J Mol Evol (2001) 53:485–495
DOI: 10.1007/s002390010239

JOURNAL OF MOLECULAR EVOLUTION

Xanthine Dehydrogenase (XDH): Episodic Evolution of a "Neutral" Protein*

Francisco Rodríguez-Trelles,[1,2] Rosa Tarrío,[1,3] Francisco J. Ayala[1]

[1] Department of Ecology and Evolutionary Biology, 321 Steinhaus Hall, University of California, Irvine, CA 92697-2525, USA
[2] Instituto de Investigaciones Agrobiológicas de Galicia (CSIC), Avenida de Vigo s/n, Apartado 122, 15780 Santiago de Compostela, Spain
[3] Misión Biológica de Galicia, CSIC, Apartado 28, 36080 Pontevedra, Spain

Received: 22 November 2000 / Accepted: 28 February 2001

Abstract. We investigated the evolution of *xanthine dehydrogenase* (*Xdh*) in 34 species from the three multicellular kingdoms, including one plant, two fungi, and three animal phyla, two classes of vertebrates, four orders of mammals, and two orders of insects. We adopted a model-based maximum-likelihood framework of inference. After accounting for among-site rate variation and heterogeneous nucleotide composition of the sequences using the discrete gamma distribution, and using nonhomogeneous nonstationary representations of the substitution process, the rate of amino acid replacement is 30.4×10^{-10}/site/year when *Drosophila* species are compared but only $\approx 18 \times 10^{-10}$/site/year when comparisons are made between mammal orders, between insect orders, or between different animal phyla and $\approx 11 \times 10^{-10}$/site/year when comparisons are made between birds and mammals, between fungi, or between the three multicellular kingdoms. To account for these observations, the rate of amino acid replacement must have been eight or more times higher in some lineages and at some times than in others. Spastic evolution of *Xdh* appears to be related to the particularities of the genomes in which the locus is embedded.

Key words: XDH protein evolution — Molecular clock — Heterogeneous nucleotide composition — Fluctuating mutation bias — Isochores — Lineage effects

Introduction

The *Xdh* locus and its encoded protein, xanthine dehydrogenase (XDH; EC 1.1.1.204), have long been one of the most studied gene–enzyme systems. XDH is a complex metalloflavoprotein that plays an important role in nucleic acid degradation in all organisms: it catalyzes the oxidation of hypoxanthine to xanthine and xanthine to uric acid, with concomitant reduction of NAD to NADH. XDH activity protects the cell against free oxygen radical-induced damage through the antioxidant action of uric acid (Xu et al. 1994). The chief physiological function of the enzyme changes, nonetheless, from one to another organism: its primary role is purine metabolism in mammals and chicken but pteridine metabolism in *Drosophila* (review by Chovnic et al. 1977). The *Xdh* locus is widely expressed in human tissues (Xu et al. 1994); in *Drosophila melanogaster* (*rosy* locus) and *Bombyx mori* it is transcribed in the fat body, midgut, and Malpighian tubules; and in *D. melanogaster* some part of the protein is transported to the eyes. *rosy* mutants that lack XDH activity have shorter life spans than wild-type flies, and they cannot synthesize the red drosopterin eye pigments, and consequently they have brownish eye color. *Bombix mori* mutants possessing no XDH activity exhibit translucent phenotypes that are sterile and shorter-lived than wild-type phenotypes (Kômoto et al. 1999). In higher plants XDH takes part in ureide biosynthesis through de novo synthesis of purines from glutamine (Sagi et al. 1998). In mammals, but not in chicken and *Drosophila*, the enzyme can be converted to the oxidase form xanthine oxidase (XO) (Hille and Nishino 1995). Defective XDH causes xanthinuria in humans; the

* This paper is dedicated to the memory of Thomas H. Jukes, molecular evolutionist par excellence and long-term collaborator on the *Journal of Molecular Evolution.*

Correspondence to: Francisco Rodríguez-Trelles c/o Francisco J. Ayala; *email:* ftrelles@iiag.cesga.es

enzyme is a target of drugs against gout and hyperuricemia and has been associated with blood pressure regulation (Enroth et al. 2000).

Unlike in mammals, the *Xdh* locus is transcribed at a low level in *Drosophila* (Riley 1989); enzyme levels in this species respond to the dose of *rosy* alleles (Grell 1962) but are quite insensitive to chromosomal position effects (Spradling and Rubin 1983). The *Xdh* locus is located in Muller element E in *Drosophila* (polytene chromosomal position 87D11 in *D. melanogaster*) and chromosomes 2 and 17 in humans (2p23-p22) and mouse, respectively (Ichida et al. 1993).

Despite functional differences across organisms, the primary structure of the protein is highly conserved. The full amino acid sequences of XDH enzymes from various organisms, including animals, plants, and fungi, consist of approximately 1330 amino acids. XDH has been characterized as a soluble protein. Efforts at resolving the three-dimensional structure of the eukaryotic enzyme are in progress (Enroth et al. 2000). The enzyme is a homodimer with a molecular mass of about 290 kDa, with each monomer acting independently in catalysis. The monomer is divided into three domains linked by short interdomains: one small N-terminal domain (~165 residues) containing two nonidentical iron sulfur centers, a flavin adenine dinucleotide (FAD) binding domain (~260), and a C-terminal molybdenum cofactor (MoCo) binding domain (~740). Unlike in eukaryotes, in *Rhodobacter capsulatus* the iron sulfur and FAD domains are located in one polypeptide, and the molybdopterin binding site is located on a separate protein (Leimkhüler et al. 1998).

The data available indicate that the *Xdh* locus is present in a single copy in most organisms in which it has been sequenced, except in *B. mori*, in which there are two copies of the gene arranged in tandem (Kômoto et al. 1999), and *Arabidopsis thaliana* (*Arabidopsis* Genome Initiative unpublished GenBank accession No. AL079347.1), where a duplication seems to have occurred very recently. The exon/intron structure of the *Xdh* gene varies widely across taxa, ranging from 36 exons (e.g., mouse and human) to 3 exons (e.g., *Neurospora crassa*). In dipterans, *Xdh* provides three examples of recently inserted introns, allegedly originated by duplication of a preexisting intron (Tarrío et al. 1998); two of these introns, physically proximal, evolve at disparate rates and in concerted fashion with their surrounding coding regions, which has been related to the local environment of the chromatin (Rodríguez-Trelles et al. 2000b).

XDH has served as a benchmark for population genetic studies in *Drosophila* (Lewontin 1985; Keith et al. 1987). Allozyme electrophoresis surveys disclosed it as one of the most highly polymorphic enzymes in *Drosophila* species (e.g., Singh et al. 1982; Buchanan and Johnson 1983; Keith 1983). Comparison of the levels of polymorphism and divergence at the nucleotide level (e.g., Riley et al. 1989; 1992) indicate that the *Xdh* locus is evolving in a neutral fashion in the populations and species where it has been investigated (Riley et al. 1992).

Here we present an analysis of the *Xdh* gene region comprising more than half (~52%) of the coding sequence (~2085 bp), corresponding to 24 codons of the FAD domain, 45–54 codons of the interdomain, and most of the MoCo domain (~95%; 613–618 codons). The study is part of a larger effort to characterize patterns of gene sequence variation across large evolutionary time scales.

