Molecular Evolution and Phylogeny of the *buzzatii* Complex (*Drosophila repleta* Group): A Maximum-Likelihood Approach

Francisco Rodríguez-Trelles, Lourdes Alarcón, and Antonio Fontdevila

Grup de Biologia Evolutiva, Departament de Genètica i Microbiologia, Universitat Autònoma de Barcelona, Barcelona, Spain

The *buzzatii* complex of the *mulleri* subgroup (*Drosophila repleta* group) consists of three clusters of species whose evolutionary relationships are poorly known. We analyzed 2,085 coding nucleotides from the *xanthine dehydrogenase* (*Xdh*) gene in the 10 available species of the complex and *Drosophila mulleri* and *Drosophila hydei.* We adopted a statistical model-fitting approach within the maximum-likelihood (ML) framework of phylogenetic inference. We first modeled the process of nucleotide substitution using a tree topology which was reasonably accurate. Then we used the most satisfactory description so attained to reconstruct the evolutionary relationships in the *buzzatii* complex. We found that a minimally realistic description of the substitution process of *Xdh* should allow six substitution types and different substitution rates for codon positions. Using this description we obtained a strongly supported, fully resolved tree which is congruent with the already-known (yet few) relationships. We also analyzed published data from three mitochondrial *cytochrome oxidases* (*CO I, II,* and *III*). In our analyses, these relatively short DNA sequences failed to discriminate statistically among alternative phylogenies. When the data of these three gene regions are combined with the *Xdh* sequences, the phylogenetic signal emerging from *Xdh* becomes reinforced. All four of the gene regions evolve faster in the *buzzatii* and *martensis* clusters than in the *stalkeri* cluster, paralleling the amount of chromosomal evolution.

Introduction

The *buzzatii* species complex is one of several complexes that comprise the *mulleri* subgroup of the *repleta* group, subgenus *Drosophila.* At present, the *buzzatii* complex consists of 11 species, which, on the basis of their polytene banding patterns and reproductive morphologies (Wasserman 1982; Ruiz and Wasserman 1993), have been classified into three clusters: *stalkeri* (including *Drosophila richardsoni* and *Drosophila stalkeri*), restricted to the Caribbean islands and Florida (Wasserman 1982; Vilela 1983); *martensis* (*Drosophila martensis, Drosophila uniseta, Drosophila venezolana,* and *Drosophila starmeri*), found in the deserts of Colombia and Venezuela (Wasserman and Koepfer 1979; Ruiz and Fontdevila 1981); and *buzzatii* (*Drosophila buzzati, Drosophila koepferae, Drosophila borborema, Drosophila serido,* and *Drosophila seriema*), inhabiting the sub-Amazonian arid regions of Brazil, Bolivia, and Argentina (Ruiz, Fontdevila, and Wasserman 1982; Wasserman and Richardson 1987; Fontdevila et al. 1988; Tido-Sklorz and Sene 1995).

The *buzzatii* complex includes endemic and subcosmopolitan species (e.g., *D. buzzatii,* a recent colonist in the Old World and Australia; Fontdevila et al. 1982; Barker et al. 1985), with instances of sympatric (the *martensis* cluster species) and allopatric (the *stalkeri* cluster) distributions. The reproductive affinities in the complex (Marín et al. 1993) suggest the occurrence of prezygotic and postzygotic isolation mechanisms. One species, *D. serido,* is considered a superspecies consisting of many semi-isolated populations in the process of split (Sene, Pereira, and Vilela 1982; Fontdevila et al. 1988). The ecological niche and population dynamics of the species are amenable to study (Hasson, Naveira, and Fontdevila 1992; Santos, Ruiz, and Fontdevila 1989; Barker 1982). For these reasons, the *buzzatii* complex has attracted much attention from evolutionary geneticists. Yet, the phylogeny of the complex remains to be elucidated.

Former analyses of the overlapping paracentric inversions of the polytene chromosome 2 led Wasserman (1982) to include the *martensis* and *buzzatii* clusters within the *mulleri* complex and to consider the *stalkeri* cluster a distinct complex within the *mulleri* subgroup. Subsequent closer inspection of the chromosomal inversion patterns forced Ruiz and Wasserman (1993) to raise *buzzatii* to the rank of a separate complex within the *mulleri* subgroup. Unfortunately, chromosomal inversion data contained too little information to settle the relationships within the clusters and their branching order (Ruiz and Wasserman 1993). The most parsimonious reconstruction, which places *stalkeri* as the oldest cluster of the complex, differs by only a single inversion from a hypothesis invoking the *martensis* cluster as the earliest derived clade (Ruiz and Wasserman 1993). The phylogeny of the *buzzatii* complex has more recently been addressed using mitochondrial DNA markers (Spicer 1995). The sequences analyzed, however, evolved much too fast to resolve the relationships on the relatively short timescale involved in the diversification of the *buzzatii* complex (Spicer 1995).

The present study seeks to determine the phylogenetic relationships in the *buzzatii* complex using molecular data. We sequenced part of the *Xdh* coding region in 10 species and also analyzed published data from three mitochondrial loci (Spicer 1995). We followed a statistical model-fitting approach within the maximum-likelihood (ML) framework of phylogenetic inference (e.g., Ritland and Clegg 1987; Yang, Lauder, and Lin 1995; Kumar 1996; Rodríguez-Trelles, Tarrío, and Ayala 1999).

Key words: *Drosophila buzzatii* complex, *xanthine dehydrogenase, cytochrome oxidase,* molecular evolution, maximum-likelihood phylogeny, molecular clock.

Address for correspondence and reprints: Francisco Rodríguez-Trelles, Instituto de Investigaciones Agrobiológicas de Galicia (CSIC), Avenida de Vigo sln, 15706-Santiago de Compostela, Spain. E-mail: ftrelles@iiag.cesga.es.

Mol. Biol. Evol. 17(7):1112–1122. 2000

Materials and Methods

Drosophila Strains

The 10 strains of the *D. buzzatii* complex are as follows (stock reference numbers for the strains obtained from the National *Drosophila* Services Resource Center in Bowling Green State University, Ohio, are given in parentheses): *buzzatii* cluster: *D. buzzatii,* from Berna, Argentina, *D. koepferae* A, Quilmes, Argentina, *D. koeferae* B, San Isidro, Bolivia, *D. borborema,* Cafarnaum, Brazil (15081-1431.1), and *D. serido,* Rio Paraguassa, Brazil (15081-1431.1); *martensis* cluster: *D. martensis* and *D. venezolana,* Guaca, Venezuela, *D. uniseta,* Salamanca, Colombia, and *D. starmeri,* Riohacha, Colombia; *stalkeri* cluster: *D. richardsoni,* La Parguera, Puerto Rico (15801-1421.0), and *D. stalkeri,* Saint Petersburg, Fla. (15081-1321.0). As outgroups, we used *D. mulleri,* Panuco, Mexico (15081-1371.1), of the *mulleri* complex, and *D. hydei,* of the *hydei* subgroup (from the Stock Center at Indiana University, Bloomington).

