Available online at www.sciencedirect.com
SCIENCE DIRECT®
Molecular Phylogenetics and Evolution 36 (2005) 695–705
MOLECULAR PHYLOGENETICS AND EVOLUTION
www.elsevier.com/locate/ympev

Phylogeny and molecular evolution of the *Drosophila hydei* subgroup (*Drosophila repleta* group) inferred from the *Xanthine dehydrogenase* gene

Tomás Morán, Antonio Fontdevila *

Grup de Biología Evolutiva, Departament de Genètica i de Microbiologia, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain

Received 15 September 2004; revised 18 March 2005
Available online 2 June 2005

Abstract

The *hydei* subgroup (*Drosophila repleta* group) consists of seven species divided into two complexes: *bifurca* and *hydei*, whose phylogenetic relationships are not well understood. To evaluate the molecular phylogeny of this subgroup, we analyzed 2085 bp of coding sequence of the *Xanthine dehydrogenase* gene in six available species of the *hydei* subgroup, with *Drosophila buzzatii* and *Drosophila mulleri* as an outgroup. For phylogenetic reconstruction we adopted a maximum-likelihood framework, based on the adjustment of descriptive models of nucleotide substitution to real data. We employed distance-based and weighted parsimony methods to construct candidate phylogenies. In all cases, we obtained only one completely resolved tree with strong statistical support for each node, that shows a phylogeny that is partially discordant with the proposed systematics of the subgroup. This tree suggests that the two species complexes are paraphyletic, as opposed to classic phylogenies using morphologic and cytologic traits. This discordance is discussed in relation to its implication for the evolutionary history of the *hydei* subgroup.

Keywords: *Drosophila*, *hydei* subgroup; *Xanthine dehydrogenase*; Phylogeny; Molecular evolution; Maximum likelihood; Molecular clock

1. Introduction

The *Drosophila repleta* group, subgenus *Drosophila*, is formed by almost a 100 species, most of them belonging to six subgroups: *mulleri*, *hydei*, *mercatorum*, *repleta*, *fasciola*, and *inca* (Rafael and Arcos, 1988, 1989; Vilela, 1983; Wasserman, 1992). Most species of this group are distributed in the arid and semiarid zones of the New World, due to their alimentary habits strongly associated with some Cactaceae (Wasserman, 1982). Throckmorton (1975) suggested that the radiation of this group occurred in North America around 30 million years ago. This time of origin is supported by Russo et al. (1995), who proposed a divergence time of the group of approximately 32 million years ago based on their analysis of the *Adh* gene in this group.

The *hydei* subgroup was defined by Wharton (1944) based on morphological characteristics and on results from interspecific crosses. The first morphological diagnosis of the subgroup was carried out by Wheeler (1949), but the present subgroup definition was proposed by Wasserman (1962, 1982, 1992) and Vilela (1983), using chromosomal and morphological criteria, respectively. In general, it is known that these species inhabit arid zones that range from southern North America (Texas, USA) to western South America (Junín, Perú; except *D. hydei* which has a cosmopolitan distribution) at altitudes ranging from 0 to over 3000 m (Vilela, 1983).

The *hydei* subgroup shows a well-defined phylogenetic origin. Chromosomally, it seems to be closely related to the *meridiana* complex of the *mulleri* subgroup

* Corresponding author. Fax: +34 93 5812726.
E-mail address: Antonio.Fontdevila@uab.es (A. Fontdevila).

1055-7903/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2005.04.009

(Wasserman, 1992). On the other hand, at the molecular level, the position of the *hydei* subgroup in the *repleta* group is not totally clear; but the studies of Russo et al. (1995) and Durando et al. (2000) suggest that the *hydei* subgroup may be closely related to the *eremophila* complex of the *mulleri* subgroup.