Materials and Methods

Taxa and Sequences. The 34 species studied and GenBank accession numbers for the *Xdh* nucleotide sequences are given in Fig. 1. We list *Dorsilopha*, *Hirtodrosophila*, and *Zaprionus* as *Drosophila* subgenera, following Tatarenkov et al. (1999), but *Scaptodrosophila* as a genus, following Grimaldi (1990), Kwiatowski et al. (1994), and Tatarenkov et al. (1999) (see also Remsen and DeSalle, 1998). The *Xdh* sequences from *D. ananassae*, *D. mimica*, and *D. busckii* were newly obtained for this study. The strategies for amplification, cloning, and sequencing are described by Tarrío et al. (1998, 2000) and Rodríguez-Trelles et al. (1999a). The *Xdh* region is duplicated in *B. mori* and *A. thaliana* (Kômoto et al. 1999); in the case of *B. mori* we use the sequence of *Xdh1*, because it manifests greater resemblance to the *Xdh* region of the remaining insects than *Xdh2*; and in the case of *A. thaliana* the duplicates are highly similar to each other, so which one is actually used is inconsequential to our study.

Homology Inferences. Alignments of the amino acid residues comprised within the FAD (24 residues) and MoCo (613–618) binding domains were readily obtained from the Protein Domain Families Database (PFAM version 5.5; http://www.sanger.ac.uk/cgi-bin/Pfam/). Residues pertaining to the linker region connecting the two domains (45–54 residues) were aligned using ClustalX version 1.81 (Thompson et al. 1997) and the alignment was adjusted by eye with the aid of the GeneDoc 2.6.001 program (Nicholas and Nicholas 1997). Gaps and ambiguous positions were removed with Gblocks version 0.73b (Castresana 2000) using the default options. The actual alignment consisted of 599 residues.

Models of Sequence Evolution. We follow a model-based maximum-likelihood framework of statistical inference. First, we model the substitution process in the *Xdh* region using a tree topology that is approximately correct, then we use the model so identified to generate nucleotide and amino acid maximum-likelihood distances between pairs of sequences. As a tree topology for model fitting we use the hypothesis shown in Fig. 1. Relationships which are not well established (e.g., the branching order of animal phyla) are set as polytomies. Use of other reasonable topologies is not expected to affect parameter estimates (Yang 1994).

We consider amino acid-, codon-, and nucleotide-based models. The amino acid substitution models used in this study are all special forms of the model of Yang et al. (1998), which is based on the matrix of Jones et al. (1992) with amino acid frequencies set as free parameters (referred to as JTT-F). Maximum-likelihood synonymous (d_S) and nonsynonymous (d_N) distances between pairs of sequences are obtained using the codon substitution model of Goldman and Yang (1994), with a single distance between any pair of amino acids; codon frequencies at equilibrium are calculated from the average nucleotide frequencies at the three codon positions (Yang 2000). Unlike approximate methods, maximum-likelihood methods properly accommodate transition/transversion rate biases and codon-usage biases, factors that are very important in the estimation of d_S and d_N (Yang 1998). At the nucleotide

Kingdom	Phylum	Class	Order	Family	Genus	Subgenus	Group	Species	Genbank #
Animalia	*Arthropoda*	*Insecta*	*Diptera*	*Drosophilidae*	*Drosophila*	*Sophophora*	*melanogaster*	*melanogaster*	Y00307
								erecta	AF169400
								ananassae	AF345901
							obscura	*pseudoobscura*	M33977
								subobscura	AF058976-7
								bifasciata	AF169402-3
							saltans	*saltans*	AF58978
								sturtevanti	AF058983
								subsaltans	AF058980
							willistoni	*willistoni*	AF093206
								nebulosa	AF93213
								sucinea	AF093211
						Drosophila	*virilis*	*virilis*	AF093215
							Hawaiian	*mimica*	AF345902
							Hirtodrosophila	*pictiventris*	AF093214
							Dorsilopha	*busckii*	AF345903
							Zaprionus	*tuberculatus*	AF093216
					Chymomyza			*amoena*	AF093217
					Scaptodrosophila			*lebanonensis*	AF058984
				Tephritidae	*Ceratitis*			*capitata*	AF093218
				Calliphoridae	*Calliphora*			*vicina*	M30316-7
			Lepidoptera		*Bombix*			*mori*	D38159
	Chordata	*Mammalia*	*Artyodactila*		*Bos*			*taurus*	X98491
					Syncerus			*caffer*	AF104501
					Tragelaphus			*oryx*	AF104502
			Rodentia		*Mus*			*musculus*	NM_011723
					Rattus			*norvegicus*	NM_017154
			Primates		*Homo*			*sapiens*	D11456
			Carnivora		*Felis*			*catus*	AF286379
		Aves			*Gallus*			*gallus*	D13221
	Nematoda				*Caenorhabditis*			*elegans*	Z83318
Fungi	*Ascomycota*				*Aspergillus*			*nidulans*	X82827
					Neurospora			*crassa*	AL391572
Plantae					*Arabidopsis*			*thaliana*	AL079347

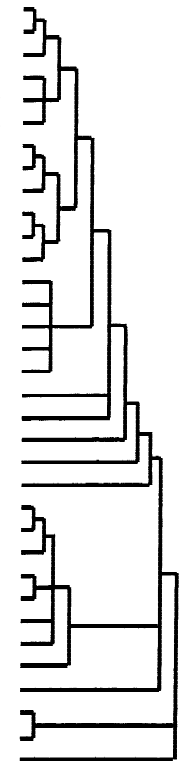

Fig. 1. Taxonomy of the species used in this study, GenBank accession numbers, and the topology used for model fitting.

level we use stationary and nonstationary substitution models. Stationary models assume that the nucleotide base composition is constant throughout the tree, whereas this assumption is relaxed in nonstationary models. For the stationary representation we use the general time-reversible [referred to as GTR (Yang 1994)] model and its nested cases; for the nonstationary models we use the maximum-likelihood implementation of the Tamura (1992) model of Galtier and Gouy (1998) and special cases of it. Variation of substitution rates across sites is accommodated into the substitution models using the discrete-gamma approximation of Yang (1996a) with eight equally probable categories of rates to approximate the continuous gamma distribution (referred to as dG models). The transition probability matrices of models and details about parameter estimation are given by Yang (2000) and Galtier and Gouy (1998).

Likelihood-ratio tests are applied to test several hypotheses of interest. For a given tree topology (i.e., Fig. 1), a model (H_1) containing p free parameters and with log-likelihood L_1 fits the data significantly better than a nested submodel (H_0) with $q = p-n$ restrictions and likelihood L_0 if the deviance $D = -2\log\Gamma = -2(\log L_1 - \log L_0)$ falls in the rejection region of a χ^2 distribution with n degrees of freedom (Yang 1996b). We use several starting values in the iterations to guard against the possible existence of multiple local optima. These analyses are conducted with the BASEML and CODEML programs from the PAML version 3.0b package (Yang 2000) and the EVAL_NH and EVAL_NHG programs from the NHML package (Galtier and Gouy 1998; Galtier et al. 1999).

