Molecular Phylogenetics and Evolution
Vol. 13, No. 1, October, pp. 110–121, 1999
Article ID mpev.1999.0631, available online at http://www.idealibrary.com on IDEAL®

Molecular Evolution and Phylogeny of the *Drosophila saltans* Species Group Inferred from the *Xdh* Gene

Francisco Rodríguez-Trelles, Rosa Tarrío, and Francisco J. Ayala

Department of Ecology and Evolutionary Biology, University of California, Irvine, California 92697-2525

Received August 11, 1998; revised December 11, 1998

The *Drosophila saltans* group of the subgenus *Sophophora* consists of five species subgroups whose phylogenetic relationships are poorly known. We have analyzed 2085 coding nucleotides from the *xanthine dehydrogenase* (*Xdh*) gene in six species, at least one from each subgroup. We follow a model-based maximum likelihood framework. We first model the substitution process using a tree topology that is approximately accurate. Then we evaluate several candidate tree topologies using a working model of nucleotide substitution. We found that a minimally realistic description of the substitution process along the *Xdh* region should allow two transition and four transversion rate parameters and different fixed rates for codon positions, which are distributed statistically according to different gamma distributions. The phylogeny obtained using this description differs in significant respects from a phylogeny based on anatomical criteria. We have also analyzed data from five additional (three nuclear and two mitochondrial) gene regions. In our analysis, these relatively short DNA sequences, either separately or jointly, fail to discriminate statistically among alternative phylogenies. When the data for these five gene regions are combined with the *Xdh* sequences, the strong phylogenetic signal emerging from *Xdh* becomes somewhat diluted rather than reinforced. The phylogeny of the species and biogeographical considerations suggest that the *D. saltans* group originated in the tropics of the New World, similarly as the closely related *D. willistoni* group.

***Key Words: Drosophila saltans* group; *Xdh;* molecular evolution; phylogeny; maximum likelihood; nucleotide substitution rate.**

INTRODUCTION

The *Sophophora* radiation of the genus *Drosophila* produced four major lineages, recognized as species groups: the *melanogaster* group in the Old World tropics, the *obscura* group in the Holarctic, and the *willistoni* and *saltans* groups in the New World. The two Old World groups, *melanogaster* and *obscura,* are more closely related to each other than the two New World groups, *willistoni* and *saltans* (Sturtevant, 1942; Patterson and Stone, 1952; Throckmorton, 1975). The *saltans* species group consists of 21 species (Spassky, 1957; Magalhães and Bjornberg, 1957; Magalhães, 1962; Mourão and Bicudo, 1967; Throckmorton, 1975) which, according to morphological criteria, have been classified into five subgroups: the *cordata* and *parasaltans* subgroups, found only in the Neotropical region, and the *elliptica, sturtevanti,* and *saltans* subgroups, which also occur in the Nearctic region (Magalhães and Bjornberg, 1957; Magalhães, 1962). Studies of the *saltans* species group carried out in the past have, for the most part, focused on representatives of the *saltans* subgroup, including analyses of chromosomal polymorphisms (Targa, 1973; Bicudo, 1973a) and reproductive isolation between species (Bicudo, 1973b).

Compared to the other groups of the subgenus *Sophophora* (reviewed in Powell, 1997), the *saltans* group has only recently received attention from molecular evolutionists: the distribution of *P*-element sequences across the group (Daniels and Strausbaugh, 1986; Clark *et al.,* 1995) and the discovery of a new intron allegedly acquired by duplication of a preexisting intron in the xanthine dehydrogenase (*Xdh*) gene (Tarrío *et al.,* 1998). *Xdh* and other nuclear gene regions reveal that the *saltans* lineage, together with its sister *willistoni* clade, exhibits mutation patterns different from those observed in the other *Sophophora* lineages (Rodríguez-Trelles *et al.,* unpublished data). So far, however, the phylogeny of this group of species has received limited attention, based on a few anatomical characters (Throckmorton and Magalhães, 1962; Throckmorton, 1975), and some misleading molecular data (O'Grady *et al.,* 1998; see Materials and Methods).

In the present study we seek to determine the phylogenetic relationships among the five major *saltans* subgroups using molecular data. We have sequenced in six species part of the *Xdh* coding region, which leads to a strongly supported tree, and also analyze published data from the *28S rRNA* region (Pélandakis and Solignac, 1993) and four other gene loci, two nuclear and two mitochondrial (O'Grady *et al.,*

1055-7903/99 $30.00

1998). We follow a statistical model-fitting approach within the maximum likelihood framework of phylogenetic inference (e.g., Ritland and Clegg, 1985; Yang, 1995; Kumar, 1996). Thus, we first model the molecular evolution of the regions relevant to phylogenetic inference employing the likelihood ratio test and then we use the most satisfactory description so attained in order to reconstruct the evolutionary relationships in the *saltans* species group. Our results challenge the current view of phylogeny in the *saltans* group.

MATERIALS AND METHODS

Drosophila Strains

The six species of the *D. saltans* group and their subgroups are as follows (all strains are from the National *Drosophila* Services Resource Center in Bowling Green State University, Ohio 43403; the stock reference numbers are in parentheses): *cordata* subgroup: *D. neocordata,* from Minas Gerais, Brazil (14041-0831); *elliptica* subgroup: *D. emarginata,* from Boquete, Panamá (14042-0841.3); *parasaltans* subgroup: *D. subsaltans,* from Belem, Brazil (14044-08720); *saltans* subgroup: *D. prosaltans,* from Turrialba, Costa Rica (14045-0901.0) and *D. saltans,* from San José, Costa Rica (14045-0911.0); and *sturtevanti* subgroup: *D. sturtevanti,* from Turrialba, Costa Rica (14043-0871.0). We have used representatives of other *Sophophora* species as outgroups, namely, the *Xdh* sequences from *D. melanogaster* (Keith *et al.,* 1987; GenBank Accession no. Y00307), *D. pseudoobscura* (Riley, 1989; M33977), and *D. subosbcura* (from Helsinki, Finland), obtained by us.

Molecular Methods

The *Xdh* gene (*rosy* in *D. melanogaster*) is one of the longest genes sequenced in *Drosophila.* The region that we have examined corresponds to positions 940–3306 in the *D. melanogaster* sequence (Keith *et al.,* 1987), including about half of exon II (1113 bp), intron II (length ranging from 55 to 537 bp), and most of exon III (972 bp). The region represents about 52% (2085 bp) of the *Xdh* gene coding region. In the *saltans* group species, the sequenced region includes 54–66 bp corresponding to a newly discovered intron located within exon II, between codons 77 and 78 (Tarrío *et al.,* 1998). However, intron divergence between the species is much too large for unambiguous alignment; therefore we consider only the coding regions in the analyses. Alignments were obtained using the default option of the program CLUSTAL W (version 1.5) (Thompsom *et al.,* 1994).

Genomic DNA was prepared from 15 to 20 flies according to the method of Kawasaki (1990). The *Xdh* region was amplified by PCR with AmpliTaq DNA polymerase (Perkin–Elmer), using high-fidelity conditions (Kwiatoski *et al.,* 1991). PCR products were purified with the Wizard PCR Preps DNA Purification kit (Promega). The amplified *Xdh* region was ligated into the pCRII vector from the TA-Cloning kit (Invitrogen) and cloned into *Escherichia coli* INV $\alpha F'$ competent cells, according to the manufacturer's protocol. Plasmid DNA was prepared for sequencing using the QIAprep kit (QIAGEN). For each species, one clone was sequenced by the Sanger's dideoxynucleotide chain-termination method for denatured double-stranded plasmid DNA, by using the Sequenase v.2.0 DNA sequencing kit (United States Biochemicals–Amersham). Compressions and ambiguities were resolved by multiple sequencing of both strands. When a nucleotide substitution appeared in only one species, additional clones from different PCRs were sequenced, in order to avoid potential mistakes introduced by the PCR (Kwiatowski *et al.,* 1991). Details on the PCR and sequencing conditions and primers are given in Tarrío *et al.* (1998).