Molecular Methods

The *Xdh* coding region investigated spans 2,085 bp (~52% of the total *Xdh* codons), including about half of exon II (1,113 bp) and most of exon III (972 bp). Genomic DNA was prepared from 0.2 g of flies according to the method of Piñol et al. (1988). The *Xdh* region was amplified with EcoTaq DNA polymerase (Ecogen). PCR products were purified with the Wizard PCR Preps DNA Purification kit (Promega). The amplified region was ligated into the pGEM-T easy vector (Promega) and cloned into *Escherichia coli* XL1-Blue or JM109 competent cells according to the manufacturer's protocol. Plasmid DNA was prepared for sequencing using the Wizard Plus Minipreps DNA Purification System kit (Promega). For each species, one clone was sequenced by Sanger's dideoxynucleotide chain-termination method for denatured double-stranded plasmid DNA. Sequencing was carried out automatically using the *A.L.F. express* DNA Sequencer (Pharmacia; Unitat de Microbiologia of the Universitat Autònoma de Barcelona), except for primer RT3 sequences, which were obtained with the AB1377 sequencer (Perkin Elmer Centre d'Investigació i Desenvolupament, CSIC, Barcelona). Compressions and ambiguities were resolved by multiple sequencing of both strands. Details on the PCR conditions and primers are given in Tarrío, Rodríguez-Trelles, and Ayala (1998). For sequencing, we used, in addition to the standard *M13*/*pUC* sequencing oligonucleotides, the following primers: *LT2,* 5′-ACGGCGARCT STWTCTGG-3′; *LT3,* 5′-CCATCGARCACRAGTCC-3′; *LT4,* 5′-GTGYTGGAYGTGATGGCAG-3′; *RT2,* 5′-TGCCATRCCRTTYAGATC-3′; *RT3,* 5′-GCAAT-CRTYGAAGCAGCG-3′; *RT5,* 5′-GATAGAARTGCY CCTGGC-3′. Sequences were aligned with CLUSTAL W, version 1.5 (Thompson, Higgins, and Gibson 1994). The *Xdh* sequences of this study were deposited in GenBank under accession numbers AF226950–AF226975.

In addition to *Xdh,* we analyzed published sequences from the mitochondrial *cytochrome oxidases I* (*CO I*) (419 bp), *II* (*CO II*) (687 bp), and *III* (*CO III*) (423 bp) (Spicer 1995). The sequences from these three regions included all the species of this study except *D. koepferae, D. uniseta,* and *D. mulleri.*

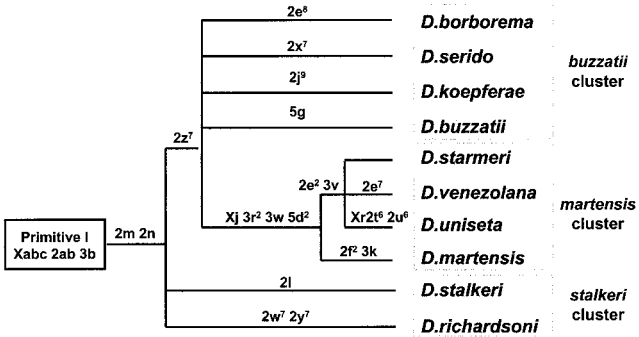

FIG. 1.—Phylogeny of the *buzzatii* complex based on homozygous paracentric inversions of chromosome 2. Redrawn from Ruiz and Wasserman (1993).

Statistical Analyses

In order to control possible errors in model fitting caused by imperfect prior knowledge of the phylogeny, we considered two tree topologies (figs. 1 and 2). Figure 2 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) using the default options. Figure 1 represents the relationships proposed by Ruiz and Wasserman (1993) on the basis of chromosomal inversion data.

Substitution models considered in this study are all special forms of the general time-reversible (REV) Markov process model (Tavaré 1986; Yang 1994). The REV model is sufficiently general for accurate estimation of the substitution pattern from the actual data (Yang, Lauder, and Lin 1995). Less complex models might, however, be desirable, because they yield estimates with smaller variances (see Kumar 1996).

Among-sites rate variation was accommodated by the models following two approaches (Yang 1996*b*): (1) allowing different rates for codon positions (referred to as C models; codon position–specific rates are considered as parameters [denoted c_1, c_2, and 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 c_1) and (2) treating rate differences among sites as a random effect using the discrete gamma distribution (eight equal-probability categories of rates, represented by the mean) with shape parameter α (denoted as dG models). The value of α is inversely related to the extent of rate variation (Yang 1996*a*). As nucleotide frequencies at equilibrium, we used the ML estimates which, for a few models examined, yielded consistently higher likelihood scores. Analyses were conducted with the BASEML program of PAML, version 2.0g (Yang 1999), and the PAUP*, version 4.0b2 (Swofford 1999), package.

The relevance of specific parameters for describing the evolution of *Xdh* was evaluated by means of the likelihood ratio test (Yang 1996*b*; Huelsenbeck and

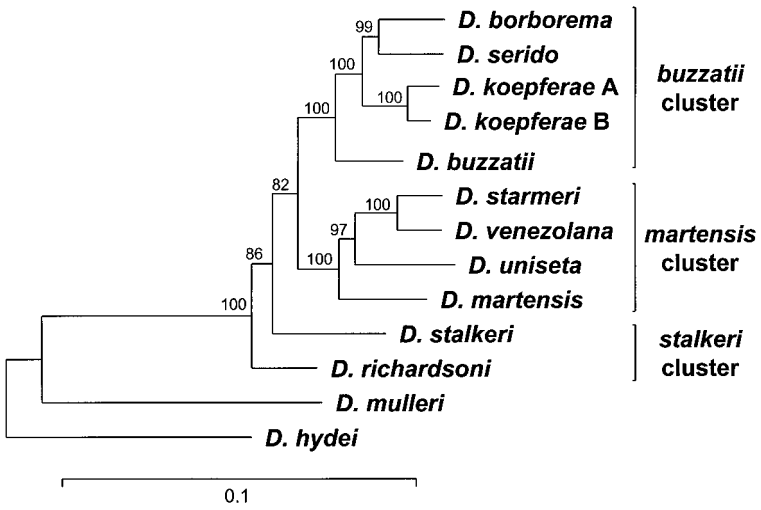

FIG. 2.—Unrooted maximum-likelihood tree of the *buzzatii* complex obtained from the *Xdh* nucleotide sequences with the REV + C model. Branch lengths are proportional to the scale, given in substitutions per nucleotide. Quartet puzzling support values are given on the nodes.

Crandall 1997). For a given tree topology (e.g., fig. 1), a model (H_1) with p parameters and 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 1996*b*). Yet, the likelihood differences of this study were all very large, so this inaccuracy of the χ^2 approximation is not expected to alter the conclusions of the tests.

Among-sites rate variation can be incorporated into the substitution models either before or after increasing the number of substitution types from one (i.e., the JC69 model) to six (the REV model). Varying the parameter addition sequence can affect best-fit model selection (Cunninghan, Zhu, and Hillis 1998). We took into account this potential source of bias by assaying different parameter addition sequences. Identified best models remained the same (results not shown).