Table 1. Percentage GC content in the first plus second and the third codon positions of the *Xdh* sequences analyzed in this study

	Codon position			Codon position	
Species	1st + 2nd	3rd	Species	1st + 2nd	3rd
D. melanogaster	50.2	65.2	*C. amoena*	48.7	44.3
D. erecta	50.6	66.6	*S. lebanonensis*	49.6	57.8
D. ananassae	49.7	62.2	*C. capitata*	46.7	42.7
D. pseudoobscura	51.8	80.5	*C. vicina*	45.4	38.1
D. subobscura	51.9	78.1	*B. mori*	46.1	31.9
D. bifasciata	51.8	72.4	*B. taurus*	47.2	64.2
D. saltans	49.2	42.8	*S. caffer*	47.5	63.4
D. sturtevanti	49.2	38.2	*T. oryx*	47.2	63.9
D. subsaltans	48.2	42.4	*M. musculus*	47.2	60.4
D. willistoni	49.5	47.4	*R. norvergicus*	47.9	60.3
D. nebulosa	49.4	50.7	*H. sapiens*	48.1	60.1
D. sucinea	49.7	51.1	*F. catus*	48.1	63.4
D. virilis	50.0	65.6	*G. gallus*	45.8	49.4
D. mimica	49.4	63.1	*C. elegans*	45.0	30.4
D. pictiventris	49.8	55.9	*E. nidulans*	50.5	57.6
D. busckii	50.4	61.6	*N. crassa*	49.4	69.1
Z. tuberculatus	49.7	59.3	*A. thaliana*	47.0	37.1

Results

Xdh *GC Content and Amino Acid Composition*

Table 1 summarizes the base composition of the *Xdh* sequences analyzed in this study (see also Rodríguez-Trelles et al. 1999b, 2000a, b). The GC content varies dramatically at third codon sites, where the highest GC percentage (80.5; corresponding to *D. pseudoobscura;* see Table 1) is ≈2.6 times greater than the lowest GC percentage (30.4; *C. elegans*). GC content variation in

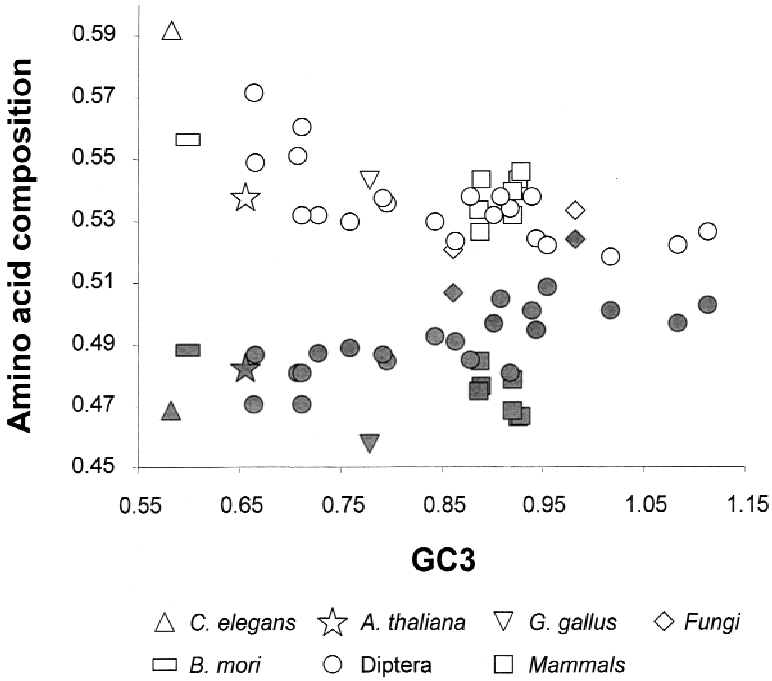

Fig. 2. Relationship between the frequencies of the high-GC (*filled symbols*) and low-GC (*open symbols*) amino acids and the frequency of GC content in third codon positions (GC3) for *Xdh*. Each data point in each group represents 1 of the 34 species studied. The GC frequencies are angularly transformed, so may reach values higher than 1.

the first and second codon positions is much narrower, ranging from 52.0% in *D. subobscura* to 45.0 in *C. elegans*, which can be attributed to stronger functional constraints on these two codon positions than on third codon positions. Note that although it is relatively young (≈65 My), the *Drosophila* genus accounts for most of the GC content variation in Table 1, which encompasses 1100 My; modern mammal orders have about the same age as the *Drosophila* genus; however, the range of GC content variation in the first plus second and in the third codon positions is ≈4 and ≈10 times larger, respectively, in *Drosophila* than in mammals (51.9–48.2 vs 48.1–47.2 and 80.5–38.2 vs 64.2–60.1; see Table 1).

Figure 2 represents variation across species in the amino acid composition of XDH as a function of GC content in the third codon position (GC3). Amino acids are classified according to the GC content in the first and second positions of their codons as high-GC (alanine, glycine, proline, and tryptophan; e.g., alanine is encoded by GCU, GCC, GCA, and GCG; represented by filled symbols in Fig. 2), low-GC (phenylalanine, isoleucine, lysine, methionine, asparagine, and tyrosine; e.g., phenylalanine is encoded by either UUU or UUC; open symbols in Fig. 2), and intermediate-GC (cysteine, aspartic acid, histidine, glutamine, serine, threonine, valine; e.g., aspartic acid is encoded by GAU or GAC; not shown). As expected, high-GC amino acids are less used by species with a low GC content (e.g., *C. elegans* and *B. mori*; triangles and rectangles in Fig. 2, respectively), while the opposite is the case for low-GC amino acids (e.g., the three species in the *Drosophila* obscura group; represented by circles corresponding to the highest three GC3% values in Fig. 2). According to Felsenstein's (1985) contrast test, which accounts for the phylogenetic inertia of the data, the correlation between the proportions of high-GC amino acids and the GC3 (variables previously transformed angularly; we assume the tree topology shown in Fig. 1) is highly significant ($r = 0.47$, $p = 0.003$; $n = 33$ contrasts), the correlation of the GC3 and the low-GC amino acids is $r = -0.36$ ($p = 0.01$), and the correlation between the GC3 and the intermediate-GC amino acids is not significant ($r = 0.13$, $p = 0.24$). Within taxonomic categories, dipterans (circles in Fig. 2) exhibit patterns of association similar to those of the complete data set, but strengthened ($r = 0.55$ and $p = 0.005$, $r = -0.46$ and $p = 0.02$, and $r = -0.14$ and $p = 0.28$ for the correlation of high-GC, low-GC, and intermediate-GC with GC3, respectively; $n = 21$ contrasts); in mammals (squares in Fig. 2) the pattern of associations is reversed, although the correlations are insignificant ($r = 0.22$ and $p = 0.70$, $r = 0.28$ and $p = 0.62$, and $r = 0.40$ and $p = 0.46$, for the correlation of high-GC, low-GC, and intermediate-GC with GC3, respectively; $n = 6$ contrasts).

Processes of Amino Acid and Nucleotide Substitution Along Xdh

Table 2 shows the log-likelihood-ratio statistic values for models of protein evolution assuming the tree topology shown in Fig. 1. Nested models are rejected compared against the next full model. The best description of the amino acid substitution process of XDH is provided by the JTT-F + dG model, which treats amino acid frequencies as free parameters and allows variable replacement

Table 2. Results of the likelihood-ratio test carried out on the *Xdh* amino acid data for the species analyzed in this study[a]

Assumption	H_0	H_1	$-2\log\Gamma$	df	p	α
Equal rate of change between any two amino acids	Poisson	JTT-F	2803.41	19	$<10^{-6}$	—
Constant rate among sites	JTT-F	JTT-F + dG	1135.52	1	$<10^{-6}$	0.839[b]
Constant rates among lineages	JTT-F + dG clock	JTT-F + dG	136.59	33	$<10^{-6}$	0.846[c]

[a] In each row, the null hypothesis (H_0) is compared with a hypothesis (H_1) that removes the assumption indicated on the left column. Log-likelihood values are obtained assuming the topology shown in Fig. 1. p represents the probability of obtaining the observed value of the likelihood-ratio test statistic ($-2\log\Gamma$) if the null hypothesis were true, with the degrees of freedom (df) indicated. Poisson, Poisson model; JTT-F, Yang et al. (1998) model; JTT-F + dG, assuming discrete gamma-distributed rates at sites; JTT-F +dG clock, assuming that all lineages evolve at equal rates.
[b] Obtained with the JTT-F + dG model.
[c] Obtained with the JTT-F +dG clock model.

rates among amino acid sites. The estimate of the parameter α of the discrete gamma distribution is not much lower than 1 ($\alpha = 0.839$; Table 2), which means that among-site rate variation along XDH is not too high.