At present, the *hydei* subgroup consists of seven species divided into two complexes; the *bifurca* complex consisting of four species: *D. bifurca*, *D. nigrohydei*, *D. novemaristata*, and *D. guayllabambae*; and the *hydei* complex comprising three species: *D. eohydei*, *D. neohydei*, and *D. hydei*. The classification into two complexes was proposed by Wasserman (1962) based in the rearrangement patterns of polytene chromosomes and reproductive morphologies. Thus, the 2z inversion is fixed in the *hydei* complex, which shows a specialized spermathecae; while species of the *bifurca* complex show primitive spermathecae and conserve the standard chromosome Primitive Sequence I (PSI: Xabc 2ab 3b 4 5).

In their classical study, Dobzhansky and Sturtevant (1938) suggested that the study of gene arrangements in polytene chromosomes is useful for uncovering the phylogenetic relationships among species. The *hydei* subgroup exhibits a clear paucity of gene rearrangements, until 1992 only one fixed inversion difference had been reported at the interspecific level (Wasserman, 1962, 1992). At that time no chromosomal data on *D. novemaristata* and *D. guayllabambae* had been reported, because these species were only known by their description and type locality. Recently, preliminary studies with these two species indicated that their metaphasic chromosomes are identical and are similar to ancestral forms observed in species of the *bifurca* complex (Morán and Mafla, 2000). In addition, their polytene chromosomes are homosequential, differing from PSI by at least two fixed inversions in chromosome 2, designated provisionally as $2a^{10}$ and $2b^{10}$ (Morán, unpublished data). These inversions confirm a close phylogenetic relationship between both species, but they do not clearly resolve their relationships with the remaining species of the subgroup. The best resolved hypothetical chromosomal phylogeny of the *hydei* subgroup is shown in Fig. 1A.