In addition to *Xdh* we have also considered published sequences from the *28S rRNA* region (541 bp; divergent domains D1 and D2) (Pélandakis and Solignac, 1993); as well as exon 2 of *Adh* (390 bp), *internal transcribed spacer-1* (*ITS1*) (163 aligned bp), and the mitochondrial *cytochrome oxidase I* (*COI*) (305 bp) and *cytochrome oxydase II* (*COII*) (688 bp). Sequences for these four genes are from O'Grady *et al.* (1998). The original publication using the *ITS1* locus refers to 785 aligned positions (see Table 2 in O'Grady *et al.,* 1998). However *ITS1* lengths are highly variable (ranging from 328 to 820 bp); we only consider the 163 bp that we could confidently align by means of CLUSTAL W.

Statistical Analyses

Maximum likelihood methods used in molecular phylogenetics assume a tree topology and a model of sequence change. In order to control possible errors due to imperfect prior knowledge of the phylogeny, we consider two phylogenetic topologies (Figs. 1 and 2). Figure 1 is based on the *Xdh* sequence data. This topology is stable after applying the computer programs DNAML and DNAPARS from the PHYLIP package (Felsenstein, 1993), with the default options used. Figure 2 represents the relationships proposed by Magalhães and Throckmorton (1962).

Substitution models considered in this study are all special forms of the general reversible Markov process model (Tavaré, 1986), referred to as REV (Yang, 1994). The REV model is sufficiently general for accurate estimation of the substitution pattern from the actual data (Yang *et al.,* 1995). Less complex models might however be desirable because they yield estimates with smaller variances (e.g., Kumar, 1996).

Rate variation among sites is accommodated into the substitution models following the approach of Yang (1994). Accordingly, the overall among-site rate variation is assumed to be made up of the contributions of two components: (i) a codon position-specific rate compo-

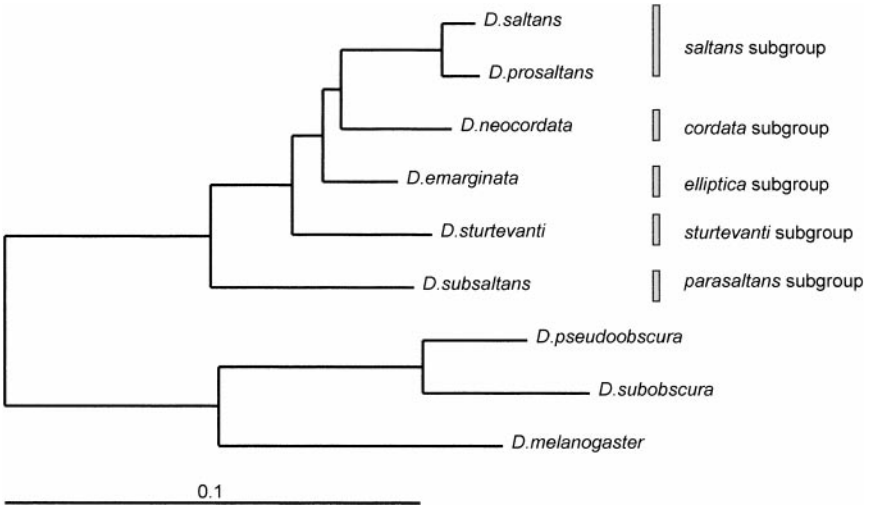

FIG. 1. Unrooted maximum likelihood tree of the *Drosophila saltans* group based on *Xdh* nucleotide sequences. The tree is obtained using the general reversible Markov process model (Yang, 1994), which allows different nucleotide frequencies and rates of substitution between nucleotides, with different fixed rates at codon positions that are randomly distributed according to different gamma distributions (REV + CdG$\pi\kappa$ model). The tree was obtained with the PAML 1.3 program (Yang, 1997). Branch lengths are proportional to the scale given in substitutions per nucleotide.

nent, which is considered as a parameter (denoted as c_1, c_2, c_3, for codon positions 1, 2, and 3, of the *Xdh* gene); c_1 is set equal to 1, so that c_2, and c_3 become rate ratios relative to the rate of the first codon position; (ii) the contribution made by the rate differences among-sites within the same codon position, which is treated as a random effect (Yang, 1996b). The random component is set by using the discrete gamma distribution (setting eight equal-probability categories of rates) with shape parameter α. The value of α is inversely related to the extent of rate variation among sites (Yang, 1996a). As nucleotide frequencies at equilibrium we use the observed averages which, for a few models examined, are very similar to the maximum likelihood estimates (results not shown). These analyses are conducted with the BASEML program from the PAML 1.3 package (Yang, 1997).

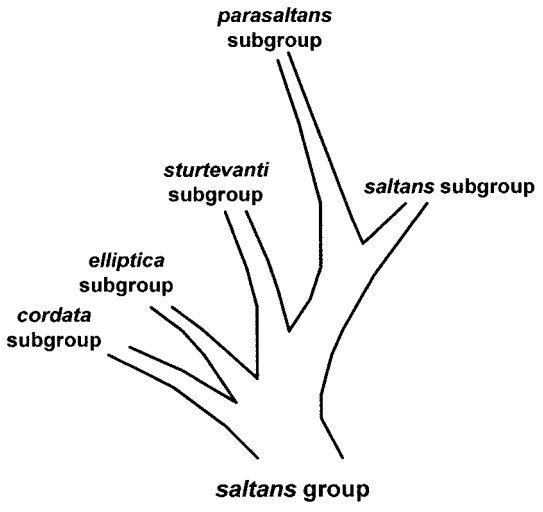

FIG. 2. Diagrammatic representation of the phylogenetic relationships among the five subgroups of the *saltans* group as proposed by Throckmorton and Magalhães (1962).

The relevance of specific parameters for describing the evolution of the *Xdh* sequences is evaluated by means of an analysis of deviance (Yang, 1996b; see also Huelsenbeck and Crandall, 1997). For a given tree topology (e.g., Fig. 1), a model (H_1) containing p 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 $D = -2 \log \Lambda = -2(\log L_1 - \log L_0)$ falls in the rejection region of a χ^2 distribution with n degrees of freedom. Specifically, for the test of rate constancy among sites, where the H_0 ($\alpha = \infty$) is equivalent to fixing α at the boundary of the parameter space of the H_1 ($\alpha < \infty$), H_0 tends to be rejected more often than expected from the nominal significance level (see Yang *et al.,* 1995). Yet, the likelihood differences of this study are all very large, so that this inaccuracy of the χ^2 approximation is not expected to alter the conclusions of the tests.

The model found to describe satisfactorily the substitution process in the *Xdh* region is used as a working hypothesis to generate candidate tree topologies by maximum likelihood. In addition, we have used distance-based neighbor-joining and weighted parsimony criteria to choose candidate trees. The estimates of α, transition/transversion bias, and codon-specific substitution rates (c's) that we use in distance computation and weighting schemes for maximum parsimony are those obtained simultaneously by the joint likelihood

comparison of all sequences in the first stage, which can be considered the most reliable (Yang, 1996a). Neighbor-joining (NJ) trees are generated using Kimura's (1980) two-parameter and Tamura and Nei's (1993) models to correct for unobserved changes. Parsimony analyses are conducted with *PAUP** (version 4d64ppc), using the branch-and-bound algorithm with the furthest addition option. Statistical support for nodes of the NJ and maximum parsimony trees is assessed using 50% majority-rule consensus trees compiled from 1000 bootstrap replications (Felsenstein, 1985). In the case of maximum parsimony, bootstrap replicates are obtained with the heuristic method (stepwise random addition with 10 replicates and tree-bisection-reconnection for branch swapping).

Phylogenetic hypotheses derived from the analyses of each gene region are compared by means of the Kishino and Hasegawa's (1989) test. For a given model of evolution, this test provides an estimate of the significance of a difference between the log likelihood scores obtained from two hypothetical trees.