Following the same methodological approach as that used above for the amino acid sequences, we found that the GTR + dG model is a representation of the nucleotide substitution process superior to any nested competitor for the complete sequences ($\ln L = -34957.27$, $\alpha = 0.531$), as well as when the third codon positions are excluded ($\ln L = -14455.03$, $\alpha = 0.633$). The GTR + dG model assumes that the nucleotide substitution pattern has remained constant over time (i.e., the homogeneity assumption) and among lineages (i.e., the stationarity assumption). These two premises are clearly untenable given the large nucleotide composition differences among sequences, at least in third codon positions (Table 1). Maximum-likelihood methods are known to be generally robust to model assumption violations. Nevertheless, to control potential estimation biases induced by nucleotide composition differences among sequences, we have used nonhomogeneous-nonstationary models. We found the T92 + dG + GC model better than any special form of it ($\ln L = -34497.00$, $\alpha = 0.500$, and $\ln L = -14460.58$, $\alpha = 0.636$, for the complete sequences and after excluding third codon positions, respectively); because they are nonnested models, the GTR + dG and the T92 + dG + GC models cannot be compared with the likelihood-ratio test.

Long-Term Evolution of XDH in Diptera

The substitution models and estimates of the among-site rate variation obtained above are now used to calculate amino acid and nucleotide distances between pairs of sequences. In addition, we use the codon substitution model of Goldman and Yang (1994; see Materials and Methods) to generate maximum-likelihood estimates of the pairwise nonsynonymous (K_a) synonymous (K_s) differences. Rows 1–5 in Table 3 give averages ($\overline{X}$) of the pairwise amino acid differences between dipterans, as well as the rate of amino acid evolution, expressed in units of 10^{-10} replacements per site per year. Figure 3 displays the number of amino acid replacements against time, and Fig. 4 shows K_a (Fig. 4A), K_s (Fig. 4B), and the number of nucleotide substitution differences obtained with the GTR + dG and the T92 + dG + GC models for the first and second codon positions (Figs. 4C and E, respectively) and the complete sequences (Figs. 4D and F, respectively). White, gray, and black circles represent averages for comparisons between the drosophilids, between *Ceratitis* and the drosophilids, and between *Calliphora* and the remaining dipterans, respectively. It is apparent that XDH evolves fairly regularly in *Drosophila*. The rates are 30.4×10^{-10} replacements per site per year between *Drosophila* groups, 29.2 between *Drosophila* subgenera, and 31.7 between *Drosophila* genera, similar enough to be acceptable as sample variation of the same stochastic clock. The average of these three rates is $\approx 30.4 \times 10^{-10}$/site/year, somewhat higher than the rate between dipteran families (25.3); the differences are graphically represented by linear regressions in Fig. 3. These differences are, however, fairly inconspicuous for the nonsynonymous substitutions (Fig. 4A), and they do not exist when we use the nucleotide distances generated by the T92 + dG + GC model on the first and second codon position data (Fig. 4E), which suggests that the heterogeneous base composition of the sequences is somewhat deflecting amino acid distance estimates.

Global Rates of XDH Evolution

Figure 1 displays the taxonomy of the 34 species investigated in this study; they encompass two orders of insects (dipterans and lepidopterans), four orders of mammals (rodents, primates, carnivores, and artiodactyls), two classes of vertebrates (mammals and birds), three metazoan phyla (arthropods, chordates, and nematodes), one phylum of fungi (ascomycetes), and three multicellular kingdoms (animal, plants, and fungi). Table 3, rows 6–11, gives the average number of amino acid differences (X) and the rate of amino acid evolution for different levels of evolutionary divergence. Figures 5A and B are plots against time of the number of amino acid differences and the number of nucleotide substitutions in

Table 3. Rate of *Xdh* evolution for increasingly divergent species[a]

Comparison	My	Amino acid replacements		Clock estimate (My)
		$\bar{X}$	Rate	
1. Within *Drosophila* groups	25–30	6.1–11.0	20.3–36.7	19–35
2. Between *Drosophila* groups	55 ± 10	16.7 ± 0.7	30.4	53
3. Between *Drosophila* subgenera	60 ± 10	17.5 ± 0.3	29.2	55
4. Between dipteran genera	65 ± 10	20.6 ± 1.7	31.7	65
5. Between dipteran families	120 ± 20	28.4 ± 1.7	25.3	90
6. Between *Arthropoda* orders	400 ± 100	72.4 ± 1.5	18.1	228
7. Between mammalian orders	70 ± 10	12.0 ± 1.2	17.1	38
8. Between mammals and birds	300 ± 50	31.5 ± 1.1	10.5	99
9. Between animal phyla	600 ± 100	115.3 ± 12.5	19.2	364
10. Between *Fungi*	300 ± 50	41.1	13.7	130
11. Between kingdoms	1100 ± 200	126.1 ± 8.5	11.5	398

[a] The species compared are listed in Fig. 1. The plus/minus values are crude estimates of error for My but are standard deviations for amino acid replacements. "My" values are modified from Ayala (1997). $\bar{X}$ values are per 100 residues for differences between species; the rates are lineage values and are expressed in units of 10^{-10} per site per year. The clock estimates of divergence time are extrapolated under the assumption that the *Drosophila* rate applies to other organisms.

the first and second codon positions inferred with the T92 + dG + GC model, respectively. The rate of amino acid replacement changes spastically from one level of evolutionary divergence to the next. A maximum-likelihood ratio test of the difference between the fit of a global clock model (i.e., XDH evolves at constant rate throughout the tree in Fig. 1) against that of a model relaxing this assumption indicates that the observed variation in rates is significant ($-2\log\Lambda = 136.59$, $p < 10^{-6}$, 33 df.; see row 3 in Table 2). When comparisons are made between kingdoms, which have evolved independently over ≈1100 My, the rate is 11.5×10^{-10}/site/year (Table 3, line 11). This rate is roughly similar to the average rate of amino acid replacement between ascomycetes (13.7) that diverged over the last 300 My (Berbee and Taylor 1993), but ≈1.7 times slower than the rate between animal phyla (19.2; 600 My), between arthropod orders (18.1; 400 My), and between mammals (17.1; 70 My), and approximately the same as the rate between birds and mammals (10.5; 300 My), but only one-third the average rate of the drosophilids (≈30.4; 60 My).