The model found to satisfactorily describe the substitution process was used to generate candidate tree topologies by ML. Statistical support values for nodes of the ML trees were assessed by the quartet puzzling method (setting 1,000 puzzling steps). Support estimates for internal branches produced by this method can be interpreted as bootstrap scores (Strimmer and von Haeseler 1996). In addition, we used distance-based neighbor-joining (NJ) and weighted parsimony criteria to choose candidate trees. Estimates of α and transition/transversion bias used in distance computation and weighing 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 1996*a*). NJ trees were generated using the best-fit model identified by the likelihood-ratio test in the ML analysis. Parsimony analyses were conducted using the branch-and-bound algorithm with options MULTREES, furthest addition, MAXTREES = 100 (PAUP*, version 4.0b2; Swofford 1999). Statistical support for nodes of the NJ and maximum-parsimony trees was assessed with the bootstrap method (retaining nodes representing >50% from 1,000 bootstrap replications; Felsenstein 1985). In the case of maximum parsimony, bootstrap replicates were obtained with the heuristic method: MULTREES, simple stepwise addition sequence, MAXTREES = 100, and TBR branch swapping (PAUP*, version 4.0b2; Swofford 1999).

Phylogenetic hypotheses derived from the analyses of each gene region are compared by the resampling estimated log-likelihood (RELL) method of Kishino, Miyata, and Hasegawa (1990) (as it is implemented in PAML, version 2.0g; Yang 1999). For a given model of evolution, this test provides an estimate of the significance of a difference between the log-likelihood scores of several candidate tree topologies.

Results

Xdh Nucleotide Composition

Most models of molecular evolution rely on the premise of stationarity of base composition. Therefore, when nucleotide frequencies are heterogeneous, sequences of similar base compositions tend to become clustered regardless the evolutionary history of the organisms (Lockhart et al. 1994). We tested the stationarity assumption with the method of Rzhetsky and Nei (1995). This test (denoted *I*) takes into account possible phylogenetic correlations thereby is more suitable than alternative approaches, such as chi-square (Rzhetsky and Nei 1995). Because 20 tests were performed (for the first, second, third, first plus second, and all three codon positions, separately for the *buzzatii* complex, the *buzzatii* complex plus the outgroups one by one, and the entire data set) the significance level for each test was adjusted using a Bonferroni correction (Rice 1989). Ta-

Table 1
Nucleotide Composition of the *Xdh* Region

Species (cluster)	First Position				Second Position				Third Position			
	A	C	G	T	A	C	G	T	A	C	G	T
repleta group												
mulleri subgroup												
buzzatii complex												
Drosophila richardsoni (*stalkerii*)	24.2	23.3	36.5	16.0	31.7	22.4	17.8	28.1	7.1	39.3	34.5	19.1
Drosophila stalkeri (*stalkeri*)	25.2	22.9	36.0	16.0	31.4	22.2	18.0	28.5	7.8	37.3	34.0	21.0
Drosophila buzzatii (*buzzatii*)	24.6	23.7	35.3	16.4	31.4	22.7	18.0	27.9	7.5	39.1	34.8	18.6
Drosophila koepferae (*buzzatii*)	24.9	22.6	35.6	16.9	30.9	22.6	17.8	28.6	8.2	37.8	33.8	20.1
Drosophila borborema (*buzzatii*)	24.6	22.7	36.0	16.7	31.2	22.9	18.0	27.9	7.8	37.8	34.0	20.4
Drosophila serido (*buzzatii*)	24.9	23.2	35.3	16.7	31.1	23.5	17.7	27.8	8.3	38.3	33.8	19.6
Drosophila martensis (*martensis*)	24.3	23.3	35.7	16.7	31.5	22.2	18.1	28.2	7.2	37.4	35.3	20.1
Drosophila uniseta (*martensis*)	25.3	23.0	34.8	16.8	31.2	22.9	17.7	28.2	9.5	36.3	33.7	20.6
Drosophila venezolana (*martensis*)	25.0	23.0	35.8	16.1	31.2	23.0	17.7	28.1	8.1	36.7	34.5	20.7
Drosophila starmeri (*martensis*)	24.7	23.2	36.0	16.1	30.8	22.7	18.1	28.3	8.1	36.7	34.1	21.2
mulleri complex												
Drosophila mulleri	24.2	23.9	36.4	15.5	31.1	22.4	18.6	27.9	5.6	43.9	35.8	14.7
hydei subgroup												
Drosophila hydei	22.7	24.3	37.8	15.1	30.6	22.2	18.8	28.3	6.6	41.2	32.9	19.3
Average	24.6	23.2	35.9	16.3	31.2	22.7	18.0	28.2	7.7	38.4	34.2	19.7

Note.—Numbers are frequencies (%) of each nucleotide.

ble 1 shows the nucleotide frequencies and their averages for the *Xdh* region. Base composition was heterogeneous for the entire data set ($I = 71.7$, $P \approx 4 \times 10^{-4}$, df = 36) due to differences between *D. mulleri* (average GC = 0.60) and the remaining species (GC = 0.57). Differences are, however, small, and they are significant because of the long sequence length (2,085 bp). Within the *buzzatii* complex, nucleotide composition can be assumed to be stationary ($I = 43.24$, $P \approx 0.06$, df = 30).

Averaged over codon positions and species, nucleotide frequencies are close to ¼ proportions (A = 0.21, T = 0.22, C = 0.28, G = 0.29). Between codon positions, base frequencies are unequal. Averaged across all species, G and A are the most frequent nucleotides in first positions (G = 0.36 and A = 0.25); in second positions, G and C are the least frequent nucleotides (G = 0.18 and C = 0.23), and these two nucleotides are the most frequent in third positions (G = 0.34 and C = 0.38).

The Process of Nucleotide Substitution Along *Xdh*

Table 2 shows the log-likelihood ratio statistic values for models obtained assuming the topology shown in figure 2. Nested models were always rejected when compared against the next full model. The best description so far of the substitution process along the *Xdh* region was provided by the REV + C model, which allows six different substitution classes (two transitions, C↔T and A↔G, and four transversions, C↔A, C↔G, T↔A, and T↔G) and different substitution rates for codon positions.

Table 3 shows parameter estimates obtained with the REV + C (c_{1st}, c_{2nd}, c_{3rd}, and R) and the REV + dG (α) models under the topology shown in figure 2. Transition bias was not too strong, with transitions occurring less than four times ($R = 1.84$) as fast as expected if transitions and transversions occur at random. From the REV + C model, third codon positions change about seven times as fast as second and about four times as fast as first codon positions. The estimated gamma distribution for rates is L-shaped ($\alpha < 1$), indicating elevated among-site variation. Table 3 also gives the corresponding parameter estimates obtained under the topology shown in figure 1 (Ruiz and Wasserman 1993). Despite the large differences between the two topolo-

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	df	$-2 \log \Lambda$	P
Equal base frequencies	JC69	F81	3	27.20	$<10^{-5}$
Transition rate equals transversion rate	F81	HKY85	1	351.38	$<10^{-6}$
Equal transitional rates	HKY85	TN93	1	23.46	$\approx 10^{-6}$
Equal transversional rates	TN93	REV	3	17.54	$<10^{-3}$
Constant rate among sites	REV	REV + dG	1	288.94	$<10^{-6}$
Constant rate among codon positions	REV	REV + C	2	532.98	$<10^{-6}$

Note.—In each row, the null hypothesis (H_0) is compared with a hypothesis (H_1) that removes the assumption indicated in the left column. Log-likelihood values are obtained assuming the topology shown in figure 2. P represents the probability of obtaining the observed value of the likelihood ratio test statistic ($-2 \log \Lambda$) if H_0 is true, with degrees of freedom (df) indicated and $\alpha = 0.01$. JC69 = Jukes-Cantor 1969; F81 = Felsenstein 1981; HKY85 = Hasegawa-Kishino-Yano 1985; TN93 = Tamura-Nei 1993; REV = general time-reversible model, with REV + C assuming different rate parameters for codon positions and REV + dG assuming discrete gamma rates at sites.