RESULTS

Xdh Nucleotide Composition

Most models currently used in phylogenetic analysis rely on the assumption of stationarity of nucleotide frequencies (Swofford *et al.,* 1996). Therefore, when the base composition of taxa varies among sequences, sequences of similar nucleotide composition tend to become clustered irrespective of the evolutionary history of the organisms (Lockhart *et al.,* 1994). We have carried out tests for the homogeneity of base composition with the method of Rzhetsky and Nei (1995). This test takes into account covariances among classes that can arise because of phylogenetic correlation or correlation among nucleotide bases and so it is more appropriate than alternative approaches, such as χ^2 (Rzhetsky and Nei, 1995). Table 1 shows the nucleotide frequencies and their averages in the *Xdh* region. Compositional heterogeneity among species is due, for the most part, to differences between the *saltans* group (average G + C = 46.7%) and the outgroup species (G + C = 59.2%) ($I = 377.7$; $P < 10^{-4}$; 24 *d.f.*). Within the *saltans* group, nucleotide composition can be assumed to be stationary ($I = 13.9$; $P \approx 0.53$; 15 *d.f.*). Hence, we do not expect the phylogenetic signal to be seriously influenced by shared species-specific compositional biases within the *saltans* species group.

Nucleotide frequencies, averaged over all the three codon positions and species, are close to 0.25 proportions (in percentage, A = 24.1, T = 25.0, C = 23.8, G = 27.1). The *saltans* species deviate less from equal nucleotide frequencies (T = 25.8, C = 27.5, A = 21.3, G = 25.4) than the outgroups (T = 20.8, C = 20.0, A = 28.7, G = 30.5). Between codon positions, however, base frequencies are clearly unequal. Averaged across all species, first positions show G and A as the most frequent nucleotides (G = 36.4 and A = 24.8), while G and C are the least frequent in the second positions (G = 19.9 and C = 22.1); in the third positions, T has the highest frequency in the *saltans* species (T = 36.4), while C is the most frequent in the outgroups (C = 39.8).

The Process of Nucleotide Substitution along Xdh

We investigate the process of substitution along the *Xdh* region by adjusting a hierarchy of parametric models to the sequence data. At each step, the fit of a simple model (H_0) is compared against a more general model (H_1). Rejection of the simpler model implies that the relaxed restriction(s) in the alternative full model contribute(s) significantly to a better description of the

TABLE 1

Nucleotide Composition of the *Xdh* Region

	First position				Second position				Third position			
Species (subgroup)	T	C	A	G	T	C	A	G	T	C	A	G
saltans Group												
D. saltans (*saltans*)	17.7	22.0	24.7	35.5	28.2	21.7	30.2	19.9	35.3	21.6	21.3	21.9
D. prosaltans (*saltans*)	17.6	21.4	25.0	36.0	28.3	21.7	29.9	20.0	35.3	21.4	22.2	21.2
D. neocordata (*cordata*)	17.8	21.4	25.2	35.5	28.8	21.7	29.8	19.7	36.7	21.6	21.3	20.4
D. emarginata (*elliptica*)	17.7	21.3	24.7	36.3	28.2	21.7	29.8	20.3	36.4	20.6	22.7	20.3
D. sturtevanti (*sturtevanti*)	17.4	21.7	25.0	35.8	28.5	21.7	29.9	19.9	38.7	17.8	24.2	19.3
D. subsaltans (*parasaltans*)	17.6	21.0	26.2	35.3	28.6	21.9	29.9	19.6	36.0	21.0	21.9	21.2
melanogaster Group												
D. melanogaster	16.4	22.0	24.3	37.3	27.8	22.7	30.2	19.3	22.0	33.5	13.5	30.9
obscura Group												
D. pseudoobscura	13.7	24.2	23.9	38.3	27.8	22.6	29.9	19.7	13.4	43.3	7.3	36.0
D. subobscura	14.2	24.5	23.9	37.4	27.5	22.9	29.2	20.4	17.4	42.4	5.3	34.8
Average	16.7	22.2	24.8	36.4	28.2	22.1	29.9	19.9	30.1	27.0	17.7	25.1

Note. Numbers are percentage frequencies of each nucleotide.

data. Table 2 shows the values of the deviances (log-likelihood ratio statistics) for models. Log-likelihood values are obtained assuming the topology in Fig. 1. As shown in the table, nested models are always rejected compared against the next full, less constrained model. The best description so far of the substitution process along the *Xdh* region is provided by the REV + CdGπκ model, which allows different substitution parameter values (including two different C ↔ T and A ↔ G transition and four different C ↔ A, C ↔ G, T ↔ A, and T ↔ G transversion rates) and different fixed rates for codon positions, which are distributed statistically according to different gamma distributions. A similar analysis of deviance, conducted separately for the subset of the *saltans* sequences indicates that the REV model also provides the best description compared to more constrained submodels (results not shown).

Table 3 shows parameter estimates obtained with the REV + CdGπκ model under the two topologies shown in Figs. 1 and 2. Figure 3 provides a diagrammatic representation of the random (*a*) and fixed (*c*) components of the among-site rate variation across the *Xdh* region. Transition bias is not too strong, with transitions in the third coding position occurring at about 4 times ($R = 2.35$) faster than expected if both, transitions and transversions, occur at random. Substitution rates for the three sites are in the proportion $c_1{:}c_2{:}c_3 = 1{:}0.39{:}5.96$ with the REV + CdGπκ model; i.e., the third codon position changes about 15 times faster than the second and 6 times faster than the first (Fig. 3). The estimated gamma distribution for rates in the first and second positions is strongly L-shaped ($\alpha \ll 1$), meaning that the rates are highly heterogeneous, with more low-variable sites in the second ($\alpha_2 = 0.20 \pm 0.06$) than in the first ($\alpha_1 = 0.37 \pm 0.06$) codon position. In the third codon position, substitution rates fit a bell-shaped gamma distribution ($\alpha_3 = 7.70 \pm 2.87$) with most sites changing at intermediate rates (Fig. 3). These patterns remain unchanged for the subset of the *saltans* sequences except for a slightly higher rate of transitions in first and third codon positions, which is expected given that the species are closely related. Table 3 also gives the corresponding parameter estimates obtained under the REV + CdGπκ assuming the topology proposed in Fig. 2 (Throckmorton and Magalhães, 1962). Parameter values are virtually the same, which validates the use of Fig. 1 as a working topology.

By using an explicit model of evolution, the maximum likelihood method simultaneously accounts for Ts/Tv rate bias, unequal nucleotide frequencies, and rate variation among sites. Joint maximum likelihood estimates of these parameters are more reliable than separate estimates obtained by using traditional parsimony procedures (Yang, 1996a). Parsimonious methods tend to overlook substitutions at the fastest changing sites, producing overestimates of α (Yang, 1996a). Table 4 shows the estimates of the overall rate variation (α) across the *Xdh* region obtained with different methods. The α values obtained with these alternative methods (namely, parsimony method of moments of Golding, 1983; Sullivan *et al.,* 1995; Yang and Kumar, 1996), are larger than the joint ML estimates obtained with the REV + dG model (Table 4).

28S rRNA

Sequences from the *28S rRNA* region are available for all the species investigated in this study, except *D. saltans, D. subsaltans,* and *D. subobscura.* The sequences span 541 bp, considerably shorter than the *Xdh* sequences. The *28S rRNA* nucleotide frequencies

TABLE 2

Results of the Analysis of Deviance Carried out on the *Xdh* Data for all *Drosophila* Species in this Study

Assumptions	H_0	:	H_1	$-2 \log \Lambda$	*d.f.*	*P*
Equal base frequencies	JC69	:	F81	13.1	3	0.004
Transition rate equals transversion rate	JC69	:	HKY85	340.8	4	$<10^{-6}$
Equal transitional rates	HKY85	:	TN93	56.0	1	$<10^{-6}$
Equal transversional rates	TN93	:	REV	40.0	3	$<10^{-6}$
Uniform rates among-sites	REV	:	REV + dG	680.1	1	$<10^{-6}$
Uniform rates among-codon positions	REV + dG	:	REV + C + dG	1049.0	2	$<10^{-6}$
Uniform rates within-codon positions	REV + C + dG	:	REV + CdG	104.4	2	$<10^{-6}$
Equal base frequencies in codon positions	REV + CdG	:	REV + CdGπ	113.4	6	$<10^{-6}$
Equal rate ratio parameter values in codon	REV + CdGπ	:	REV + CdGπκ	89.3	10	$<10^{-6}$

Note. 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 ($-2 \log \Lambda$) if the null hypothesis were true, with degrees of freedom (*d.f.*) indicated and $\alpha = 0.01$. JC69: Jukes-Cantor, 1969; F81: Felsenstein, 1981; HKY85: Hasegawa-Kishino-Yano, 1985; TN93: Tamura-Nei, 1993; REV: general reversible model; REV + dG; assuming discrete gamma rates at sites; REV + C + dG; different fixed rates for codon positions and similar discrete gamma distributed rates for sites; REV + CdG: different fixed rates and gamma distributions for codon positions; REV + CdGπ: different fixed rates, gamma distributions, and nucleotide frequencies for codon positions; and REV + CdGπκ: different fixed rates, gamma distributions, nucleotide frequencies, and rate ratio parameters for codon positions.