The differences become more apparent when we take into account that these rates apply to largely overlapping lineages. For example, the average number of amino acid differences between vertebrates (as inferred from the divergence between birds and mammals; see Table 3) and nematodes, which diverged some 600 My ago, is 21.4×10^{-10}/site/year. We infer, however, that the rate between birds and mammals is two times slower than this value (i.e., 10.5). In addition, the rate between vertebrates and arthropods is 15.1; because arthropods evolve at a slightly higher rate (18.1) it seems reasonable to assume that the average rate of evolution of XDH in vertebrates during the 600 My elapsed since their split from nematodes has been 10.5. This means that to attain an average of 21.4 between vertebrates and nematodes, XDH had to have evolved at a prevailing rate three times higher in nematodes than in vertebrates (32.3×10^{-10}/site/year); this is roughly the same as the average rate of XDH evolution in *Drosophila.* A similar reasoning can be applied to the case of mammals and birds: if we admit that the lineage of mammals has evolved at an average rate of 17.1 since they became separated from birds, to attain an average of 10.5 between mammals and birds, the latter

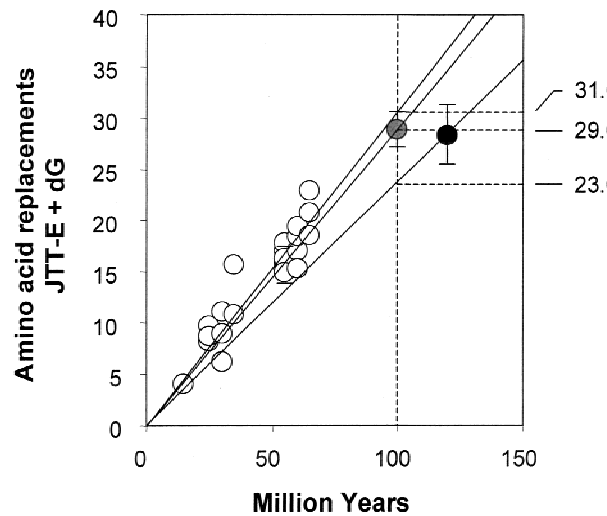

Fig. 3. Rate of XDH amino acid replacement in dipterans. Time (abscissa) is in million years. *White circles* indicate comparisons made between drosophilids; *gray circles*, between *Ceratitis* and the drosophilids; and *black circles*, between *Calliphora* and the remaining dipterans. Averages (with their standard errors) are calculated to minimize the impact of the phylogenetic structure of the sequence data shown in Fig. 1. Thus, for example, for the *Drosophila melanogaster* species group, the average amino acid distance (0.0811±0.0115) is the arithmetic mean of the pairwise distances *D. ananassae* to *D. melanogaster* (0.0925) and *D. ananassae* to *D. erecta* (0.0696). The rates on the right are replacements × 10^{-10} per site per year. The rates (31.0 between the drosophilids, 29.0 between *Ceratitis* and the drosophilids, and 23.0 between *Calliphora* and the remaining dipterans) are obtained by linear regression, with the intercept constrained to be the origin; the slopes of the regression lines are 0.0031, 0.0029, and 0.0023, respectively.

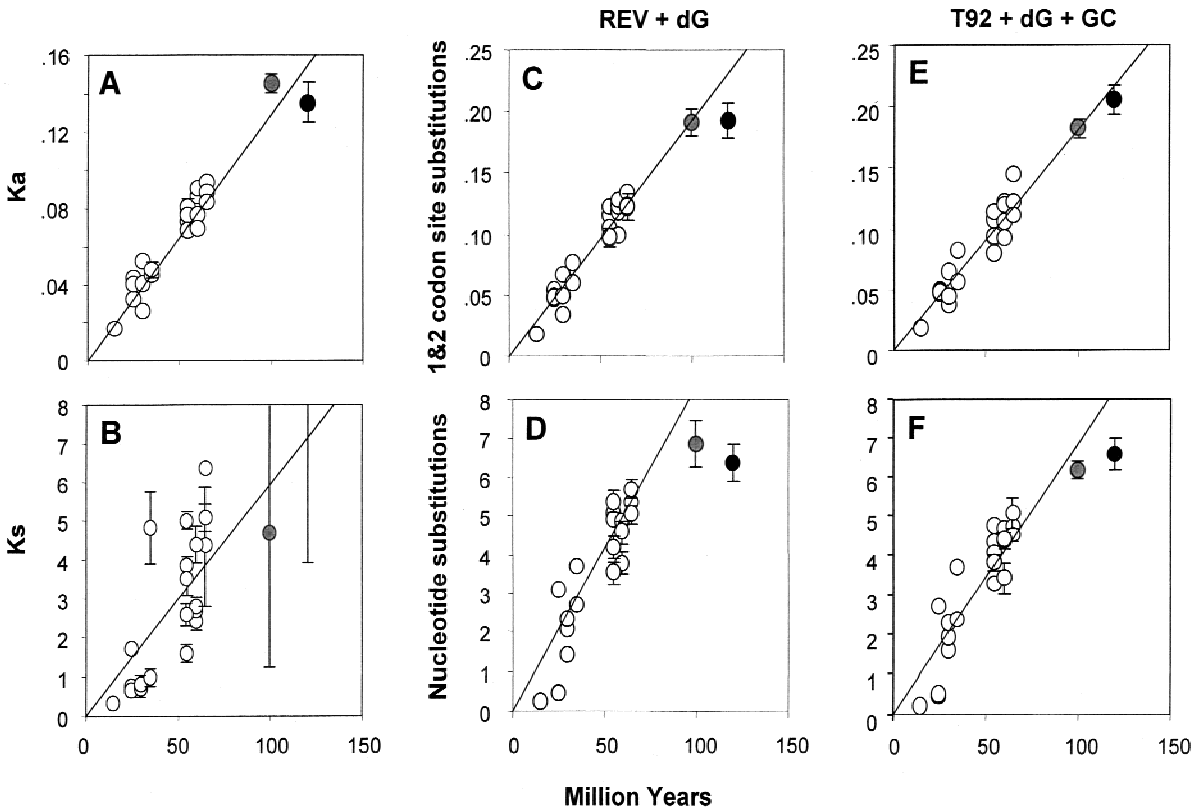

Fig. 4. *Xdh* rates of nonsynonymous **(A)** and synonymous **(B)** substitutions, and rates of nucleotide substitution in the first plus second codon positions, and in the complete sequences obtained with the REV + dG (**C** and **D,** respectively) and the T92 + dG + GC (**E** and **F,** respectively) models, in dipterans. *Straight lines* represent linear regressions of the rates between the drosophilids on time. The regressions are conducted as in Fig. 3, and the slopes are, from A to F, 0.0014, 0.0609, 0.0019, 0.0075, 0.0019, and 0.0072, respectively. Symbols and annotations are as in Fig. 3.

lineage must have evolved at a rate of only 3.9, i.e., a rate approximately eight times slower than that of *Drosophila*; similarly, lepidopterans would have evolved one-sixth more slowly than dipterans in order to yield a rate of 18.1 in arthropods. The fact that they hold after applying the T92 + dG + GC to the first and second nucleotide codon positions (see Fig. 5B) indicates that the observed rate differences are not an artifact caused by undue nucleotide composition variation between the sequences.

Discussion

Several factors can potentially distort evolutionary distance estimates derived from molecular data. Two of the most serious ones are variation in the evolutionary rate from site to site (Yang 1996a) and heterogeneous nucleotide base composition between sequences (Yang and Roberts 1995; Galtier and Gouy 1995; Tourasse and Li 1999). Ignoring among-site rate variation when it is present tends to produce underestimates of sequence distances, with the bias being more severe for larger distances than for small ones (Yang 1996a). Omission of nucleotide composition differences can lead to inflated distances because of changes in nucleotide composition rather than changes in substitution rate (Tourasse and Li 1999). We have identified these two factors as significant features of the data by adopting a model-based maximum-likelihood approach, then we have taken them into account in the calculation of distances using the discrete gamma distribution and nonhomogeneous nonstationary representations of the substitution process. That our results hold under a broad range of model assumptions and conditions increases the confidence that the observed variation in the long-term evolutionary rates of XDH is real.