Table 3.
Analysis of the *Xdh* Rate Variation Among Sites

Topology	c_{1st}	c_{2nd}	c_{3rd}	α	*R*
1	1	0.507	3.585	0.427	1.837
2	1	0.519	3.741	0.320	1.860

NOTE.—Topologies 1 and 2 are those shown in figures 2 and 1, respectively. Estimates were obtained with REV + C model for rates at codon positions and with REV + dG for parameters α and *R*.

gies, parameter estimates are very similar, which validates the use of figure 2 as a working topology.

CO I, II, and *III*

CO nucleotide frequencies were statistically homogeneous across taxa except for the third codon position of *CO III* when the ingroup sequences were combined with *D. hydei* ($I = 48.20$, $P < 10^{-4}$, df = 16). Typical of insect mitochondrial genes, base composition is very low in GC, with average frequencies over codon positions of T = 0.38, C = 0.19, A = 0.27, and G = 0.16 for *CO I*; T = 0.40, C = 0.14, A = 0.33, and G = 0.13 for *CO II*; and T = 0.41, C = 0.17, A = 0.30, and G = 0.12 for *CO III*.

As for *Xdh*, the most satisfactory description of the substitution process in the three mitochondrial regions was attained with the REV + C model (results not shown; ln $L = -1{,}289.09$, $-1{,}870.18$, and $-1{,}207.77$, for *CO I*, *CO II*, and *CO III*, respectively).

Figure 3 represents the rate variation among codon positions (first, second, and third) and regions (*Xdh*, *CO I*, *CO II*, and *CO III*) as inferred with the REV + C model (allowing 12 categories of rates) using the topology shown in figure 2. Estimates are scaled to the substitution rate in first codon positions of *Xdh*. *CO I* exhibits the greatest heterogeneity, with third codon positions evolving >200 times as fast as second positions (~7.31:0.03). First and second codon positions evolve the fastest in *Xdh*, followed by *CO III*. Even though rate differences among codon positions were largest for *CO I*, α values obtained separately for each region indicated that the overall among-sites rate variation was largest for *CO II* (α = 0.209 vs. α = 0.192 for *CO I* and *CO II*, respectively). This seemingly contradictory result reflects the different ways in which *c* and α interpret the rate variation among sites. *c* values represent substitution rates for the average site of a codon position, whereas α measures the extent of among-sites rate variation regardless the codon position of the site (Yang 1996*b*). A likelihood ratio test of the null hypothesis that the four regions fit the same gamma distribution indicated differences among genes in their extent of among-sites rate variation ($-2 \log \Lambda = 8.76$, df = 3, $P \approx 0.02$).

Analysis of the among-sites rate variation within third codon positions revealed that it was largest in *CO I* (α = 0.893), followed by *CO II* (0.972); third codon positions of *Xdh* and *CO III* exhibited greater values of α (1.039 and 2.497, respectively), therefore, they were expected to be more informative phylogenetically. Estimated transition/transversion ratios for the three mitochondrial genes were larger than those of *Xdh*, ranging between *R* = 2.25, for *CO I* and *R* = 3.29 for *CO III*.

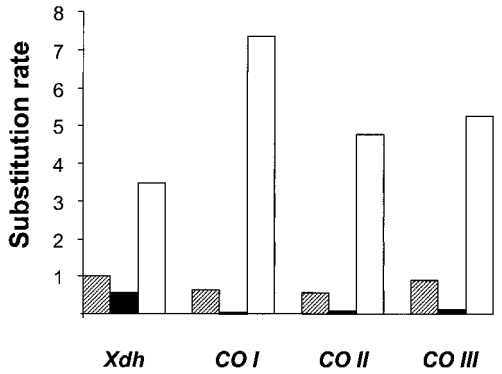

FIG. 3.—Estimated substitution rates in first (cross-hatched bars), second (black bars), and third (white bars) codon positions of the *Xdh*, *CO I*, *CO II*, and *CO III* regions, scaled to the rate of substitution in the first codon position of *Xdh*.

Phylogenetic Relationships in the *buzzatii* Complex

Figure 2 shows the ML tree obtained from the *Xdh* data set with the REV + C model (table 2). This is a fully resolved tree with well-supported nodes (quartet puzzling scores >80 in all cases), which coincides with the working topology also presented by this figure. Accordingly, the *stalkeri* cluster is the oldest lineage and is paraphyletic, with *D. richardsoni* derived before *D. stalkeri*. Next, the *martensis* and *buzzatii* clusters, which are monophyletic sister clades, split. Within the *martensis* cluster, *D. martensis* derived the earliest, followed by *D. uniseta* and *D. starmeri* and *D. venezolana*. Within the *buzzatii* cluster, *D. buzzatii* derived first, followed by *D. koepferae* and *D. borborema* and *D. serido*. These relationships are fully congruent with the chromosomal evolution of the complex as proposed by Ruiz and Wasserman (1993; fig. 1).

Figure 4*A* shows the NJ tree based on the REV model (assuming constant rates for sites; see table 2) and the complete *Xdh* data set. The NJ tree was topologically identical to the ML tree, with well-supported nodes (by the 70% or greater criterion of Hillis and Bull [1993]) except for the node determining the paraphyly of the *stalkeri* cluster (retained 454 times in the bootstrap test; not shown). The relatively low support for this node might reflect an inconsistency of the NJ algorithm that could be attributed to not taking into consideration the *Xdh* among-sites rate variation for the distance model. Using the REV + dG model with α set to 0.427 (see table 3) yielded a similar NJ topology (fig. 4*B*), but the bootstrap support for the paraphyly of the *stalkeri* cluster was strengthened. Analogously, unweighted maximum parsimony yielded two equally most-parsimonious trees 922 steps long that corresponded to the ML topology (fig. 2) and a topology clustering *D. martensis* and *D. uniseta* as a monophyletic group within the *martensis* cluster (fig. 4*C*). Setting the overall transition/transversion ratio *R* = 1.8 (see table 3; stepmatrix 9 by 5 in PAUP*), the single most-parsimonious

