaxed, e.g., in pseudogenes (Li et al. 1984). When synonymous positions are considered by groups, the predicted ratio is no longer 0.5. The organization of the genetic code predicts the overall transition-to-transversion bias for synonymous substitutions of 0.86, not the previously reported 0.5. As the five genes discussed in the present paper conform very closely to the predicted ratios of each type of codon and as they also conform, with the possible exception of *D. pseudoobscura Xdh,* very closely to the average codon bias in *Drosophila,* it would be predicted that, given equal probabilities of transition and transversion mutational events, all four genes should display ratios of 0.86. The overall transition-to-transversion ratio for the four genes is 1.13. This represents a significant bias in favor of transition events, observed to expected, of 1.30, so there are more transitional events than expected. At *Xdh,* groups 3 and 4 were shown not to deviate from expected ratios; therefore the bias (1.05) at this gene must result from the fact that, after correcting for actual codon frequencies in *Xdh,* there were too many group 2 substitutions compared with the number expected. The bias observed in group 6 arginine codons cannot account for the 1.05 overall bias, as group 6 was observed to have a transversion—not a transition—bias. The observed bias at *Adh* (1.26) must result from a different factor, as group 2 codons occur as frequently as in *Xdh* (37% of total synonymous codons) and are substituted at approximately one-half the level observed at *Xdh.* Thus, transitional biases are observed for the five genes, are not as large as previously suggested, and appear to be due to different causes, at least in *Xdh* and *Adh.*

F. Codon Usage Bias in *Xdh*

Codon usage bias has been observed in several organisms (Ikemura 1985; Sharp and Li 1986). It has been suggested that there is codon optimization, particularly in genes that are heavily transcribed, as a response to selection for efficiency of translation (e.g., see Gouy and Gautier 1982; Ikemura 1985). *Xdh* is transcribed at very low levels and might be expected to experience only weak codon bias selection pressures. It is interesting to note that *Xdh* codon usage in *D. melanogaster* and *D. pseudoobscura* is significantly different. In contrast, *Adh* is a very highly expressed gene, and, according to the optimization hypothesis, it is expected to display strong codon preference. Codon usage comparisons between *D. melanogaster* and *D. pseudoobscura* for *Adh* reveal similar codon bias in the two species (Schaeffer and Aquadro 1987), an indication that there may be codon optimization at *Adh.*

An additional explanation offered for the differences in codon bias pressures experienced at the two loci is that there is an interaction of silent and replacement changes in coding sequences, as suggested by Lipman and Wilbur (1985). The authors compared conserved with unconserved regions of the same proteins and found that poorly conserved regions have less-biased codon use. They suggest that replacement changes themselves are responsible for the reduced bias and that changing the context of adjacent codons could increase the rate of silent substitution in positions adjacent to the replacements.

It was noted for *Adh* that less well-conserved regions of the protein also tended to harbor an increased level of synonymous substitutions (Bodmer and Ashburner 1984; Schaeffer and Aquadro 1987). According to the Lipman and Wilbur hypothesis, the overall increased level of replacement substitutions at *Xdh,* relative to *Adh,* may account for the weaker codon bias observed, and the higher overall rate of synonymous substitution may follow from the disrupted codon context produced by these replacement changes.

G. Divergence Estimates Based on Synonymous Substitution Levels

Synonymous nucleotide substitution values have frequently been employed to estimate divergence time between *Drosophila* species (Ashburner et al. 1984; Bodmer and Ashburner 1984; Cohn 1985; Blackman and Meselson 1986; Moriyama 1987; Schaeffer and Aquadro 1987). The assumption inherent in these calculations is that there is a constant (or near constant) rate of silent substitutions, e.g., v, the silent rate constant of Hayashida and Miyata (1983) calculated to be 5.49×10^{-9} on the basis of a large number of mammalian genes. It is clear from the present analysis and previous work (Kreitman 1983; Gillespie 1986; Schaeffer and Aquadro 1987; Sharp and Li 1987) that silent substitution rates vary from gene to gene and between synonymous codon groups. Thus, estimates of divergence times employing silent rate constants are ballpark figures at best.

As there is no silent-rate constant available for *Drosophila* and since the levels of substitution experienced at different synonymous codons within and between genes is shown to vary, it seems unlikely that synonymous substitution rates are going to be useful in estimating times since divergence for species of *Drosophila.* It may be more informative to simply give corrected mean number of evolutionary events per synonymous codon group by using a common correction formula for species comparison. An average for each codon group can be obtained for each gene compared,

and these values can then serve as an indication of the level of divergence between different pairs of species.