Population genetics studies of *Drosophila* had suggested that XDH might be a good candidate for the study of long-term rates of molecular evolution. The *Xdh* locus, one of the largest analyzed in *Drosophila*, is highly variable. The variation conforms to the expectations of the neutral theory (Riley 1989; Riley et al. 1992), so that we would expect the protein to evolve in a clock-like fashion over evolutionary time (Kimura 1983). We have observed, on the contrary, that the rate of XDH evolution changes spastically from one to another level of evolutionary divergence. For instance, applying the *Drosophila* rate, the origin of the multicellular kingdoms would be timed at about 400 My ago (see Table 3), clearly a gross underestimate.

The rate appears at first to be fairly uniform in *Dro-*

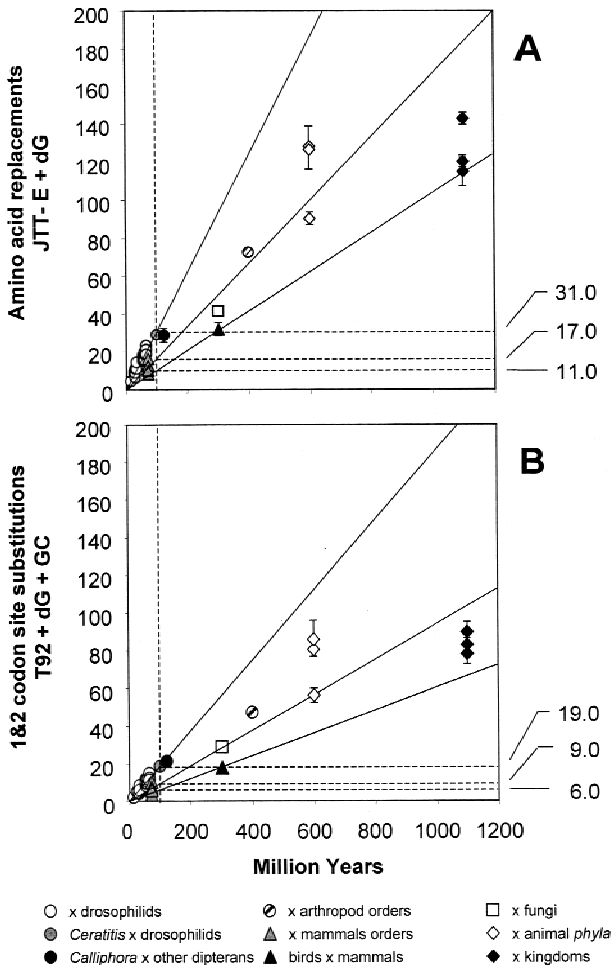

Fig. 5. Global rates of amino acid replacement for XDH **(A)** and nucleotide substitution rates in the first plus second codon positions as inferred with the T92 + dG + GC model (**B**). The rates on the right are replacements × 10^{-10} per site per year; the rates correspond to the comparisons between the drosophilids (31.0 and 19.0 for A and B, respectively), between mammal orders (17.0 and 9.0), and between birds and mammals (11.0 and 6.0). The regressions are conducted as in Fig. 3, and the slopes are 0.0031, 0.0017, and 0.0011 for A and 0.0019, 0.0009, and 0.0006 for B. The amino acid replacement rate (×10^{-10} per site per year) is 19.2 between animal phyla and 11.5 between the kingdoms; see Table 3 (regression lines are not shown).

sophila. This is, however, a false perception that emanates from the fact that we are dealing with long-term averages. Consider, for example, the case of the *Sophophora* subgenus of *Drosophila*. The fastest rate within this subgenus (36.7 × 10^{-10}/site/year; corresponding to the *saltans* species group) is about twice as high as the slowest (20.0 × 10^{-10}/site/year; *obscura* species group); this is approximately the same difference that exists between the evolutionary rate of XDH in *Drosophila* (30.4) and that in mammals (17.1). Implications of the observed rate variation within the subgenus *Sophophora* are discussed below.

Previous studies from our laboratory have focused on the long-term evolutionary rates of SOD and GPDH using a set of species largely overlapping with the ones of this study and have unveiled dramatic variations in evolutionary rate through time and among lineages (see Ayala 1997, 1999, and references therein). Importantly, the rates do not change following similar patterns in both loci. This observation allowed us to eliminate lineage effects, such as variable generation times or effective population sizes as explanations for the rate variations. Comparison of the previous and current results cannot be made with respect to the precise values obtained, however, given that we previously used a simple Poisson model of protein evolution (see Table 2) and did not take into account among-site rate variation or heterogeneous nucleotide base composition among sequences (Ayala et al. 1996; Kwiatowski et al. 1997).

Has the physiological function of XDH changed in evolution? This does not seem to be the case with respect to the role of the enzyme as a catalyzing agent in nucleic acid degradation; also, the domains and their arrangement in the quaternary structure of the protein appear to be well conserved throughout the eukaryote kingdom [in bacteria the iron sulfur and FAD domains and the molybdopterin binding site are located on separate polypeptides (Leimkhüler et al. 1998)]. XDH has, nonetheless, evolved extra functions, which vary from one to another organism. For example, as noted in the Introduction, in mammals the primary metabolic role of XDH is in purine catabolism, whereas in *Drosophila* it is in the metabolism of pteridines (Hille and Nishino 1995); the latter compounds are, however, virtually absent in the dipteran *Calliphora* (Houde et al. 1989). In higher plants, XDH plays a role in the biosynthesis of ureides (Sagi et al. 1998); in mammals, but seemingly not in birds and insects, XDH can be interconverted to the xanthine oxidase (XO) (Hille and Nishino 1995) form. The mechanical basis for these functional differences is at present insufficiently understood (Doyle et al. 1996; Glatigny et al. 1998; Enroth et al. 2000).

It seems unlikely that adaptation alone has been the primary factor responsible for the spastic long-term evolution of XDH. As mentioned above, studies of the intraspecific variation in the *Xdh* locus of *Drosophila* have shown that a large fraction of the protein sites appears to be evolving free of functional constraints, which has been related to the fact that XDH is large [i.e., may tolerate numerous amino acid replacements that are effectively neutral in their physiological effect (Riley et al. 1992)] and soluble [i.e., free of the constraints imposed, for example, by transmembrane structural requirements (Riley et al. 1989)]. Moreover, in almost all proteins where positive selection has been demonstrated, only a few amino acid sites have been responsible for the adaptive evolution (see Nielsen and Yang 1998). In the case of XDH, it seems unlikely that adaptive selection on a few sites would have a large effect on estimates of the global evolutionary rate of the protein.

It may be suggested that the proportion of constrained

residues in our XDH sequences is increased because of the strategy adopted for the alignment. In fact, in addition to 39 columns with alignment gaps, we removed from analysis 69 columns that were hypervariable, and thus putatively were hosting sites of dubious homology, according to the default criterion of the GenBlock program (Castresana 2000). However, after reanalysis of the data including these 69 sites, the patterns described in this study remain virtually unchanged (results not shown).

We now focus on the *Xdh* sequences of drosophilids and modern mammal orders. These two sets of lineages originated about 65–70 My ago; with three genera (*Drosophila, Chymomyza,* and *Scaptodrosophila*; see Fig. 1) and four orders (artiodactyls, carnivores, primates, and rodents), respectively, they are reasonably well represented in our data set.