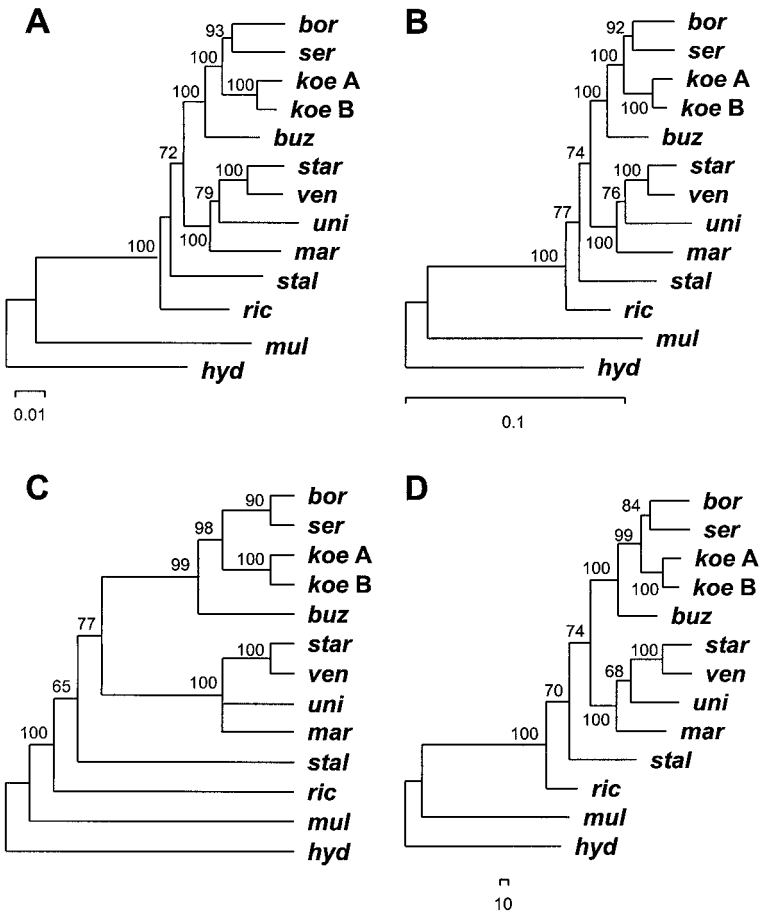

FIG. 4.—*A* and *B*, Neighbor-joining trees based on the REV distance, setting equal rates for sites (*A*) and allowing rate variation among sites ($\alpha = 0.427$; *B*). *C*, Strict consensus of the two equally most-parsimonious trees derived by unweighted parsimony. *D*, The single most-parsimonious tree obtained with a weighing 1 transition to 1.8 transversions. Branch lengths are proportional to the scale, given in substitutions per nucleotide. Percentage bootstrap values are given on the nodes.

tree is figure 4*D* (and fig. 2); bootstrap values for this tree are very similar to those for the NJ tree shown in figure 4*B*.

Figure 5 shows the trees obtained from the three cytochrome oxidase regions. Separately, the three mitochondrial genes support the monophyly of the *buzzatii* cluster, with *D. buzzatii* separated before *D. borborema* and *D. serido*, and place *D. starmeri* and *D. venezolana* as derived sister clades. These relationships are consistent across phylogenetic methods, with high nodal statistical support, and match the ones obtained for these species from *Xdh* (figs. 2 and 4). With regard to the remaining relationships, *CO III* favors the same topology as *Xdh*, whereas *CO I* places *D. martensis* in the base of the *buzzatii* complex, and *CO II* yields unstable topologies that are poorly supported statistically. As discussed above in connection with the substitution processes, *CO I* and *CO II* are the slowest evolving regions in first and second codon positions (fig. 3) and exhibit the third codon positions with the largest among-sites rate variation. Presumably, both circumstances reduce the phylogenetic utility of these genes. After combining the three mitochondrial regions, ML yields the same topology as the single most-parsimonious tree, which also matches the topologies obtained from *Xdh* and *CO III* (figs. 2, 4, and 5*G–I*).

Figure 6 combines all the information. The same topology is obtained with ML under the REV + C model, with the NJ algorithm based on the REV + dG distance (setting $\alpha = 0.300$) and with the weighted parsimony criterion (estimated $R = 1.40$; step-matrix 2 by 1 in PAUP*), which is similar to the topology shown in figure 2 but with fewer species. Quartet puzzling scores and bootstrap values clearly support the *stalkeri* cluster as the first derived and paraphyletic clade.

Figure 7*A–C* shows the results of Kishino, Miyata, and Hasegawa's (1990) RELL test for different phylogenetic hypotheses of interest about the *stalkeri* cluster: ''star-burst'' phylogeny (*A*), monophyly (*B*), and paraphyly, with *D. richardsoni* derived earliest (*C*). Tests were conducted using the REV + C model. Separately, all four of the regions support hypothesis *C* (a paraphyletic *stalkeri* cluster) as the most likely. Hypothesis *C* is statistically superior to hypothesis *A* for *Xdh* ($P \approx 0.01$) and *CO III* ($P \approx 0.05$), and to hypothesis *B* for *CO I* ($P \approx 0.04$) and *CO III* ($P \approx 0.07$). With all the evidence combinined, hypotheses *A* and *B* are clearly rejected ($P < 0.01$ and $P \approx 0.03$, respectively).

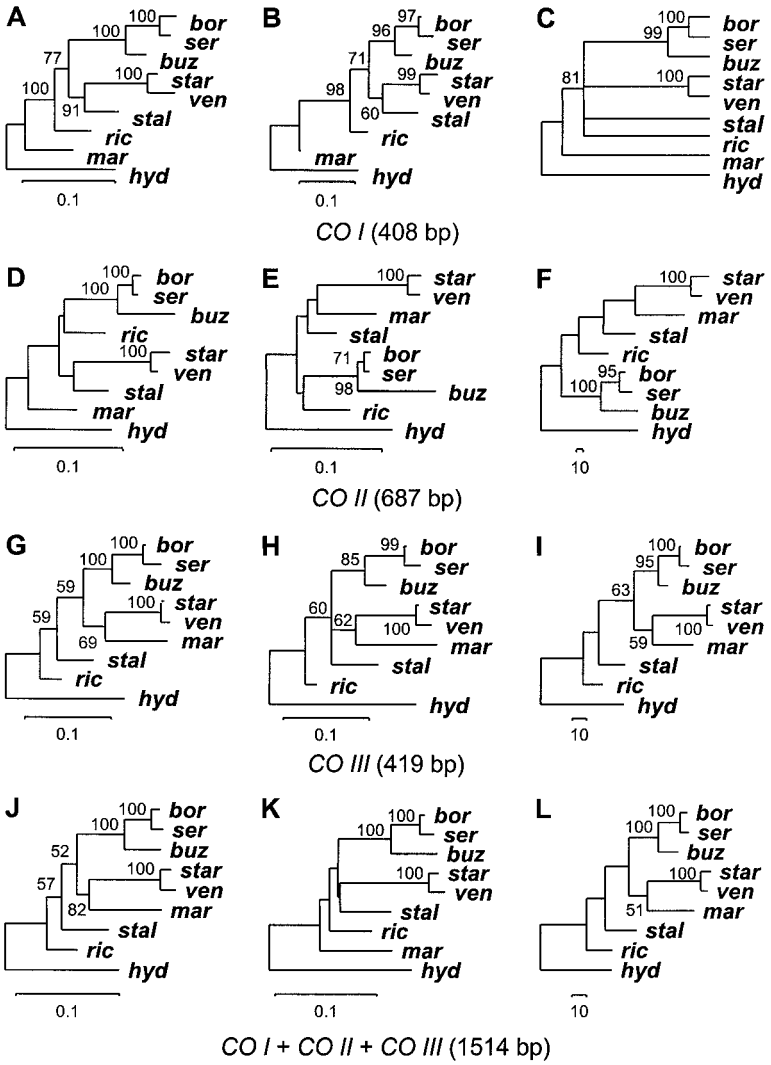

FIG. 5.—Maximum-likelihood (ML), distance, and parsimony trees for *CO I*, *II*, and *III* analyzed separately (*A–I*) and in combination (*J–L*). *A*, *D*, *G*, and *J* are ML trees obtained using the REV + C model; *B*, *E*, *H*, and *K* are neighbor-joining trees based on the REV distance, allowing rate variation among sites ($\alpha = 0.209$, 0.192, 0.279, and 0.210, respectively); *C*, *F*, *I*, and *L* are parsimony trees derived with transition : transversion weights of 1:2, 2:5, 1:3, and 2:5, respectively. The single most-parsimonious tree is presented except where several equally most-parsimonious trees were obtained (tree *C*), in which case the strict consensus is shown. Branch lengths are proportional to the scale, given in substitutions per nucleotide. Quartet puzzling (for the ML tree) and bootstrap values are given on the nodes.