In conclusion, the *Xdh* nucleotide substitution data suggest additional complexity that must be incorporated into any unifying theory of gene evolution in *Drosophila.* Constant rates of substitution at synonymous positions appear to represent averages across nucleotide positions experiencing a number of distinct selection pressures, including transition-to-transversion bias, codon usage bias, codon usage optimization, replacement/silent position interactions, and differential mutation rates. Because of the inherent difficulties in estimating the actual number of substitution events from the observed number (see Gillespie 1986), caution must be exercised in the interpretation of nucleotide substitution data when species as divergent as *D. melanogaster* and *D. pseudoobscura* are compared. However, these data point out patterns and complexities of substitution at functionally distinct positions that will require further analysis using population-level sequence comparisons, where the effect of multiple substitution events is insignificant.

Acknowledgments

I thank T. Keith for providing the fly stock and R. Lewontin, M. Kreitman, P. Sharp, S. Schaeffer, and M. Cummings for their comments on the manuscript. D. Curtis, W. Bender, and A. Chovnick provided information on *l(3)s12* and aided in the sequence alignment. This work was supported by grants from the National Institutes of Health to R. C. Lewontin.

LITERATURE CITED

ASHBURNER, M., M. BODMER, and F. LEMEUNIER. 1984. On the evolutionary relationships of *Drosophila melanogaster.* Devel. Genet. **4**:295–312.

BANKIER, A. T., and B. G. BARREL. 1983. Shotgun DNA sequencing. Laboratory of Molecular Biology, MRC Centre, Cambridge.

BIGGINS, M. D., J. J. GIBSON, and G. F. HONG. 1983. Buffer gradient gels and 35-S label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. USA **80**:3963–3965.

BLACKMAN, R. K., and M. MESELSON. 1986. Interspecific nucleotide sequence comparisons used to identify regulatory and structural features of the *Drosophila Hsp*82 gene. J. Mol. Biol. **188**:499–515.

BODMER, M., and M. ASHBURNER. 1984. Conservation and change in the DNA sequences coding for alcohol dehydrogenase in sibling species of *Drosophila.* Nature **309**:421–430.

BUCHANON, B. A., and D. L. E. JOHNSON. 1983. Hidden electrophoretic variation at the xanthine dehydrogenase locus in a natural population of *Drosophila melanogaster.* Genetics **104**:301–315.

CLARK, S. H., and A. CHOVNICK. 1986. Studies of normal and position-affected expression of *rosy* region genes in *Drosophila melanogaster.* Genetics **114**:819–840.

COHN, V. H. 1985. Organization and evolution of the alcohol dehydrogenase gene in *Drosophila.* Ph.D. diss., University of Michigan, Ann Arbor.

COYNE, J. A. 1976. Lack of genetic similarity between two sibling species of *Drosophila* as revealed by varied techniques. Genetics **84**:593–607.

COYNE, J. A., and M. KREITMAN. 1986. Evolutionary genetics of two sibling species *Drosophila simulans* and *D. sechellia.* Evolution **404**:673–691.

DALE, R. M. K., B. A. MCCLURE, and J. P. HOUCHINS. 1985. A rapid single-stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: application to sequencing the corn mitochondrial 18 S rDNA. Plasmid **13**:31–40.

FITCH, W. M. 1967. Evidence suggesting a non-random character to nucleotide replacements in naturally occurring mutations. J. Mol. Biol. **26**:499–507.

GILLESPIE, J. H. 1986. Variability of evolutionary rates of DNA. Genetics **113**:1077–1091.

GOUY, M., and C. GAUTIER. 1982. Codon usage in bacteria: correlation with gene expressivity. Nucleic Acids Res. **10**:7055–7074.

HAYASHIDA, H., and T. MIYATA. 1983. Unusual evolutionary conservation and frequent DNA segment exchange in class I genes of the major histocompatibility complex. Proc. Natl. Acad. Sci. USA **80**:2671–2675.

HENIKOFF, S., and M. K. EGHTEDARZADEH. 1987. Conserved arrangement of nested genes at the *Drosophila Gart* locus. Genetics **117**:711–725.

HILLIKER, A. J., S. H. CLARK, A. CHOVNICK, and W. GELBART. 1980. Cytogenetic analysis of the chromosomal region immediately adjacent to the rosy locus in *Drosophila melanogaster.* Genetics **95**:95–110.

HOPP, T. P., and K. R. WOODS. 1981. Prediction of protein antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci. USA **78**:3824–3828.

IKEMURA, T. 1985. Codon usage and tRNA content in unicellular and multicellular organisms (review). Mol. Biol. Evol. **2**:13–34.

JUKES, T. H. 1980. Silent nucleotide substitutions and the molecular evolutionary clock. Science **210**:973–978.

———. 1987. Transitions, transversions, and the molecular evolutionary clock. J. Mol. Evol. **26**:87–98.

JUKES, T. H., and C. R. CANTOR. 1969. Evolution of protein molecules. Pp. 21–132 *in* H. N. MUNRO, ed. Mammalian protein metabolism III. Academic Press, New York.

JUKES, T. H., and J. L. KING. 1979. Evolutionary nucleotide replacements in DNA. Nature **281**:605–606.