TABLE 3

Analysis of the *Xdh* Rate Variation among Sites

Species	Topology	α_{1st}	α_{2nd}	α_{3rd}	c_{1st}	c_{2nd}	c_{3rd}	R_1	R_2	R_3
All	1	0.37 ± 0.06	0.20 ± 0.06	7.70 ± 2.87	1	0.39 ± 0.06	5.96 ± 0.58	1.51	0.90	2.35
	2	0.31 ± 0.05	0.14 ± 0.04	6.12 ± 1.98	1	0.40 ± 0.06	5.76 ± 0.58	1.62	0.86	2.36
Saltans	1	0.22 ± 0.06	0.21 ± 0.13	4.47 ± 2.14	1	0.35 ± 0.07	4.93 ± 0.63	2.06	0.88	2.68
Group	2	0.21 ± 0.05	0.12 ± 0.07	3.58 ± 1.37	1	0.37 ± 0.07	4.94 ± 0.63	2.09	0.89	2.63

Note. Topologies 1 and 2 are those shown in Figs. 1 and 2, respectively. All estimates are obtained with the REV + CdGπκ model.

are statistically homogeneous across taxa ($I = 12.63$, $P > 0.6$, 15 *d.f.*), with average frequencies of T = 0.35, C = 0.17, A = 0.36, and G = 0.12, even though these frequencies deviate substantially from 0.25 proportions. Table 5 shows the results of the analysis of deviance. The most satisfactory description of the substitution pattern between nucleotides is obtained with the HKY85 model ($L_0 = -980.4$), which allows unequal nucleotides frequencies at equilibrium, as well as different rates for transitions and transversions. Assuming two rates of transition ($a \neq f$) in TN93 or allowing six rate ratios ($a \neq b \neq c \neq d \neq e \neq f$) in REV does not significantly improve the likelihood ($-2 \log \Lambda = 1.5$; $P > 0.2$ and $-2 \log \Lambda = 7.6$; $P > 0.10$, respectively; data not shown in the table). Substitution rates can be assumed to be homogeneous along the *28S rRNA,* since HKY + dG is not significantly better than HKY85 ($-2 \log \Lambda = 2.6$; $P > 0.1$).

Table 6 shows the maximum likelihood estimates of parameters for only the *28S rRNA* region (under the HKY + dG model, top row) and combining the *28S rRNA* data with the *Xdh* data (under the REV + CdGπκ model, bottom row). Combining both data sets has little effect on the parameter estimates. The α value (0.40 ± 0.38) indicates that the *28S rRNA* substitution rates are very heterogeneous, which agrees with the reported existence of very short hyper-variable regions interspersed with large highly conserved domains in this region (Pélandakis and Solignac, 1993). The large standard error of α (attributable to the short length of the sequences) would explain why the variation of rates cannot be distinguished from a uniform distribution in the analysis of deviance. The overall rate of evolution of the *28S rRNA* sequences ($c = 0.25 \pm 0.05$) is slower than the corresponding rates in the second codon positions of *Xdh* ($c_2 = 0.39 \pm 0.06$, see Table 3), although the difference is not significant.

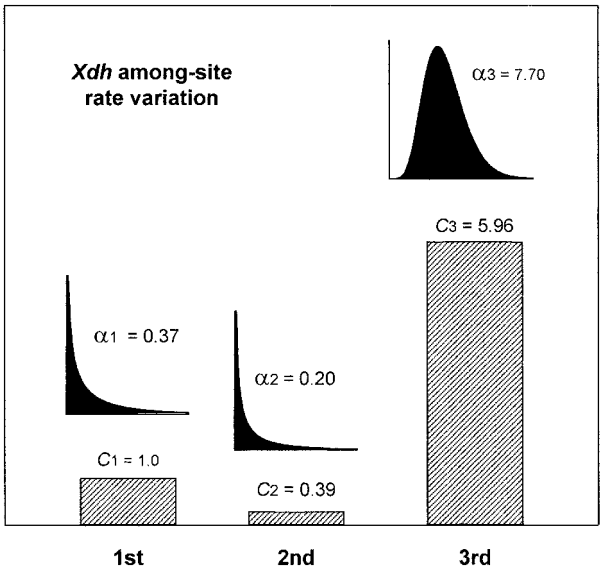

FIG. 3. Diagrammatic representation of the fixed (rate parameters) and random (gamma distributions) components of the rate variation among sites in *Xdh.* Vertical bars are scaled proportionally to rate parameter estimates for the first, second, and third codon positions (c_1, c_2, and c_3). Gamma distributions above the bars represent the rate variation among sites pertaining to a given codon position.

Adh, ITS1, COI, and COII

Sequences for all these four genes have been analyzed in order to resolve the *saltans* group phylogeny (O'Grady *et al.,* 1998). We found a satisfactory description (see below), the HKY85 model, assuming gamma-distributed rates for the conserved domain of the *ITS1* untranslated region (HKY + dGπ model; log L = −375.9) and allowing different categories of rates and nucleotides frequencies in codon positions for *Adh, COI,* and *COII* (HKY + Cπ + dG model; log L = −1307.3, log L = −858.6, log L = −1790.9, respectively). Parameter estimates were $R = 1.72$ and

TABLE 4

Estimates of the Rate Variation among Sites (α) across *Xdh* According to Different Methods

Method	α
Maximum parsimony	
Method of moments	1.41
Sullivan et al., 1995	0.98
Yang and Kumar, 1996	0.74
Maximum likelihood	
REV + dG model	0.43

Note. Maximum likelihood estimates are obtained assuming the topology in Fig. 1.

TABLE 5

The Results of the Analysis of Deviance Carried out on the *28S rRNA* Data (Pélandakis and Solignac, 1993)

Null hypothesis	H_0	:	H_1	$-2 \log \Lambda$	*d.f.*	*P*
Equal base frequencies	JC69	:	F81	107.4	3	$<10^{-6}$
Transition rate equals transversion rate	F81	:	HKY85	19.0	1	$<10^{-3}$
Equal transitional rates	HKY85	:	TN93	1.5	1	0.23
Uniform rates among sites	HKY85	:	HKY85 + dG	2.6	1	0.11

Note. 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 ($-2 \log \Lambda$) if the null hypothesis were true, with degrees of freedom (*d.f.*) indicated and $\alpha = 0.01$. JC69: Jukes-Cantor, 1969; F81: Felsenstein, 1981; HKY85: Hasegawa-Kishino-Yano, 1985; HKY + dG: assuming discrete gamma distributed rates for sites; TN93: Tamura-Nei, 1993.

$\alpha = 0.36 \pm 0.15$ for *ITS1*; $R = 1.41$, $\alpha = 2.84 \pm 1.74$, $c_2 = 0.44 \pm 0.14$, and $c_3 = 5.15 \pm 1.11$ for *Adh;* $R = 6.52$, $\alpha = 0.65 \pm 0.24$, $c_2 = 0.28 \pm 0.13$, and $c_3 = 14.64 \pm 5.97$ for *COI;* and $R = 2.79$, $\alpha = 0.42 \pm 0.10$, $c_2 = 0.40 \pm 0.13$, and $c_3 = 7.32 \pm 1.87$ for *COII*. It must be noticed, however, that the sequences from all these four regions are much shorter than those of *Xdh,* spanning 163, 390, 305, and 687 nucleotides for *ITS1, Adh, COI,* and *COII,* respectively, which is expected to yield a reduced power of the likelihood-ratio test. This circumstance could preclude more complex models (e.g., TN93 or REV) from fitting sequence evolution even though they are more realistic.