The question arises, Why does XDH evolve at a rate two times faster in drosophilids than in mammals? Is the rate increased in drosophilids, decreased in mammals, or a combination of both? The answer(s) might be concealed in Fig. 2 (see also Table 1). The nucleotide and amino acid composition of *Xdh* varies between broad limits. However, it is apparent that drosophilids and mammals behave very differently in this respect: compositional variation among drosophilids is large compared to that of mammals. Either *Xdh* is "unusually" variable compositionally in drosophilids or the locus is compositionally greatly constrained in mammals. In our view, both alternatives may be involved.

The key to the puzzle is unlikely to be purifying selection for metabolic function in mammals, because, as is apparent in Fig. 2, a functional XDH molecule can accept conspicuously different amino acid compositions. Variable long-term population sizes in the drosophilids is also unlikely to be the answer, because effective population sizes are probably more variable in mammals; in any case, it is unclear why long-term effective population sizes should be so different among drosophilids, many of which share geographical distributions and ecological affinities. More likely, the answer lies in different dynamics of the genomes of drosophilids and mammals.

Patterns of compositional variation across drosophilids similar to those discussed here for *Xdh* have been reported for other nuclear regions, which suggests that they reflect genomewide processes (Rodríguez-Trelles et al. 1999b, 2000a, c). The base composition in the last common ancestor of *Sophophora* has recently been inferred for several genes (Rodríguez-Trelles et al. 2000c). Compared to extant representatives of the subgenus, the GC content has remained approximately constant in the lineage that evolved into the present *melanogaster* and *obscura* groups but decreased drastically during the evolution of the *saltans* and *willistoni* groups (Rodríguez-Trelles et al. 2000c). The shift in composition is associated with increased rates of substitution in synonymous and (less so) nonsynonymous sites toward low-GC content codons (Rodríguez-Trelles et al. 1999b, 2000a). The accelerated evolutionary rate in the *saltans–willistoni* stem is not an artifact emanating from the phylogenetic distribution of compositional differences accumulated at a constant rate (see Tourasse and Li 1999), because the acceleration remains significant after accounting for heterogeneous base composition with the nonhomogeneous-nonstationary LogDet distance relative rate test (Rodríguez-Trelles et al. 2000c). These observations are best explained by a shift in the pattern of point mutation that occurred in the ancestor of the *saltans–willistoni* offshoot, after it split from the lineage that gave rise to the *melanogaster* and *obscura* groups, possibly reinforced by synergistic effects of reduced population numbers (Rodríguez-Trelles et al. 1999b, 2000a, c). After inspection of Fig. 2, it is tempting to speculate that changeability of the pattern of point mutation might not be a distinctive characteristic of just a limited group of species within the subgenus *Sophophora,* but a general feature of the drosophilid genome. In addition to the switch just mentioned, the pattern of point mutation could have shifted, for example, when the drosophilids split from the tephritids, because *Scaptodrosophila* has an extremely high GC content compared to *Ceratitis* (see Table 1), or when the genus *Drosophila* split from *Chymomyza,* because the latter has a very low GC content compared to the prevailing situation in the former. Fluctuating mutation bias could, in this way, have speeded the evolution of the whole lineage.

Compared to drosophilids, the genome of mammals is highly structured compositionally in GC-rich and GC-poor regions [the so-called isochores (Bernardi et al. 1985)]. Fryxell and Zuckerkandl (2000) have recently shown that the evolution of the base composition of isochores is dominated by strong mutation bias resulting from the interplay between deamination of cytosine and the DNA repair machinery. The limited GC content evolution in the *Xdh* region in mammals could be a reflection of such mutational constraints. In this regard, XDH illustrates nicely how much a protein can vary (i.e., drosophilids) and how little it may actually vary (i.e., mammals) because of the constraints imposed by mutational biases.

We note that the explanatory hypothesis just proposed for XDH does not apply to the previous genes we have investigated. In the case of GPDH, the rate of evolution is higher (by a factor of five!) between mammal orders than between *Drosophila* subgenera: 5.3 versus 1.1×10^{-10} per site per year (Ayala et al. 1996; Ayala 1997, 1999). In the case of SOD the rate of evolution is virtually identical between mammal orders and between *Drosophila* subgenera: 16.2 versus 17.2×10^{-10} per site per year (Ayala et al. 1996; Ayala 1997, 1999).

Whether or not the explanatory hypothesis suggested above is correct for *Xdh,* it may well be the case that the

evolutionary dynamics of *Xdh* reflects better than other genes the imprinting of the genome in which is embedded. This property may render this gene particularly useful for investigating lineage effects on the limits of the variation of the rate of molecular evolution.

Whichever may be the underlying processes of the spastic pattern of XDH evolution, it is apparent that relying on XDH as an evolutionary clock would lead to erroneous, even grossly erroneous inferences. As shown in Table 3, the assumption that the XDH *Drosophila* rate prevails in other organisms would lead to estimations that the divergence of the animal phyla occurred just 354 My ago and that the multicellular kingdoms diverged as recently as 398 My, just a bit earlier than the divergence of the animal phyla. Our investigations of SOD and GPDH evolution over similar broad ranges of lineages and large spans of evolutionary time have also shown erratic patterns of evolution, which, moreover, are different from one gene to the other (Ayala 1997, 1999, and references therein) and from those of XDH. One may reasonably conclude that the hypothesis of the molecular clock is just a mirage (Ayala 1999). It is appropriate, then, to suggest that whenever molecular evidence is the only evidence available, or the most reliable, caution is called for. Moreover, one should use as many loci as reasonably possible. There is, of necessity, a change/time correlation—the longer the time elapsed, the larger the number of nucleotide substitutions or amino acid replacements that is likely to have occurred. As the number of loci increases, the "law of large numbers" will lead toward convergence on an average estimate that will likely approximate the correct value.

Acknowledgments. We are indebted to Nicolas Galtier and Nicolas Tourasse for making their software available. F.R.-T. has received support from the Spanish Council for Scientific Research (Contrato Temporal de Investigación) and Grant XUGA 400001B98 to A.M. Vieitez. The research was supported by NIH Grant GM42397 to F.J.A.

References

Ayala FJ (1997) Vagaries of the molecular clock. Proc Natl Acad Sci USA 94:7776–7783

Ayala FJ (1999) Molecular clock mirages. BioEssays 21:71–75

Ayala FJ, Barrio E, Kwiatowski J (1996) Molecular clock or erratic evolution? A tale of two genes. Proc Natl Acad Sci USA 93:11729–11734

Berbee ML, Taylor JW (1993) Dating the evolutionary radiations of the true fungi. Can J Bot 71:1114–1127

Bernardi G, Olofsson B, Filipski J, Zerial M, Salinas J, Cuny G, Meunier-Rotival M, Rodier F (1985). The mosaic genome of warm-blooded vertebrates. Science 228:953–958

Buchanan BA, Jonson DEL (1983) Hidden electrophoretic variation at the xanthine dehydrogenase locus in a natural population of *Drosophila melanogaster*. Genetics 104:301–315

Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17: 540–552

Chovnick A, Gelbart W, McCarron M (1977) Organization of the *rosy* locus in *Drosophila melanogaster.* Cell 11:1–10

Doyle WA, Burke JF, Chovnick A, Dutton FL, Whittle RS, Bray RC (1996) Properties of xanthine dehydrogenase variants from *rosy* mutant strains of *Drosophila melanogaster* and their relevance to enzyme's structure and mechanism. Eur J Biochem 239:782–795

Enroth C, Eger BT, Okamoto K, Nishino T, Nishino T, Pai EF (2000) Cystal structures of bovine milk xanthine dehydrogenase and xanthine oxidase: Structure-based mechanism of conversion. Proc Natl Acad Sci USA 97:10723–10728