KAFATOS, F. C., A. EFSTRATIADIS, B. G. FORGET, and S. M. WEISSMAN. 1977. Molecular evolution of human and rabbit β-globin mRNAs. Proc. Natl. Acad. Sci. USA **74**:5618–5622.

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

KEITH, T. P., L. D. BROOKS, R. C. LEWONTIN, J. C. MARTINEZ-CRUZADO, and D. L. RIGBY. 1985. Nearly identical allelic distributions of xanthine dehydrogenase in two populations of *Drosophila pseudoobscura.* Mol. Biol. Evol. **2**:206–216.

KEITH, T. P., M. A. RILEY, M. KREITMAN, R. C. LEWONTIN, D. CURTIS, and G. CHAMBERS. 1987. Sequence of the structural gene for xanthine dehydrogenase (rosy locus) in *Drosophila melanogaster.* Genetics **116**:67–73.

KELLER, E. B., and W. A. NOON. 1985. Intron splicing: a conserved internal signal in introns of *Drosophila* pre-mRNA's. Nucleic Acids Res. **13**:4971–4981.

KIMURA, M. 1977. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. Nature **267**:275–276.

———. 1987. Molecular evolutionary clock and the neutral theory. J. Mol. Evol. **26**:24–33.

KING, J. L., and T. H. JUKES. 1969. Non-Darwinian evolution. Science **164**:788–798.

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

LEE, C. S., D. CURTIS, M. MCCARRON, C. LOVE, M. GRAY, W. BENDER and A. CHOVNICK. 1987. Mutations affecting expression of the *rosy* locus in *Drosophila melanogaster.* Genetics **116**:55–66.

LEWONTIN, R. C. 1985. Population genetics. Annu. Rev. Genet. **19**:81–102.

———. 1988. Inferring the number of evolutionary events from DNA coding sequence differences. Mol. Biol. Evol. **6**:15–32.

LI, W.-H., C.-I. WU, and C.-C. LUO. 1984. Nonrandomness of point mutations as reflected in nucleotide substitutions and its evolutionary implications. J. Mol. Evol. **21**:58–71.

———. 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide

substitution considering the relative likelihood of nucleotide and codon changes. Mol. Biol. Evol. **2**:150–174.

LIPMAN, D. J., and W. J. WILBUR. 1985. Interaction of silent and replacement changes in eucaryotic coding sequences. J. Mol. Evol. **21**:161–167.

MANIATIS, T., E. F. FRITSCH, and J. SAMBROOK. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

MEYEROWITZ, E. M., and C. H. MARTIN. 1984. Adjacent chromosomal regions can evolve at very different rates: evolution of the *Drosophila* 68C glue gene cluster. J. Mol. Evol. **20**:251–264.

MIYATA, T., T. YASUNAGA, and T. NISHIDA. 1980. Nucleotide sequence divergence and functional constraints in mRNA evolution. Proc. Natl. Acad. Sci. USA **77**:7328–7332.

MORIYAMA, E. N. 1987. Higher rates of nucleotide substitution in *Drosophila* than in mammals. Jpn. J. Genet. **62**:139–147.

MOUNT, S. M. 1982. A catalogue of splice junction sequences. Nucleic Acids Res. **16**:459–472.

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

PUSTELL, J., and F. C. KAFATOS. 1984. A convenient and adaptable package of computer programs for DNA and protein sequence management, analysis and homology determination. Nucleic Acids Res. **12**:643–655.

SANGER, F., S. NICKLER, and A. R. COULSON. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.

SCHAEFFER, S. W., and C. F. AQUADRO. 1987. Nucleotide sequence of the alcohol dehydrogenase region of *Drosophila pseudoobscura:* evolutionary change and evidence for an ancient duplication. Genetics **117**:61–73.

SHARP, P. M., and W.-H. LI. 1986. An evolutionary perspective on synonymous codon usage in unicellular organisms. J. Mol. Evol. **24**:28–38.

———. 1987. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. **15**:1281–1295.

STADEN, R. 1982. Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. Nucleic Acids Res. **10**:4731–4751.

———. 1984. A computer program to enter DNA gel reading data into a computer. Nucleic Acids Res. **12**:499–504.

VOGEL, F. 1972. Non-randomness of base replacement in point mutation. J. Mol. Evol. **1**:334–367.

WILDE, C. D., and M. AKAM. 1987. Conserved sequence elements in the 5′ region of the ultrabithorax transcription unit. EMBO J. **6**:1393–1401.

ZWIEBEL, L. J., V. H. COHN, D. R. WRIGHT, and G. P. MOORE. 1982. Evolution of single-copy DNA and the ADH gene in seven drosophilids. J. Mol. Evol. **19**:62–71.

BARRY G. HALL, reviewing editor

Received March 7, 1988; revision received July 15, 1988