Phylogenetic Relationships of the saltans Group Species

Several simple methods for tree reconstruction (see Materials and Methods) yield the topology shown in Fig. 1 with the *Xdh* data. We have used this topology and the one proposed by Throckmorton and Magalhães (1962), shown in Fig. 2, as working hypotheses for modeling the molecular evolution of the sequence data by means of an analysis of deviance. The models identified will now be used to perform a finer reconstruction of the phylogenetic relationships in the *saltans* group. Figure 1 shows the maximum likelihood tree obtained for the *Xdh* data set with the REV + CdG$\pi\kappa$ model (see Table 3). This is a fully resolved or strictly bifurcating tree, which coincides with the working topology also presented by this Fig. 1. According to this maximum likelihood tree, the *parasaltans* subgroup (represented by *D. subsaltans*) is the oldest lineage, followed in successive progression by the *sturtevanti* subgroup (*D. sturtevanti*), the *elliptica* subgroup (*D. emarginata*), the *cordata* subgroup (*D. neocordata*), and the *saltans* subgroup (*D. saltans* and *D. prosaltans*). These relationships are consistent with the phylogenetic hypothesis proposed by Throckmorton and Magalhães (1962) shown in Fig. 2 only in the position of the *saltans* subgroup as the latest derived clade (also in that the *elliptica* and *cordata* subgroups are adjacent to one another). In contrast with the maximum likelihood tree derived from the *Xdh* data (Fig. 1), these authors place the *elliptica* and *cordata* subgroups as the most primitive, followed by the *sturtevanti* and then the *parasaltans* subgroups. The *parasaltans* subgroup, as noted, is the most ancient in the *Xdh* phylogeny.

We have also used estimates of the rate variation among sites and transition bias obtained with the REV + CdG$\pi\kappa$ model for phylogenetic reconstruction by distance and maximum parsimony criteria. Figure 4A shows the NJ tree based on the TN93 model (see Table 2; Tamura and Nei, 1993) and using the complete *Xdh* data set. The NJ tree differs from the maximum likelihood tree only in that the *cordata* and the *elliptica* subgroups appear as a monophyletic cluster, with high bootstrap support. Nevertheless, this branching may reflect inconsistencies in the NJ algorithm that arise from homoplasic changes at rapidly evolving third codon positions. If only transversions are used (which are less affected by unobserved changes than transitions) and setting $\alpha = 0.43$ (according to the maximum likelihood estimate shown in Table 4), the NJ tree obtained (Fig. 4B) has a topology identical to that of maximum likelihood. Analogously, unweighted maximum parsimony analysis yields two equally most parsimonious trees that are 1678 steps long and correspond to the topologies shown in Figs. 4A and 4B. If we set the overall transition/transversion ratio (R) equal to 1.6 (average of R_1, R_2, and R_3 values in Table 3) (stepmatrix 8 by 5 in *PAUP*) and weighing 3:15:1 the first, second, and third codon positions data (according to estimated *c* values shown in Table 3), the most parsimonious tree is Fig. 4C (and Fig. 1); bootstrap values for this tree are almost identical to the bootstrap proportions for the NJ tree shown in Fig. 4B.

Figure 4D shows the NJ tree derived from the *28S rRNA* sequences (Pélandakis and Solignac, 1993; the sequences for *D. saltans* and *D. subsaltans* are not available). The same topology is obtained with maxi-

TABLE 6

Parameter Estimates for the *28S rRNA* Region

Data set	α	*c*	R
28S rRNA[a]	0.40 ± 0.38	—	1.60
28S rRNA + *Xdh*[b]	0.43 ± 0.42	0.25 ± 0.05	1.58

[a] Estimates obtained with the HHY + dG model.
[b] Estimates obtained with the REV + CdG$\pi\kappa$ model.

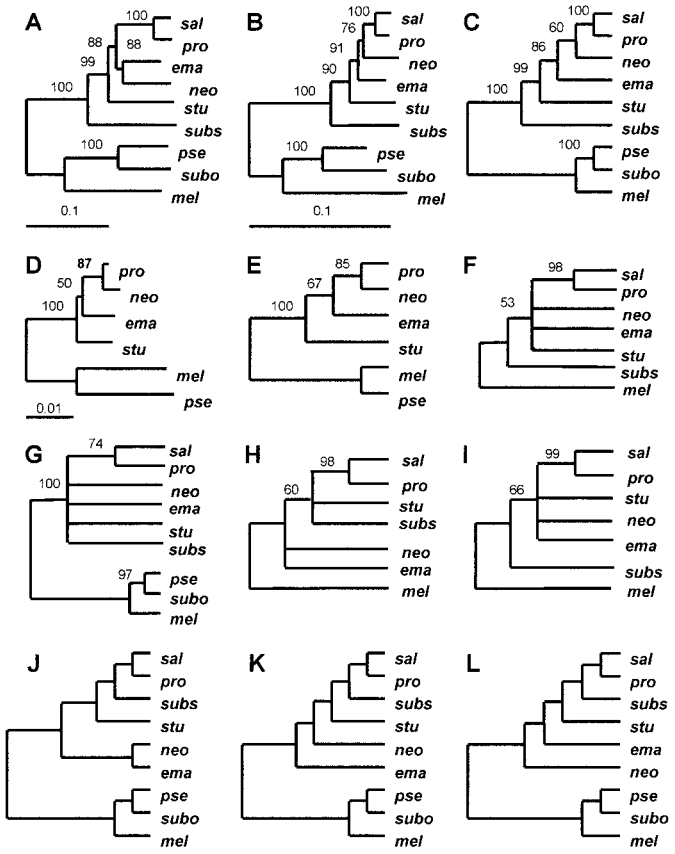

FIG. 4. (A and B) Neighbor-Joining (NJ) trees based on the Tamura–Nei distances for *Xdh,* assuming equal rates for sites and using all substitutions (A) and using only transversions with $\alpha = 0.43$ (B); (C) the single most parsimonious tree derived from *Xdh,* weighing 5 transitions and 8 transversions and weighing 3, 15, and 1 substitutions at first, second, and third codon positions, respectively; (D) NJ tree based on *28S rRNA* sequences using the HKY85 model; (E) the single most parsimonious tree derived from *28S rRNA,* weighing 5 transitions and 8 transversions; (F–I) consensus maximum-parsimony trees derived from *ITS1* (F), *Adh* (G), *COI* (H), and *COII* (I); (J–L) three binary topologies consistent with the tree shown in Fig. 2, proposed by Throckmorton and Magalhães (1962). Branch lengths are proportional to the scale given in substitutions per nucleotide. Percentage bootstrap values (based on 1000 pseudo replications) are given on the nodes for trees A–I.

mum likelihood under the HKY85 model and with the NJ algorithm based on the maximum likelihood distance implemented in the Felsenstein's DNAML program using a transition/transversion ratio $R = 1.6$ (Table 7). It is also the most parsimonious tree obtained under the maximum parsimony criterion (Fig. 4E). This topology is the same shown in Fig. 4B but with fewer species. Figures 4F–4I are the maximum parsimony trees derived from *ITS1, Adh, COI,* and *COII,* respectively. In accordance with previous results (O'Grady *et al.,* 1988), none of these genes allows statistically significant discrimination among any of the trees shown in Fig. 4 (see below, Table 7). The trees shown in Figs. 4J, 4K, and 4L correspond to Fig. 2, the phylogenetic hypothesis of Throckmorton and Magalhães (1962).

Figure 5 shows the NJ trees derived from four different combinations of the data sets, including *Xdh + Adh + ITS1* alone, combined separately with *COI* and *COII,* and all the sequence data pooled together. The same topologies are obtained with maximum likelihood under the REV + dG model and with the NJ algorithm based on the TN93 distance and assuming heterogeneous substitution rates among-sites. The four data sets yield the same topology, which is similar to the topology shown in Fig. 4A but with fewer species. Bootstrap values clearly support the position of the *parasaltans* subgroup as the first derived clade.