Discussion

Because phylogenetic analyses are sensitive to the assumptions they make, we decided to adopt an ML framework. ML methods make mathematically explicit assumptions about the substitution process and provide a rationale for choosing among progressively sophisticated models of evolution through the likelihood ratio test (Yang 1994; Swofford et al. 1996; Huelsenbeck and Crandall 1997). Following this approach to assess the phylogenetic relationships of the *buzzatii* complex has led ML, distance, and parsimony criteria to converge to the same fully resolved tree.

All four of the gene regions analyzed in this study showed large substitution rate differences among sites. Furthermore, the degrees of among-sites rate variation were different for different genes. It is not circumstantial that the two regions exhibiting the greatest extent of among-sites rate variation, *CO I* and *CO II*, also produce disparate estimates of the phylogeny relative to what we consider the best estimate (see fig. 5). High levels of among-sites rate variation indicate that most substitutions occur in only a few positions, while the majority of sites never experience substitutions (Yang 1996*a*). Since neither sites with very few changes nor sites saturated with substitutions contain much phylogenetic information, it is not surprising that *CO I* and *CO II* yielded unreasonable topologies. It is noteworthy, however, that despite their different degrees of among-sites rate variation, combining the four genes under a common gamma distribution of rates resulted in a reinforced phylogenetic signal for the tree assumed by us to be the right tree. Apparently, there is phylogenetic signal in *CO I* and *CO II*. Presumably, this signal is present in slowly evolving sites, but it is overridden by some systematic

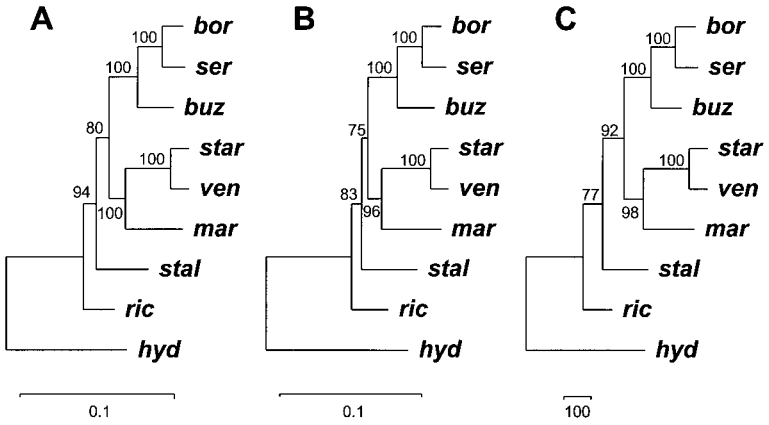

FIG. 6.—Maximum-likelihood (ML), distance, and parsimony trees for the *Xdh, CO I, CO II,* and *CO III* data sets pooled together. *A,* The ML tree obtained using the REV + C model, which allows different rate parameters for codon positions. *B,* The neighbor-joining tree based on the REV distance allowing rate variation among sites ($\alpha = 0.300$). *C,* The single most-parsimonious tree derived with a weighting one transitions to two transversions. Branch lengths are proportional to the scale, given in substitutions per nucleotide. Quartet puzzling (for the ML tree) and bootstrap values are given on the nodes.

bias at rapidly evolving sites, even though the among-sites rate variation seems to have been adequately accounted for by our best-fit model. In this respect, our conclusions strengthen the results from other studies (Barret, Donogue, and Sober 1991; Sullivan, Holsinger, and Simon 1995), indicating that phylogenetic signal can be additive when data from genes with different evolutionary processes are analyzed under a common reconstruction model.

That four independent molecular data sets, representing nuclear (*Xdh*) and mitochondrial (*CO I, CO II,* and *CO III*) protein-coding regions, converge to the same topology, which is furthermore fully consistent with the relationships inferred from chromosomal inversion data, supports the conclusion that the correct phylogeny of the *buzzatii* complex has been determined. Yet, our molecular analyses are more informative than the extensive cytological surveys conducted so far in the complex (Wasserman 1982; Ruiz and Wasserman 1993). Inversion data provide weak support for the basal position of the *stalkeri* cluster: it is the shortest of two nearly equal-length (one step difference) paths for the evolution of chromosome 2 (Ruiz and Wasserman 1993). The primitiveness of the *stalkeri* cluster is further attested to by the fact that its X, 3, 4, and 5 chromosomes seem to have remained unchanged relative to the hypothetical ancestral caryotype of the complex (Ruiz and Wasserman 1993). Chromosomal evolution also favors the monophyly of the *martensis* cluster, with *D. uniseta, D. venezolana,* and *D. starmeri* being more closely related to each other than to *D. martensis* (Ruiz and Wasserman 1993).

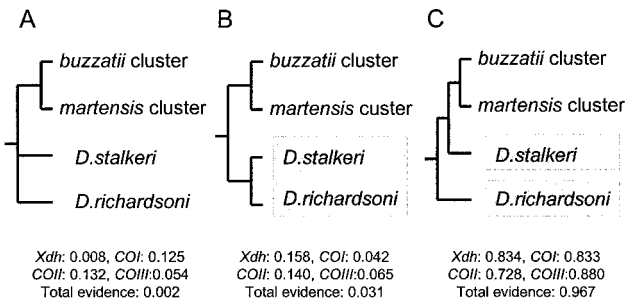

FIG. 7.—Hypotheses for the phylogenetic relationships of the *stalkeri* cluster with their corresponding RELL support values, obtained from each data set separately and in combination. RELL support values were obtained using the REV + C model.

The phylogeny of *buzzatii* has recently been addressed using data from three cytochrome oxidases (*CO I, CO II,* and *CO III*) (Spicer 1995). Neither separately nor combined do these sequences allow statistical discrimination of the branching order in the complex, apart from the identification of the *buzzatii* cluster and the closeness between *D. venezolana* and *D. starmeri* (Spicer 1995; results herein). Estimates of the degree of among-sites rate variation obtained by Spicer (1995) are larger than those obtained by us (Spicer's estimates vs. ours for *CO I, CO II,* and *CO III,* respectively: 0.346 vs. 0.209, 0.371 vs. 0.192, and 0.481 vs. 0.279). The discrepancy can be attributed to the different analytical methods used. Spicer's (1995) estimates of α differ from our ML estimates in that they are based on the number of changes at sites inferred by parsimony, which tends to overlook substitutions at the fastest-evolving sites, thus producing overestimates of α (Yang 1996*a*).