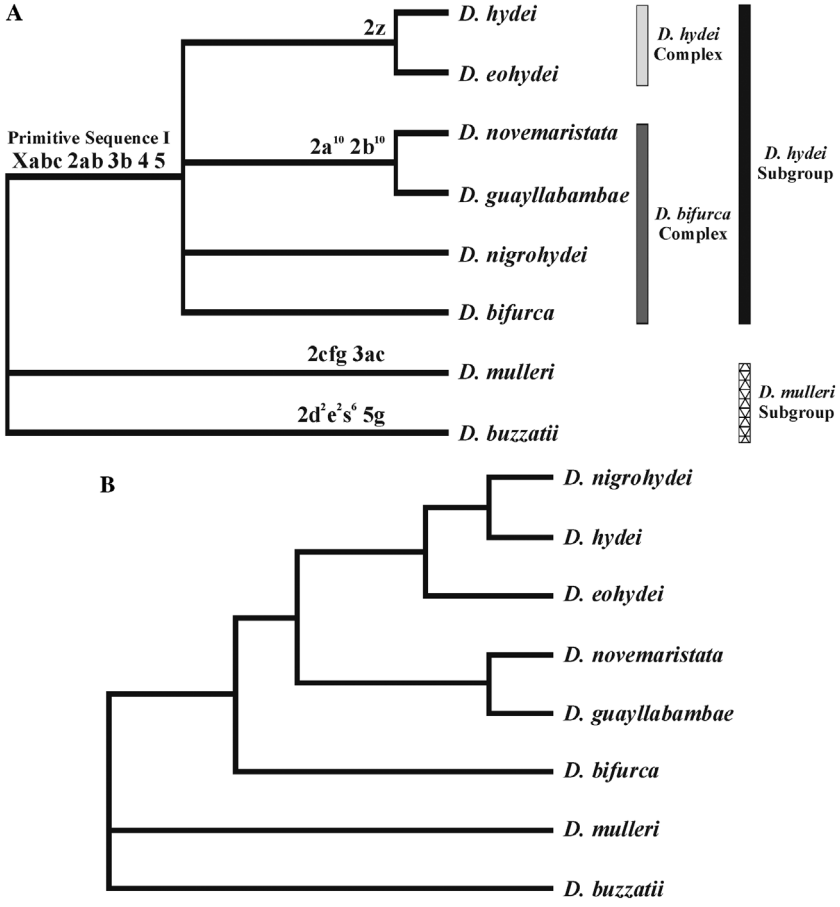

Fig. 1. Initial topologies of the *hydei* subgroup phylogeny. (A) Topology T_2, based on homozygous paracentric inversions (constructed from Spicer and Pitnick, 1996; Wasserman, 1992; Morán, unpublished data on inversions $2a^{10}$ and $2b^{10}$); the vertical bars correspond to the subgroup and complex assignment for all species. (B) Topology T_1, consensus of three different trees constructed by ML, maximum parsimony and neighbor joining distances.

Interspecific crosses in the subgroup do not shed more light on phylogenetic relationships either, and only indicate close relationships between species. In the *hydei* complex, *D. hydei* and *D. neohydei* are capable of producing hybrids in both directions and they have been tested up to the F_2 (Gregg, 1978; Schäfer, 1976; Wasserman, 1962, 1992). In the *bifurca* complex, *D. novemaristata* and *D. guayllabambae* also crossbreed in both directions and their fertility has also been demonstrated up to the F_2 (B. Larrea and V. Rafael, personal communication). The only hybridization case between species of distinct complexes occurs between *D. nigrohydei* males and *D. eohydei* females, tested up to the F_2 (the reciprocal cross does not produce offspring) (Wasserman, 1962, 1992) and, under very enforced conditions, crosses between *D. hydei* males and *D. nigrohydei* females produce a few progeny (two viable crosses out of 40 tested), yielding only sterile females (Bock, 1984; Patterson and Stone, 1952). Inter-complex hybridization raises some doubts about the complex authenticity as a monophyletic unit, and reveals the scarce knowledge about the species phylogenetic relationships. This contradiction between crossability and classic taxonomy is not uncommon among species of the *Drosophila repleta* group (see Marín et al., 1993).

Spicer and Pitnick (1996), employing mitochondrial DNA markers, showed concordance between chromosomal data and molecular analysis in the *hydei* subgroup, supporting the monophyletic origin of the *hydei* complex with strong statistical significance; but providing little support for the *bifurca* complex. Other studies on the molecular phylogeny of the *repleta* group, in which at least two species of the *hydei* subgroup are included, confirm that this subgroup is a monophyletic unit (Durando et al., 2000; Russo et al., 1995; Tatarenkov and Ayala, 2001).

This study is an attempt to elucidate the present questions regarding the evolutionary relationships among the species of the *hydei* subgroup, using molecular data and including species that have never been studied at the molecular level, such as *D. novemaristata* and *D. guayllabambae*. We analyzed 2085 coding nucleotides from the *Xanthine dehydrogenase* (*Xdh*) nuclear gene, using a maximum-likelihood (ML) approach for phylogenetic inference (e.g., Kumar, 1996; Ritland and Clegg, 1987; Rodríguez-Trelles et al., 1999; Yang et al., 1995). Previous reports have shown that *Xdh* gene is a good marker for phylogenetic studies among *Drosophila* species that diverged recently (Rodríguez-Trelles et al., 1999, 2000). Some characteristics that support these studies include the suitable rate of *Xdh* change to obtain a sufficient phylogenetic signal, its large inter- and intraspecific variability and its low functional constraints (Comeron and Aguadé, 1996; Riley, 1989; Riley et al., 1992; Rodríguez-Trelles et al., 1999, 2000).

2. Materials and methods

2.1. Drosophila stocks

All the species of the *hydei* subgroup were studied, except *D. neohydei* for which there are no available stocks. As an outgroup we used *D. buzzatii* (*buzzatii* complex) and *D. mulleri* (*mulleri* complex), because they belong to the *mulleri* subgroup, a sister taxon of the *hydei* subgroup (Russo et al., 1995; Tatarenkov and Ayala, 2001). The sequences of *D. hydei*, *D. buzzatii*, and *D. mulleri* were previously obtained by Rodríguez-Trelles et al. (2000). Accession numbers, stock references, and the geographical origin of these species are shown in Table 1.