Table 7 shows the results of the Kishino–Hasegawa's (1989) test for the different phylogenetic hypotheses shown in Fig. 4. The tests include the outgroups and are conducted using models as implemented in *PAUP** (version 4d64ppc, Swofford, 1998): REV + dG for *Xdh,* HKY85 for *28S rRNA,* and HKY85 + dG for the *ITS1, Adh, COI,* and *COII* data sets. The data in Table 7 (see also Fig. 4) show that *Xdh* contains a strong phylogenetic signal, showing hypotheses A and B as statistically superior to J, K, and L. None of the other five genes allows statistical discrimination among the hypotheses shown in Fig. 4. When the data for all genes

TABLE 7

Kishino-Hasegawa's (1989) Tests for the Different Hypotheses about the Phylogeny of *saltans*

Hypotheses	*Xdh* (2085 bp)	*28S rRNA* (541 bp)	*ITS1* (163 bp)	*Adh* (390 bp)	*COI* (303 bp)	*COII* (687 bp)	Combined (3642 bp)
A	Best	8.07	1.10	2.44	7.50	Best	Best
B	0.34	Best	Best	5.27	7.62	1.47	6.54
J	28.29**	10.49	1.62	0.35	0.36	3.15	17.08†
K	33.40**	10.49	1.62	0.35	0.27	3.78	18.31†
L	32.86**	8.36	1.30	Best	Best	1.94	16.04

Note. Hypotheses A, B, and J–L are according to Fig. 4. Hypotheses A (all substitutions) and B (only transversions) are based on *Xdh*; hypotheses J–L are variations of the relationships proposed by Throckmorton and Magalhães (1962); hypotheses D–I in Fig. 4 are not included in this table because D and E are topologically identical to B but with fewer species and because F–I are less resolved versions of A, B, or J–L. Log-likelihood values are obtained with the REV + dG model for the *Xdh* and combined data sets, the HKY85 model for the *28S rRNA* data set, and the HKY + dG for the *ITS1, Adh, COI,* and *COII* data sets. In the case of the *28S rRNA*, tests are conducted without *D. saltans* and *D. subsaltans.* Numbers between parentheses are the lengths of the sequences. * $P < 0.05$, ** $P < 0.01$, † $P < 0.07$.

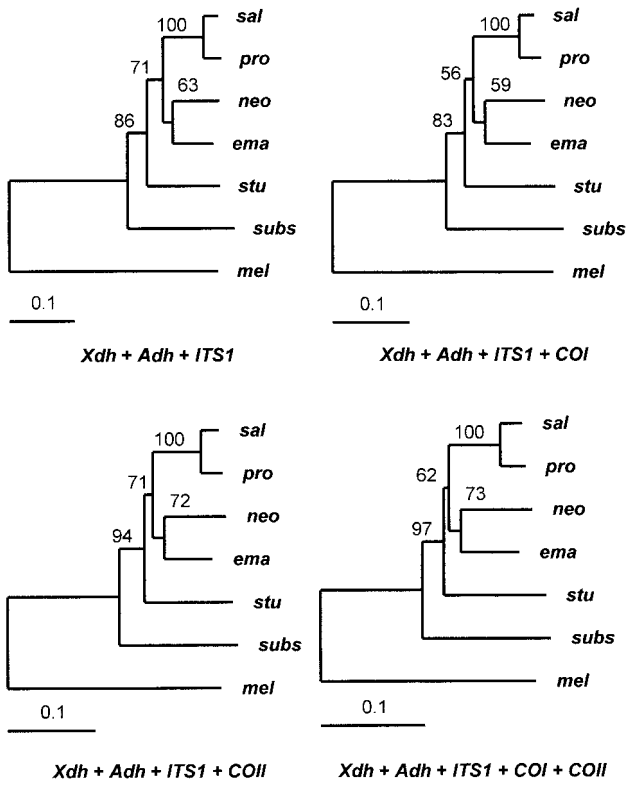

FIG. 5. Neighbor-joining trees based on Tamura–Nei distances for four different combinations of five gene sequences, using all substitutions and assuming gamma-distributed rates among-sites. Estimates of the parameter α of the gamma distribution for each data set combination are obtained by maximum likelihood under the REV + dG model; they were 0.38 ± 0.03 for *Xdh + Adh + ITS1* with and without *COI* and 0.35 ± 0.02 for *Xdh + Adh + ITS1 + COII* and for the five data sets pooled together.

(except the *28S rRNA,* containing fewer sequences) are combined with *Xdh,* trees A and B remain the best, but the strong phylogenetic signal in *Xdh* becomes diluted so that the combined data are less discriminating than *Xdh* alone (see Fig. 5). In any case, the *Xdh* data lead to the rejection of the phylogeny proposed by Throckmorton and Magalhães (1962; Fig. 2) in any of the alternative representations displayed in Figs. 4J, 4K, and 4L.

DISCUSSION

Phylogenetic methods may produce biased phylogenies when the methods' assumptions are violated. Assuming equal nucleotide frequencies across taxa, a single substitution rate for sites or constant evolutionary rates for lineages are common premises that are frequently contravened (review in Swofford *et al.* 1996). A phylogenetic methodology should, therefore, include some means of testing the assumptions implicit in the analyses (e.g., Ritland and Clegg, 1987; Yang, 1995; Kumar, 1997; Huelsenbeck and Crandall, 1997). It is for this reason that we have chosen to use maximum likelihood methods (although see Rzhetsky and Nei, 1995). Through the statistical comparison of hierarchical explicit models of sequence change, these methods provide critical information about relevant aspects of the evolutionary process in the *Xdh* and other genes that we have analyzed. Circularity may occur because testing the assumptions depends on knowing the true phylogenetic relationships among the species, the *saltans* group in our case. We avoid the pitfalls of circularity because the two tentative topologies at stake (Figs. 1 and 2), which are substantially different, yield virtually identical results concerning the relevant evolutionary parameters. In this respect, our conclusions strengthen the results from other studies (Yang, 1994; Yang *et al.* 1994, 1995), indicating that, in general, tree topology differences have only a minor effect on parameter estimates. Thus, we concluded that a "realistic" description of the evolution of *Xdh* should incorporate significant transition bias and extreme rate variation among sites along the region. We took these results into account before proceeding to test alternative phylogenetic hypotheses.

The five subgroups of the *D. saltans* group had been unambiguously defined morphologically (Magalhães, 1962). But their phylogenetic relationships have received little attention compared to other groups of the *Sophophora* subgenus (Throckmorton, 1975). On the basis of anatomical features, primarily male and female genitalia (Magalhães, 1962), Throckmorton and Magalhães (1962) proposed the phylogeny shown in Fig. 2, which places the *cordata* and *elliptica* lineages as the oldest, followed successively by *sturtevanti, parasaltans,* and *saltans.* These relationships were proposed, however, tentatively, noting that the only unequivocal conclusion was that the *saltans* subgroup was the most derivative (Throckmorton, 1975). Based on this phylogeny, combined with evidence on the contemporary species distribution and geological information, Throckmorton (1975) proposed that the ancestor of the *saltans* species group originated in North America, where the lineage diversified giving rise to the ancestors of the primitive *cordata* and *elliptica* subgroups. Subsequently, prior to the rising of the present day isthmus of Panama, a progenitor from North America crossed into South America, where further diversification produced the more derived *sturtevanti, parasaltans,* and *saltans* subgroups. According to this scenario, the *saltans* group would have had a different geographical origin than the closely related *willistoni* group, which is thought to have originated in tropical South America (Throckmorton, 1975).

Earlier analyses of the *28S rRNA* region did not support Throckmorton's hypothesis, clustering *D. sturtevanti* with *D. emarginata* and *D. prosaltans* with *D. neocordata* into two separated monophyletic clades (Pélandakis *et al.,* 1991; Pélandakis and Solignac, 1993). However, these molecular studies were aimed at

ascertaining the phylogenetic relationships among the species groups of *Sophophora,* as well as other higher taxonomic categories within *Drosophila* and related genera. Consequently the authors disregarded hypervariable, allegedly saturated sites and relied on more conserved positions (Pélandakis *et al.,* 1991; Pélandakis and Solignac 1993). The sequences analyzed contained little information to resolve the relationships on the relatively short time-scale involved in the diversification of the *saltans* subgroups.