Our results challenge some current ideas on the evolution of the *buzzatii* complex. After combining the inversion data with information about the contemporary distribution of the species, Ruiz and Wasserman (1993) postulated that the ancestor of the complex inhabited the region now occupied by the *martensis* cluster (Venezuela and Colombia). From this primitive lineage, there was an early invasion of the Caribbean islands by migrants which evolved into the present *stalkeri* cluster. According to this scenario, the *stalkeri* cluster should be monophyletic. On the contrary, we infer that the *stalkeri* cluster is paraphyletic, which requires at least two independent colonizing events: one by migrants which evolved into the *D. richardsoni* lineage, followed by a second one that gave birth to *D. stalkeri.* Subsequently, the continental lineage stemmed into southern South America,

where it evolved into the present *buzzatii* cluster, whereas the northern South American area continued to change and is now the *martensis* cluster.

Until the work of Fontdevila et al. (1988), who identified it as a distinct species, *D. koepferae* had been erroneously considered as a geographical form of the morphologically polytypic *D. serido* (Ruiz, Fontdevila, and Wasserman 1982). In our study, the Argentinian (A) and Bolivian (B) strains of *D. koepferae* cluster in a monophyletic group clearly separated from *D. serido* (figs. 2 and 3). The *Xdh* sequence of *D. koepferae* A is more diverged from *D. serido* (0.0089; fig. 2) than the *Xdh* sequence of *D. koepferae* B (0.0065; fig. 2). Differences are nonsignificant ($P \approx 0.87$), but they occur in the same direction as for allozyme and interspecific crossability data (Fontdevila et al. 1988).

This study eliminates doubts about the use of *D. mulleri* as an outgroup for the phylogeny of *buzzatii*. Crossability experiments had revealed unsuspected levels of intercomplex mating between several species of the *buzzatii* and *mulleri* complexes (Ruiz and Wasserman 1993). Females of *D. mulleri* strains produced third-instar larvae when crossed with males of *D. buzzatii, D. martensis,* and *D. venezolana.* Analogously, males of *D. mulleri* produced first- or second-instar larvae when crossed to *D. borborema* females. Also, before the work of Ruiz and Wasserman (1993), the *buzzatii* and *martensis* clusters were classified within the *mulleri* complex. According to our analyses, however, *D. mulleri* falls well outside the *buzzatii* complex.

In order to examine the assumption that rates of substitution are constant along different parts of the tree, we carried out likelihood ratio tests of the molecular-clock hypothesis for *Xdh, CO I, CO II,* and *CO III* separately and for all four regions lumped together. Strictly speaking, this comparison is valid only if the likelihood values are calculated using the true topology. The topology shown in figure 2 is well supported by the analyses of this study and is consistent with the chromosomal phylogeny of the complex; therefore, it seems a reasonable hypothesis for tests of the molecular clock. We used the REV + C model to calculate the likelihood values either with or without the clock assumption. The REV + C clock hypothesis was rejected for each region taken separately ($-2 \log \Lambda = 43.22$, df = 12, $P < 10^{-4}$; $-2 \log \Lambda = 19.38$, df = 8, $P \approx 0.013$; $-2 \log \Lambda = 24.90$, df = 8, $P \approx 0.002$; and $-2 \log \Lambda = 34.76$, df = 8, $P < 10^{-4}$ for *Xdh, CO I, CO II,* and *CO III,* respectively) and for all four regions combined ($-2 \log \Lambda = 82.98$, df = 8, $P < 10^{-4}$). This result does not depend on the inclusion of *D. hydei* in the tests, because after removing this species, the outcome remained virtually unchanged. However, when only the species of the *martensis* and *buzzatii* clusters were considered in the analysis, all of the data sets can be assumed to fit the clock assumption ($-2 \log \Lambda = 6.70$, df = 8, $P \approx 0.57$; $-2 \log \Lambda = 4.06$, df = 5, $P \approx 0.54$; $-2 \log \Lambda = 5.24$, df = 5, $P \approx 0.39$; $-2 \log \Lambda = 11.46$, df = 5, $P \approx 0.04$; and $-2 \log \Lambda = 3.06$, df = 5, $P \approx 0.69$ for *Xdh, CO I, CO II, CO III,* and the combined data set, respectively). Hence, the deviation from the clock model detected above for the entire *buzzatii* complex must be caused by differences in the rate of evolution between the *stalkeri* cluster and the *martensis* and *buzzatii* clusters. It is apparent in figures 1 and 7 that the latter two clusters evolve faster than the former. Interestingly enough, from a cytogenetical standpoint, both the *martensis* and the *buzzatii* clusters have diverged more from the hypothetical ancestral caryotype than the *stalkeri* cluster (Ruiz and Wasserman 1993). This observation suggests the existence of a positive association between the rates of molecular and chromosomal evolution. Further analyses of the substitution rates by classes (e.g., synonymous vs. nonsynonymous, etc.) might help to clarify this question.

For a molecular marker to be useful for phylogenetic inference its amount of evolution should match the relevant divergence times. The *Xdh* region analyzed here had successfully been used before to resolve the phylogenetic relationships among the major subgroups of the *Drosophila saltans* species group (Rodríguez-Trelles, Tarrío, and Ayala 1999). Because it comprises intermediate-evolving first and second codon positions and fast-changing third codon positions with little rate variation from site to site, *Xdh* was highlighted as a suitable marker for studies seeking to resolve evolutionary relationships among recently derived taxa (i.e., within species groups or subgroups). The successful application of *Xdh* to the resolution of the evolutionary relationships among the closely related species of the *buzzatii* complex in this study strengthens this conclusion.

Acknowledgments

We thank G. Spicer for making available to us the *cytochrome oxidase* sequences used in this study. We are also grateful to E. Zouros for valuable suggestions, H. Cerdá, E. Hasson, and M. Ordoñez for help on collecting trips, and H. Laayouni and M. Peiró for technical assistance. This research has been supported by project PB96-1136 to A.F. from the Ministerio de Educación y Cultura (M.E.C.), Spain. L.A. was financed by a scholarship from the M.E.C., and F.R.-T. worked under a Contrato de Reincorporación funded by the M.E.C. This work was financed in part by Generalitat de Catalunya through a grant (1998 SGR-00050) to the Grup de Biologia Evolutiva.

LITERATURE CITED

BARKER, J. S. F. 1982. Population genetics of Opuntia breeding *Drosophila* in Australia. Pp. 209–224 *in* J. S. F. BARKER and W. T. STARMER, eds. Ecological genetics and evolution. Academic Press, Australia.

BARKER, J. S. F., F. M. SENE, P. D. EAST, and M. A. Q. R. PEREIRA. 1985. Allozyme and chromosomal polymorphisms of *Drosophila buzzatii* in Brazil and Argentina. Genetica **67**:161–170.

BARRET, M., M. J. DONOGHUE, and E. SOBER. 1991. Against consensus. Syst. Zool. **40**:486–493.

CUNNINGHAN, C. W., H. ZHU, and D. M. HILLIS. 1998. Best-fit maximum-likelihood models for phylogenetic inference: empirical tests with known phylogenies. Evolution **52**:978–987.

FELSENSTEIN, J. 1981. Evolutionary tress from DNA sequences: a maximum likelihood approach. J. Mol. Evol. **17**:368–376.

———. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39**:783–791.

———. 1993. PHYLIP, phylogenetic inference package and documentation. Version 3.5c. Distributed by the author, Department of Genetics, University of Washington, Seattle.

FONTDEVILA, A., C. PLA, E. HASSON, M. WASSERMAN, A. SÁNCHEZ, H. NAVEIRA, and A. RUIZ. 1988. *Drosophila koepferae*: a new member of the *Drosophila serido* (*Diptera*: *Drosophilidae*) superspecies taxon. Ann. Entomol. Soc. Am. **81**:380–385.