2.2. Molecular methods

Genomic DNA was extracted either from 0.2 g of live flies as described by Piñol et al. (1988), or from alcohol preserved flies (Latorre et al., 1986), and kept at −20 °C. Polymerase chain reactions (PCR) were performed to amplify about 52% of all *Xdh* codons (2085 bp), including about half of exon II (1113 bp) and almost all of exon III (972 bp), following conditions and primers in Tarrío et al. (1998). PCR products were purified with GENECLEAN Kit (Q-BIO gene, *BIO101 systems*), ligated into the pGEM-T Easy Vector (Promega) and cloned into *Escherichia coli* DH5α competent cells, following the manufacturer's methods. Plasmid DNA for

Table 1
Drosophila species examined in this study

Taxon	Source locality[a]	GenBank Accession Nos.
D. hydei subgroup		
D. hydei complex		
D. hydei	Indiana, Bloomington Stock Center	AF226974 AF226975
D. eohydei	Bucaramanga—Colombia (15085–1631.1)	AY757914 AY757915
D. bifurca complex		
D. bifurca	Metztitlán—Mexico (15085–1621.0)	AY757912 AY757913
D. nigrohydei	Guayllabamba—Ecuador	AY757918 AY757919
D. novemaristata	Queminján—Perú	AY757910 AY757911
D. guayllabambae	Loja—Ecuador	AY757916 AY757917
D. mulleri subgroup		
D. buzzatii complex		
D. buzzatii	Berna—Argentina	AF226958 AF226959
D. mulleri complex		
D. mulleri	Pánuco—Mexico (15081–1371.1)	AF226972 AF226973

[a] Numbers in parenthesis refer to Tucson Stock Center numbers.

sequencing was isolated using the QIAprep Spin Miniprep Kit (QIAGEN). Reactions for sequencing were performed for only one clone of each species by the Sanger's dideoxynucleotide chain-termination method, using the ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems). To assure accuracy of sequences, the *Xdh* fragment was sequenced on both strands for each clone, or by multiple sequencing of the same clone. The sequencing primers (additional to the standard pGEM-T Easy Vector sequencing oligonucleotides) were: *LTM2*, 5′-GTGATYGTGACCATTG AGCAGG-3′; *LTM3*, 5′-ACCAATCTGCCCTCSA ATAC-3′; *RTM2*, 5′-CGAMARCAGCACCGATCCA T-3′; *RTM3*, 5′-CCGGCRTTGTTGTAGCACTC-3′. Sequence editing and alignment were performed with BioEdit, version 5.0.9 (Hall, 1999).

2.3. Phylogenetic inference

A maximum-likelihood (ML) framework was adopted for phylogenetic inference. This methodology makes use of all the information contained in sequences, incorporates concrete mathematical models to explain evolutionary patterns, allows statistical tests to evaluate different hypotheses and has been shown to be robust to violation of restrictions imposed by different models (Felsenstein, 1981; Page and Holmes, 1998; Yang, 1994). The likelihood for each candidate model is computed using a hypothetical initial topology. The likelihood ratio test is applied to determine which model fits to the molecular data best. The most satisfactory description of the substitution process and its derived parameters is used to make phylogenetic inferences by means of different algorithms of tree construction based on ML, maximum parsimony and genetic distance methods.

To detect important deviations in parameter estimation, without knowing the true *hydei* subgroup phylogeny, we started our analyses with two initial hypothetical topologies. The first topology (named T_1, Fig. 1B) corresponds to the strict consensus tree of topologies generated when applying the computer programs DNAml, DNAdis, and DNApars from PHYLIP package version 3.6a2 (Felsenstein, 2001), with default options. The second topology (named T_2, Fig. 1A) shows the species relationships based on chromosomal rearrangements.