The phylogeny of *saltans* has been more recently addressed using data from the *Adh, ITS1, COI,* and *COII* nucleotide regions (O'Grady *et al.,* 1998). The sequences analyzed are relatively short, so that separately no region allows one to ascertain the branching order among the major subgroups of *saltans.* Combining all the evidence provides support for the position of *elliptica* and *cordata* as the primitive subgroups, in agreement with Throckmorton and Magalhães (1962; Throckmorton, 1975). However, (i) this result is based on a maximum parsimony analysis that does not take into account differences in substitution rates among and within the regions. According to our maximum likelihood estimates, there is extensive variation in substitution rates, ranging from 0.40 ± 0.13 substitutions per site for the second codon positions of *COII* to 8.57 ± 1.97 for the third codon positions of *COI* (i.e., sites evolving fastest change ~20 times faster than the slowest sites); (ii) the study includes an alignment of the complete *ITS1* region (785 bp long according to O'Grady *et al.,* 1988); given the lengths of the *ITS1* region in the *saltans* group species, which are distinctively shorter (ranging from 328 to 419 bp) than this region lengths in *D. melanogaster* (727 bp; Schlotterer, 1994) and *D. yakuba* (820 bp; Tautz *et al.,* 1988), the inference of homology is probably wrong for most of the alignment; we can confidently align only 163 bp; (iii) the study refers to 771 bp aligned for *Adh* (see Table 2 in O'Grady *et al.,* 1998), even though only exon 2 (i.e., 405 bp) was sequenced by the authors, and the sequence for *D. subsaltans* is even shorter (390 bp) owing mainly to unknowns in the most variable, i.e. most informative, third codon positions; with sequences this short in length, absence of conflict between data partitions reported by the authors can simply result from partition homogeneity tests lacking enough power to detect differences; (iv) the study does not consider alternative phylogenetic hypotheses.

The *Xdh* coding region yields two slightly different phylogenies that are statistically superior to the alternatives (Table 7). The greater capability of the *Xdh* sequences for discriminating among competing hypothesis is probably because they comprise a substantial number of nucleotides (2085 bp). Moreover, the remaining molecular evidence (Pélandakis *et al.,* 1991; Pélandakis and Solignac, 1993; O'Grady *et al.,* 1998), either separately (3 out of 5 genes) or combined into a single data set, favors the same two phylogenies (Fig. 5, Table 7). Stability of the same basic topology across several different gene regions suggests that it reflects the true phylogenetic relationships. According to this phylogeny, the earliest derived taxon is the *parasaltans* subgroup, rather than the *elliptica* and *cordata* subgroups, as proposed by Throckmorton and Magalhães (1962). These two taxa originate later, after the further split of the *sturtevanti* subgroup. The *Xdh* phylogeny agrees, however, with that of Throckmorton and Magalhães (1962) in the position of the *saltans* subgroup as the most derivative and also in that *elliptica* and *cordata* are subgroups closely related to one another. The relative position of these two taxa remains to be elucidated. From the evidence at hand either (i) they could cluster in a monophyletic group or (ii) *elliptica* could have originated first. The first possibility is supported by the distribution of *P* elements in the *saltans* group: *elliptica* and *cordata* are the only subgroups lacking *P* elements (Clark and Kidwell, 1997), which can be most parsimoniously explained invoking one single loss in their common ancestor, rather than two independent losses.

The geographic distribution of the *parasaltans* subgroup, oldest in the phylogeny, seems to be circumscribed to Brazil (*D. subsaltans* and *D. parasaltans*) and some Caribbean islands (*D. pulchella*) (Magalhães, 1962). The phylogeny in Fig. 1 thus would favor tropical South America rather than North America as the place of origin of the *saltans* group. Accordingly, the *saltans* group shared the same South American origins as the *willistoni* group, with which it is closely related.

The molecular clock is an assumption that rates of substitution are constant along different parts of the tree. Previous results using a relative rate test across pairs of species, with *Scaptodrosophila lebanonensis* as outgroup, have revealed that the *saltans* lineage is evolving about twice as fast as the lineage leading to the *melanogaster* and *obscura* groups in the *Xdh* region (Rodríguez-Trelles *et al.,* unpublished results). In order to know whether substitution rates among lineages are homogeneous within the *saltans* species group, we have conducted likelihood ratio tests of the molecular clock hypothesis for the six *saltans* species separately. Strictly speaking, this comparison is valid only if the likelihood values are calculated using the true topology, and caution is needed when the phylogeny is uncertain (Yang *et al.,* 1995). We alleviate this problem by using tree topologies A and B, rated the best in Table 7, for the comparison. The REV + CdG$\pi\kappa$ model is used to calculate the likelihood values either with or without the clock assumption. The REV + CdG$\pi\kappa$ clock hypothesis is rejected for the topology B, and the difference is marginally significant for the topology A ($-2 \log \Lambda = 21.3$; $P < 10^{-3}$, and $-2 \log \Lambda = 9.5$; $P < 10^{-3}$, respectively, each test with five degrees of freedom). In principle, the molecular clock hypothesis

appears to be a statistically insufficient description of the evolutionary process of the *Xdh* sequences in the *saltans* species group.

Lack of congruence between molecular and morphological trees is a commonplace in systematics (Patterson *et al.,* 1993; Olmstead and Sweere, 1994). Topological conflicts may, in fact, be useful for unveiling interesting underlying evolutionary processes. The retention of hypothetically ancestral anatomical features in *elliptica,* which led Throckmorton (1975) to consider this and the *cordata* subgroup as primitive, may be related to a slower rate of molecular evolution of *D. emarginata* at the *Xdh* locus (Fig. 1). A similar state of affairs does not, however, occur in *D. neocordata,* which is rather one of the fastest evolving *saltans* species at this locus (Fig. 1).

One final observation: choosing the right gene is a principle of economics in molecular phylogenetics. The phylogenetic utility of a gene is intimately related to its rate of evolution. Characters are phylogenetically informative when their rate of evolution matches the relevant divergence times. If the characters evolve too slowly they will be uninformative, and if they evolve too fast, homoplasy will obscure the phylogenetic signal. The *Xdh* gene has been employed for the first time in this study as a molecular marker for phylogeny. High rates of synonymous substitution have been previously reported for this region (Riley *et al.,* 1992). Indeed, we find that substitution rates are extremely heterogeneous, with third codon positions changing about 15 times faster than second codon positions. The use of *Xdh* as a nuclear DNA marker may be particularly useful in *Drosophila* for phylogenetic studies that seek to resolve evolutionary relationships among recently derived taxa (i.e., within species groups or subgroups).

ACKNOWLEDGMENTS

We are indebted to Ziheng Yang for advice in using PAML and to Carlos Machado and Andrei Tatarenkov for valuable suggestions. F.R.-T. has received support from Ministerio de Educación y Cultura (Spain) (Contrato de Reincorporación) and Grant PB96-1136 to A. Fontdevila. Research supported by NIH Grant GM42397 to F.J.A.

REFERENCES

Bicudo, H. E. M. C. (1973a). Chromosomal polymorphism in the *saltans* group of *Drosophila.* I. The *saltans* subgroup. *Genetica* **44:** 550–552.

Bicudo, H. E. M. C. (1973b). Reproductive isolation in the *saltans* group of *Drosophila.* I. The *saltans* subgroup. *Genetica* **44:** 313–329.

Clark, J. B., and Kidwell, M. G. (1997). A phylogenetic perspective on *P* transposable element evolution in *Drosophila. Proc. Natl. Acad. Sci. USA* **94:** 11428–11433.

Clark, J. B., Altheide, T. K., Schlosser, M. J., and Kidwell, M. G. (1995). Molecular evolution of *P* transposable elements in the genus *Drosophila.* I. The *saltans* and *willistoni* Species Groups. *Mol. Biol. Evol.* **12:** 902–913.