FONTDEVILA, A., A. RUIZ, J. OCAÑA, and G. ALONSO. 1982. Evolutionary history of *Drosophila buzzatii.* II. How much has chromosomal polymorphism changed in colonization? Evolution **36**:843–851.

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

HASEGAWA, M., H. KISHINO, and T. YANO. 1985. Dating the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. **22**:160–174.

HASSON, E., H. NAVEIRA, and A. FONTDEVILA. 1992. The breeding sites of Argentinian cactophilic species of the *Drosophila mulleri* complex (subgenus *Drosophila-repleta* group). Rev. Chilena Hitoria Nat. **65**:319–326.

HILLIS, D. M., and J. J. BULL. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. **42**:182–192.

HUELSENBECK, J. P., and K. A. CRANDALL. 1997. Phylogeny estimation and hypothesis testing using maximum likelihood. Ann. Rev. Ecol. Syst. **28**:437–466.

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

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., T. MIYATA, and M. HASEGAWA. 1990. Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. J. Mol. Evol. **31**:151–160.

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

LOCKHART, P. J., M. A. STEEL, M. D. HENDY, and D. PENNY. 1994. Recovering evolutionary trees under a more realistic model of sequence evolution. Mol. Biol. Evol. **11**:605–612.

MARÍN, I., A. RUIZ, C. PLA, and A. FONTDEVILA. 1993. Reproductive relationships among ten species of the *Drosophila repleta* group from South America and the West Indies. Evolution **47**:1616–1624.

PIÑOL, J., O. FRANCINO, A. FONTDEVILA, and O. CABRÉ. 1988. Rapid isolation of Drosophila high molecular weight DNA to obtain genomic libraries. Nuceic Acids Res. **16**:2736–2737.

RICE, W. R. 1989. Analyzing tables of statistical tests. Evolution **43**:223–225.

RITLAND, K., and M. T. CLEGG. 1987. Evolutionary analysis of plant DNA sequences. Am. Nat. **130**:S74–S100.

RODRÍGUEZ-TRELLES, F., R. TARRÍO, and F. J. AYALA. 1999. Molecular evolution and phylogeny of the *Drosophila saltans* species group inferred from the *Xdh* gene. Mol. Phylogenet. Evol. **13**:110–121.

RUIZ, A., and A. FONTDEVILA. 1981. Ecología y evolución del subgrupo *mulleri* de *Drosophila* en Venezuela y Colombia. Acta Cient. Venez. **32**:338–345.

RUIZ, A., A. FONTDEVILA, and M. WASSERMAN. 1982. The evolutionary history of *Drosophila buzzatii.* III. Cytogenetic relationships between two sibling species of the *buzzatii* cluster. Genetics **101**: 503–518.

RUIZ, A., and M. WASSERMAN. 1993. Evolutionary cytogenetics of the *Drosophila buzzatii* complex. Heredity **70**:582–596.

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

SANTOS, M., A. RUIZ, and A. FONTDEVILA. 1989. The evolutionary history of the *Drosophila buzzatii*. XIII. Random differentiation as a partial explanation of chromosomal variation in a structured natural population. Am. Nat. **133**:183–197.

SENE, F. M., M. A. Q. R. PEREIRA, and C. R. VILELA. 1982. Evolutionary aspects of cactus breeding *Drosophila* in South America. Pp. 97–106 *in* J. S. BARKER and W. T. STARMER, eds. Ecological genetics and evolution. The cactus-yeast *Drosophila* model system. Academic Press, Sydney, Australia.

SPICER, G. 1995. Phylogenetic utility of the mitochondrial *cytochrome oxidase* gene: molecular evolution of the *Drosophila buzzatii* species complex. J. Mol. Evol. **41**:749–759.

STRIMMER, K., and A. VON HAESELER. 1996. Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. Mol. Biol. Evol. **13**:964–969.

SULLIVAN, J., K. E. HOLSINGER, and C. SIMON. 1995. Among-site rate variation and phylogenetic analysis of the *12S rRNA* in sigmodontine rodents. Mol. Biol. Evol. **12**:988–1001.

SWOFFORD, D. L. 1999. PAUP*: phylogenetic analysis using parsimony (and other methods). Version 4.0b2. Sinauer, Sunderland, Mass.

SWOFFORD, D. L., G. J. OLSEN, P. J. WADDELL, and D. M. HILLIS. 1996. Phylogenetic inference. Pp. 407–514 *in* D. M. HILLIS, C. MORITZ, and B. K. MABLE, eds. Molecular systematics. Sinauer, Sunderland, Mass.

TAMURA, K., and M. NEI. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. **10**:512–526.

TARRÍO, R., F. RODRÍGUEZ-TRELLES, and F. J. AYALA. 1998. New *Drosophila* introns originate by duplication. Proc. Natl. Acad. Sci. USA **95**:1658–1662.

TAVARÉ, S. 1986. Some probabilistic and statistical problems on the analysis of DNA sequences. Pp. 57–86 *in* R. M. MUIR, ed. Lectures in mathematics in the life sciences. Vol. 17. American Mathematical Society, Providence, R.I.

THOMPSON, J. D., D. G. HIGGINS, and T. J. GIBSON. 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.

TIDO-SKLORZ, R., and F. M. SENE. 1995. *Drosophila seriema* n. sp.: new member of the *Drosophila serido* (*Diptera*: *Drosophilidae*) superspecies taxon. Ann. Entomol. Soc. Am. **88**:139–142.

VILELA, C. R. 1983. A revision of the *Drosophila repleta* species group (*Diptera*: *Drosophilidae*). Rev. Bras. Entomol. **27**:1–114.

WASSERMAN, M. 1982. Evolution of the *repleta* group. Pp. 61–139 *in* M. ASHBURNER, H. L. CARSON, and J. N. THOMPSON, eds. The genetics and biology of *Drosophila.* Vol. 3b. Academic Press, London.

WASSERMAN, M., and H. R. KOEPFER. 1979. Cytogenetics of South American *Drosophila mulleri* complex: the *martensis* cluster. More sharing of inversions. Genetics **93**:935–946.

WASSERMAN, M., and R. H. RICHARDSON. 1987. Evolution of the Brazilian *Drosophila mulleri* complex species. J. Hered. **78**:282–286.

YANG, Z. 1994. Estimating the pattern of nucleotide substitution. J. Mol. Evol. **39**:105–111.

———. 1996*a*. The among-site rate variation and its impact on phylogenetic analyses. TREE **11**:367–372.

———. 1996*b*. Maximum likelihood models for combined analyses of multiple sequence data. J. Mol. Evol. **42**:587–596.

———. 1999. PAML: phylogenetic analysis by maximum likelihood. Version 2.0g. Distributed by the author, Department of Biology, Galton Laboratory, University College London.

YANG, Z., N. GOLDMAN, and N. E. FRIDAY. 1994. Comparison of models for nucleotide substitution used in maximum likelihood phylogenetic estimation. Mol. Biol. Evol. **11**: 316–324.

YANG, Z., I. J. LAUDER, and H. J. LIN. 1995. Molecular evolution of the hepatitis B virus genome. J. Mol. Evol. **41**: 587–596.

ELEFTHERIOS ZOUROS, reviewing editor

Accepted March 31, 2000