All the candidate substitution models are special forms of the general time-reversible Markov process model (GTR) (Tavaré, 1986), also known as REV (Yang, 1994). The simpler nested models result from the successive inclusion of restrictions to the GTR model; and because these models usually have smaller variances, they are preferred over more general models if the adjustment to real data is optimal (Kumar, 1996).

Most of these models share some basic premises, whose violation can carry serious consequences. Thus, for example, it is known that when the nucleotide frequencies are heterogeneous, sequences tend to group by their nucleotide composition and the ensuing tree does not resemble the true evolutionary history of taxa (Lockhart et al., 1994; Tarrío et al., 2001). The *I* statistic was used in order to test the stationarity assumption, according to the method of Rzhetsky and Nei (1995). This test contemplates possible phylogenetic correlations and therefore is more trustworthy than other approaches like Chi-square. The level of significance for each test was adjusted by Bonferroni correction, fixing error of type I to 5×10^{-4} (Rice, 1989).

To estimate among-site rate variation, we followed two methodological approaches (Yang, 1994, 1996a,b): the first includes the random effect of a variable with discrete gamma distribution (Γ models), where substitution rates are fractioned in eight equally probable categories, and whose shape depends on parameter α (α is inversely related to the magnitude of substitution rate variation from site to site); the second one employs specific substitution rates for each codon position (C models), namely *c1*, *c2*, and *c3*, for first, second, and third codon position, respectively (the *c1* value is fixed to 1, so *c2* and *c3* are rates relative to first position). The analyses and parameter estimation by ML were made with the BASEML program of PAML, version 3.13a—2002 (Yang, 1997).

The likelihood values computed for each model were compared in a deviation analysis by the hierarchical likelihood ratio test (Huelsenbeck and Crandall, 1997; Yang, 1996b). Under this test, for a given topology a model (H_1), with p parameters and log-likelihood L_1, describes the observed 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 zone of a χ^2 distribution with n degrees of freedom (Yang, 1996b).

Variation in among-site substitution rates can be incorporated into the models before or after increasing the number of substitution types (from one in JC69 to six in GTR). Taking into account the effect of parameter inclusion order in model selection (Cunninghan et al., 1998), we evaluated different parameter addition sequences. In all cases the best model was always the same and therefore we only show one of the tested sequences. In our case, when the same model showed a good fit after the incorporation of different among-site rate variation parameters, the model selected was the one demonstrating the smallest *P* values.

We carried out three different phylogenetic tree algorithms: ML, neighbor-joining distances, and weighted parsimony, using the options implemented in PAML version 3.13a—2002 (Yang, 1997) and PAUP* version 4.0b2 (Swofford, 1999). In general, searches were made with the branch and bound (in PAUP*) and star-decomposition (semi-automatic option in PAML) algorithms. In the case of topologies constructed with PAUP*, the node statistical support was evaluated by the quartet

puzzling method (Strimmer and von Haeseler, 1996), fixing the option to 1000 puzzling steps.

An exhaustive search was used for distance methods (under the minimum evolution criterion) and also for weighted parsimony, including the options MulTrees, furthest addition and MaxTrees = 100 (PAUP*, version 4.0b2, Swofford, 1999). The support for nodes was evaluated by the bootstrap method, retaining nodes represented in more than 50% of the generated trees from 1000 bootstrap resamplings (heuristic method with the options of MulTrees, simple stepwise addition sequence, MaxTrees = 100 and TBR branch swapping; PAUP*, version 4.0b2, Swofford, 1999).

3. Results

3.1. Nucleotide composition

The average nucleotide frequencies across species deviate minimally from the expected equilibrium of 1/4 (A = 20.5%, C = 28.8%, G = 29.8%, and T = 20.9%), due to a small increase in the GC composition (58.6%). Yet nucleotide frequencies at each codon position deviate greatly from expected. For first codon positions the most frequent nucleotides are G and C (G = 37.4%, C = 23.9%), for second positions are A and T (A = 30.9%, T = 28.2%) and for third positions are C and G (C = 40.3%, G = 33.5%). Nucleotide frequencies are very similar among species ($\chi^2 = 10.807839$, df = 21, $P = 0.96639951$). Nevertheless, this comparison does not exclude the possibility that phylogenetic correlations in nucleotide composition may exist.