Daniels, S. B., and Strausbaugh, L. D. (1986). The distribution of *P*-element sequences in Drosophila: The *willistoni* and *saltans* Species Groups. *J. Mol. Evol.* **23:** 138–148.

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: A maximum likelihood approach. *J. Mol. Evol.* **17:** 368–376.

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39:** 783–791.

Felsenstein, J. (1993). PHYLIP, phylogenetic inference package and documentation. Version 3.5c. Distributed by the author, Department of Genetics, Univ. of Washington, Seattle.

Golding, G. B. (1983). Estimates of DNA and protein sequence divergence: An examination of some assumptions. *Mol. Biol. Evol.* **1:** 125–142.

Goldman, N. (1993). Statistical tests of models of DNA substitution. *J. Mol. Evol.* **36:** 182–198.

Hasegawa, M., Kishino, H., and Yano, T. (1985). Dating the human–ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22:** 160–174.

Huelsenbeck, J. P., and Crandall, K. A. (1997). Phylogeny estimation and hypothesis testing using maximum likelihood. *Annu. Rev. Ecol. Syst.* **28:** 437–466.

Jukes, T. H., and Cantor, C. R. (1969). Evolution of protein molecules. *In* "Mammalian Protein Metabolism" (H. N. Munro, Ed.), pp. 21–132. Academic Press, New York.

Kawasaki, E. S. (1990). Sample preparation from blood, cells, and other fluids. *In* "PCR Protocols: A Guide to Methods and Applications" (M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White, Eds.), pp. 146–152. Academic Press, San Diego.

Keith, T. P., Riley, M. A., Kreitman, M., Lewontin, R. C., Curtis, D., and Chambers, G. (1987). Sequence of the structural gene for xanthine dehydrogenase (rosy locus) in *Drosophila melanogaster. Genetics* **116:** 67–73.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16:** 111–120.

Kishino, H., and Hasegawa, M. (1989). Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *J. Mol. Evol.* **29:** 170–179.

Kwiatoski, J., Skarecky, D., Hernández, S., Pham, D., Quijas, F., and Ayala, F. J. (1991). High fidelity of the polymerase chain reaction. *Mol. Biol. Evol.* **8:** 884–887.

Kumar, S. (1996). Patterns of nucleotide substitution in mitochondrial protein coding genes of vertebrates. *Genetics* **143:** 537–548.

Lockhart, P. J., Steel, M. A., Hendy, M. D., and Penny, D. (1994). Recovering evolutionary trees under a more realistic model of sequence evolution. *Mol. Biol. Evol.* **11:** 605–612.

Magalhães, L. E. (1962). Notes on the taxonomy, morphology, and distribution of the *saltans* group of *Drosophila,* with descriptions of four new species. *Univ. Tex. Pub.* **6205:** 135–154.

Magalhães, L. E., and Björnberg, A. J. S. (1957). Estudo da genitalia masculina de *Drosophila* (Diptera). *Rev. Bras. Biol.* **16:** 273–280.

Mourão, C. A., and Bicudo, H. E. M. C. (1967). Duas novas espécies de *Drosophila* do grupo *saltans* (*Drosophilidae,* Diptera). *Pap. Dept. Zool. S. Paulo* **20:** 123–134.

O'Grady, P. M., Clark, J. B., and Kidwell, M. G. (1988). Phylogeny of the *Drosophila saltans* Species Group based on combined analysis of nuclear and mitochondrial DNA sequences. *Mol. Biol. Evol.* **15:** 656–664.

Olmstead, R. G., and Sweere, J. A. (1994). Combining data in phylogenetic systematics: An empirical approach using three molecular data sets in the *Solanaceae. Syst. Biol.* **43:** 467–481.

Patterson, C., Williams, D. M., and Humphries, C. J. (1993). Congruence between molecular and morphological phylogenies. *Annu. Rev. Ecol. Syst.* **24:** 153–188.

Patterson, J. T., and Stone, W. M. (1952). "Evolution in the Genus *Drosophila,*" MacMillan Co., New York.

Pélandakis, M., and Solignac, M. (1993). Molecular phylogeny of *Drosophila* based on ribosomal RNA sequences. *J. Mol. Evol.* **37:** 525–543.

Pélandakis, M., Higgins, D. G., and Solignac, M. (1991). Molecular phylogeny of the subgenus *Sophophora* of *Drosophila* derived from large subunit of ribosomal RNA sequences. *Genetica* **84:** 87–94.

Powell, J. R. (1997). "Progress and Prospects in Evolutionary Biology: The *Drosophila* Model," Oxford Univ. Press, New York.

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

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

Ritland, K., and Clegg, M. T. (1987). Evolutionary analysis of plant DNA sequences. *Am. Nat.* **130:** S74–S100.

Rzhetsky, A., and Nei, M. (1995). Tests of the applicability of several substitution models for DNA sequence data. *Mol. Biol. Evol.* **12:** 131–151.

Schlotterer, C. (1994). Comparative evolutionary analysis of rDNA regions in *Drosophila. Mol. Biol. Evol.* **11:** 513–522.

Spassky, B. (1957). Morphological differences between siblings species of *Drosophila. Univ. Tex. Pub.* **5721:** 48–61.

Sturtevant, A. H. (1942). The classification of the genus *Drosophila,* with descriptions of nine new species. *Univ. Tex. Pub.* **4213:** 6–51.

Sullivan, J., Holsinger, K. E., and Simon, C. (1995). Among-site rate variation and phylogenetic analysis of the *12S rRNA* in sigmodontine rodents. *Mol. Biol. Evol.* **12:** 988–1001.

Swofford, D. L. (1998). PAUP: phylogenetic analysis using parsimony. Version 4. Smithsonian Institution, Washington, DC.

Swofford, D. L., Olsen, G. J., Waddell, P. J., and Hillis, D. M. (1996). Phylogenetic inference. *In* "Molecular Systematics" (D. M. Hillis, C. Moritz, and B. K. Mable, Eds.), pp. 407–514. Sinauer, Sunderland, MA.

Tamura, K., and Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpazes. *Mol. Biol. Evol.* **10:** 512–526.

Targa, H. J. (1973). Comparisons of gene arrangements in four species of the *saltans* group of *Drosophila. Rev. Bras. Biol.* **33:** 353–360.

Tarrío, R., Rodríguez-Trelles, F., and Ayala, F. J. (1998). New *Drosophila* introns originate by duplication. *Proc. Natl. Acad. Sci. USA* **95:** 1658–1662.

Tautz, D., Hancock, J. M., Webb, D. A., Tautz, C., and Dover, G. A. (1988). Complete sequences of the *rRNA* genes of *Drosophila melanogaster. Mol. Biol. Evol.* **5:** 366–376.

Tavaré, S. (1986). Some probabilistic and statistical problems on the analysis of DNA sequences. *Lect. Math. Life Sci.* **17:** 57–86.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). *CLUSTAL W:* Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22:** 4673–4680.

Throckmorton, L. H. (1975). The phylogeny, ecology and geography of *Drosophila. In* "Handbook of Genetics" (R. C. King, Ed.), Vol. 3, pp. 421–436. Plenum, New York.

Throckmorton, L. H., and Magalhães, L. E. (1962). XVI. Changes with evolution of pteridine accumulations in species of the saltans group of the genus *Drosophila. Univ. Tex. Pub.* **6205:** 489–505.

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. (1997). PAML: phylogenetic analysis by maximum likelihood. Version 1.3. Distributed by the author, Department of Integrative Biology, University of California, Berkeley.

Yang, Z., and Kumar, S. (1996). New parsimony-based methods for estimating the pattern of nucleotide substitution and the variation of substitution rates among sites and comparison with likelihood methods. *Mol. Biol. Evol.* **13:** 650–659.

Yang, Z., Lauder, I. J., and Lin, H. J. (1995). Molecular evolution of the Hepatitis B virus genome. *J. Mol. Evol.* **41:** 587–596.

Yang, Z., Goldman, N., and Friday, N. E. (1994). Comparison of models for nucleotide substitution used in maximum likelihood phylogenetic estimation. *Mol. Biol. Evol.* **11:** 316–324.