Fryxell KJ, Zuckerkandl E (2000) Cytosine deamination plays a primary role in the evolution of mammalian isochores. Mol Biol Evol 17:1371–1383

Felsenstein J (1985) Phylogenies and the comparative method. Am Nat 125:1–15

Galtier N, Gouy M (1995) Inferring phylogenies from DNA sequences of unequal base compositions. Proc Natl Acad Sci USA 92:11317–11321

Galtier N, Gouy M (1998) Inferring the pattern and process: Maximum-likelihood implementation of a nonhomogeneous model of DNA sequence evolution for phylogenetic analysis. Mol Biol Evol 15: 871–879

Galtier N, Tourasse N, Gouy M (1999) A nonhyperthermophilic common ancestor to extant life forms. Science 283:220–221

Glatigny A, Hof P, Romão MJ, Huber R, Scazzocchio C (1998) Altered specificity mutations define residues essential for substrate positioning in xanthine dehydrogenase. J Mol Biol 278:431–438

Goldman N, Yang Y (1994) A codon-based model of nucleotide substitutions for protein-coding DNA sequences. Mol Biol Evol 11: 725–736

Grell EH (1962) The dosage effect of ma-l^+ and ry^+ on xanthine dehydrogenase activity in *Drosophila melanogaster*. Z VererbLehre 93:371–377

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

Hille R, Nishino T (1995) Xanthine oxidase and xanthine dehydrogenase. FASEB J 9:995–1003

Houde M, Tiveron M-C, Brégégère F (1989) Divergence of the nucleotide sequences encoding xanthine dehydrogenase in *Calliphora vicina* and *Drosophila melanogaster.* Gene 85:391–402

Ichida K, Amaya Y, Noda K, Minoshima S, Hosoya T, Sakai O, Shimizu N, Nishino T (1993) Cloning of the cDNA encoding human xanthine dehydrogenase (oxidase): Structural analysis of the protein and chromosomal location of the gene. Gene 133:279–284

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. CABIOS 8:275–282

Keith TP, Riley MA, Kreitman M, Lewontin RC, Curtis D, Chambers G (1987) Sequence of the structural gene for xanthine dehydrogenase (*rosy* locus) in *Drosophila melanogaster.* Genetics 116:67–73

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge

Kômoto N, Yukuhiro K, Tamura T (1999) Structure and expression of tandemly duplicated xanthine dehydrogenase genes of the silkworm (*Bombyx mori*). Insect Mol Biol 8:73–83

Kwiatowski J, Skarecky D, Bailey K, Ayala FJ (1994) Phylogeny of *Drosophila* and related genera inferred from the nucleotide sequence of the Cu,Zn *Sod* gene. J Mol Evol 38:443–454

Kwiatowski J, Krawczyk M, Jaworski M, Skarecky D, Ayala FJ (1997) Erratic evolution of glycerol-3-phosphate dehydrogenase in *Drosophila*, *Chymomyza*, and *Ceratitis.* J Mol Evol 44:9–22

Leimkühler S, Kern M, Solomon PS, McEwan AG, Schwarz G, Mendel RR, Klipp W (1998) Xanthine dehydrogenase from the phototrophic purple bacterium Rhodobacter caopsulatus is more similar to its eukaryotic counterparts than to prokaryotic molibdenum enzymes. Mol Microbiol 27:853–869

Lewontin RC (1985) Population genetics. Annu Rev Genet 19:81–102

Nicholas KB, Nicholas HB Jr (1997) GeneDoc: A tool for editing and

annotating multiple sequence alignments vs. 2.6.2001. Distributed by the authors

Nielsen R, Yang ZH (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148:929–936

Remsen J, DeSalle R (1998) Character congruence of multiple data partitions and the origin of the Hawaiian *Drosophilidae*. Mol Phylogenet Evol 9:225–235

Rodríguez-Trelles F, Tarrío R, Ayala FJ (1999a) Molecular evolution and phylogeny of the *Drosophila saltans* species group inferred from the *Xdh* gene. Mol Phylogenet Evol 13:110–121

Rodríguez-Trelles F, Tarrío R, Ayala FJ (1999b) Switch in codon bias and increased rates of amino acid substitution in the *Drosophila saltans* species group. Genetics 153:339–350

Rodríguez-Trelles F, Tarrío R, Ayala FJ (2000a) Fluctuating mutation bias and the evolution of the base composition in *Drosophila.* J Mol Evol 50:1–10

Rodríguez-Trelles F, Tarrío R, Ayala FJ (2000b) Disparate evolution of paralogous introns in the *Xdh* gene of *Drosophila.* J Mol Evol 50:123–130

Rodríguez-Trelles F, Tarrío R, Ayala FJ (2000c) Evidence for a high ancestral GC content in *Drosophila.* Mol Biol Evol 17:1710–1717

Riley MA (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

Riley MA, Hallas ME, Lewontin RC (1989) Distinguishing the forces controlling genetic variation at the Xdh locus in *Drosophila pseudoobscura.* Genetics 123:359–369

Riley MA, Kaplan SR, Veuille M (1992) Nucleotide polymorphism at the xanthine dehydrogenase locus in *Drosophila pseudoobscura.* Mol Biol Evol 9:56–69

Sagi M, Omarov RT, Lips SH (1998) The Mo-hydroxylases xanthine dehydrogenase and aldehyde oxidase in ryegrass as affected by nitrogen and salinity. Plant Sci 135:125–135

Singh RS, Hickey DA, David J (1982) Genetic differentiation between geographically distant populations of *Drosophila melanogaster.* Genetics 104:301–315

Tamura K (1992) Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C content biases. Mol Biol Evol 9:678–687

Tarrío R, Rodríguez-Trelles F, Ayala FJ (1998) New *Drosophila* introns originate by duplication. Proc Natl Acad Sci USA 95:1658–1652

Tarrío R, Rodríguez-Trelles F, Ayala FJ (2000) Tree rooting with outgroups when they differ in their nucleotide composition from the ingroup: The *Drosophila saltans* and *willistoni* groups, a case study. Mol Phylogenet Evol 16:344–349

Tatarenkov A, Kwiatowski J, Skarecky D, Barrio E, Ayala FJ (1999) On the evolution of *Dopa decarboxylase (Ddc)* and *Drosophila* systematics. J Mol Evol 48:445–462

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 24:4876–4882

Tourasse NJ, Li WH (1999) Performance of the relative-rate test under nonstationary models of nucleotide substitution. Mol Biol Evol 16:1068–1078

Xu P, Hueckseadt TP, Harrison R, Hoidal JR (1994) Molecular cloning, tissue expression of human xanthine dehydrogenase. Biochem Biophys Res Commun 199:998–1004

Yang Z (1994) Estimating the pattern of nucleotide substitution. J Mol Evol 39:105–111

Yang Z (1996a) The among-site rate variation and its impact on phylogenetic analyses. TREE 11:367–372

Yang Z (1996b) Maximum likelihood models for combined analyses of multiple sequence data. J Mol Evol 42:587–596

Yang Z (1998) Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol Biol Evol 15: 568–573

Yang Z (2000) Phylogenetic analysis by maximum likelihood (PAML), vs. 3.0a University College London, London

Yang Z, Roberts D (1995) On the use of nucleic acid sequences to infer branchings in the tree of life. Mol Biol Evol 12:451–458

Yang Z, Nielsen R, Hasegawa M (1998) Models of amino acid substitution and applications to mitochondrial DNA evolution. Mol Biol Evol 15:1600–1611