Table 2 shows stationary statistic values (I) for different taxon groupings and codon positions. The significant I values indicate that the evolutionary change cannot be considered homogeneous for all species and positions ($I = 71.69$, $P = 1.878 \times 10^{-7}$), nor for the *hydei* subgroup alone ($I = 56.43$, $P = 1.025 \times 10^{-6}$). For third positions, the homogeneity premise is also rejected ($I = 79.03$, $P = 1.172 \times 10^{-8}$), but not for first and second positions. Thus, our results could be accounted for by a high substitution rate of third codon positions and not necessarily by an overall heterogeneous change (Rzhetsky and Nei, 1995).

3.2. The ML phylogenetic reconstruction

We approached the problem of non-stationarity by using three models: (i) an homogeneous substitution process for the complete sequence, (ii) an homogeneous process for the first and second codon positions, and (iii) an heterogeneous process for the whole region.

The results of the likelihood ratio test are shown in Table 3. Our results indicate that the approximate descriptions that best represent the change process in the examined region are those contributed by models GTR C (first approach: when the complete sequence is considered and it is assumed that the process was stationary), HKY85 Γ (second approach: when only the first two codon positions are considered) and T92 Γ GC (third approach: when it is assumed that the change was heterogeneous). Therefore, an approximate change description of these sequences must take into account different substitution types (6 for GTR C and 2 for HKY85 Γ and T92 Γ GC), different substitution rates for each codon position, among-site rate variation, or difference in GC contents among lineages. The computed parameters derived from the three different models to accommodate among-site substitution rate variation (α, c_1, c_2, c_3, and R) are shown in Table 4.

All trees constructed from these models using the inferred parameters converged to a single highly supported tree (Figs. 2A and B), that is stable independent of the applied methodology (ML, distance or maximum parsimony). It is exactly the same as T_1 and supports the monophyletic origin of the *hydei* subgroup. Thus, *D. bifurca* represents the oldest lineage and its sister clade branches off to divide into two groups. One group gives rise to the ancestor from which *D. novemaristata* and *D. guayllabambae* diverged and the other is comprised of *D. eohydei*, which originated earlier, and a lineage that gave rise to *D. hydei* and *D. nigrohydei* later on. If these phylogenetic relationships depict the true evolutionary history of the *hydei* subgroup, the taxonomic division of the group into the *hydei* and *bifurca* complexes remains absolutely artificial. Both complexes appear paraphyletic, reflecting a discordance between the molecular phylogeny and the proposed phylogenetic relationships based on morphology and chromosomal markers (Figs. 1A and 2).

Table 2
Stationarity statistic values (I) of nucleotide composition (Rzhetsky and Nei, 1995)

Taxa	Codon position					
	1st	2nd	3rd	1st and 2nd	All	df
Ingroup	17.04	6.19	55.57*	11.76	56.43*	15
Ingroup and *D. mulleri*	21.56	6.94	74.90*	15.73	62.83*	18
Ingroup and *D. buzzatii*	31.61	10.17	58.57*	26.59	62.16*	18
All species	32.42	10.42	79.03*	27.08	71.69*	21

Note. Ingroup, all species of *hydei* subgroup.
* $P < 5 \times 10^{-4}$.

Table 3
Likelihood ratio test results of hierarchical substitution model comparison

Assumptions	H_0	H_1	df	Topology T_1		Topology T_2	
				$-2\log\Lambda$	P	$-2\log\Lambda$	P
1. First approach: Homogeneous model and three codon positions							
Equal base frequencies	JC69	:F81	3	77.51	1.05×10^{-16}	72.93	1.01×10^{-15}
Equal transition and transversion rates	F81	:HKY85	1	247.08	1.13×10^{-55}	278.66	1.47×10^{-62}
Equal transition rates	HKY85	:TN93	1	28.34	1.02×10^{-07}	22.59	2.00×10^{-06}
Equal transversion rates	TN93	:GTR	3	17.00	0.71×10^{-04}	21.22	9.46×10^{-05}
Uniform rate among sites	GTR	:GTR Γ	1	127.11	1.76×10^{-29}	329.98	9.70×10^{-74}
Uniform rate among codon positions	GTR	:GTR C	2	439.15	4.37×10^{-96}	495.93	2.00×10^{-108}
2. Second approach: Homogeneous model and the two first codon positions							
Equal base frequencies	JC69	:F81	3	13.22	4.18×10^{-03}	12.25	6.58×10^{-03}
Equal transition and transversion rates	F81	:HKY85	1	23.68	1.14×10^{-06}	24.61	7.01×10^{-07}
Equal transition rates	HKY85	:TN93	1	1.63	0.20093^{ns}	2.29	0.13006^{ns}
Uniform rate among sites	HKY85	:HKY85 Γ	1	32.45	1.22×10^{-08}	111.10	5.63×10^{-26}
Uniform rate among codon positions	HKY85	:HKY85 C	1	30.83	2.82×10^{-08}	29.46	5.70×10^{-08}
3. Third approach: Heterogeneous model and three codon positions							
Equal transition and transversion rates	JC69	:K80	1	234.73	5.52×10^{-53}	264.96	1.42×10^{-59}
GC & AT content in equilibrium	K80	:T92	1	91.48	1.13×10^{-21}	87.98	6.60×10^{-21}
Uniform rate among sites	T92	:T92 Γ	1	130.00	4.10×10^{-30}	341.50	3.01×10^{-76}
Equal nucleotide composition among branches	T92	:T92 GC	13	58.93	8.15×10^{-08}	53.36	7.82×10^{-07}
Equal nucleotide composition among branches	T92 Γ	:T92 Γ GC	13	62.40	1.95×10^{-08}	47.16	9.08×10^{-06}

Note. In each row, the null hypothesis (H_0) is compared with the hypothesis (H_1) that removes the assumption indicated in the first column. P indicates the probability of obtaining the observed value of the Likelihood Ratio Test statistic ($-2\log\Lambda$) if H_0 is true, with indicated degrees of freedom (df). JC69, Jukes and Cantor (1969); F81, Felsenstein (1981); K80, Kimura (1980); HKY85, Hasegawa et al. (1985); TN93, Tamura and Nei (1993); T92, Tamura (1992); GTR, general time reversible model; with Γ, assuming discrete gamma rates at sites; with C, assuming different rate parameters for codon positions; with GC, assuming nucleotide composition variation among tree branches.

Table 4
Parameters derived from the analysis of among site rate variation in *Xdh*

Topology	α	c_{1st}	c_{2nd}	c_{3rd}	R
1. First approach. Models GTR Γand GTR C					
T_1	0.39699	1	0.38905	3.80409	1.9502
T_2	0.22622	1	0.41480	3.98652	1.9629
2. Second approach. Models HKY85 Γand HKY85 C					
T_1	0.26269	1	0.42222	—	2.20374
T_2	0.10900	1	0.45122	—	2.27621
3. Third approach. Model T92 ΓGC					
T_1	0.37820	—	—	—	4.51652
T_2	0.21369	—	—	—	5.15068

Note. Topologies T_1 and T_2 are shown in Figs. 1B and A, respectively. The models to which all values refer are specified for each approach (models Γ, C or Γ GC). See text for symbol significance.

Using the first and second approaches, the trees obtained from matrix distances or by maximum parsimony (with or without adjusting for among-site rate heterogeneity with gamma distribution, or weighting transversions 2 times the value of transitions, $R=1.95$) are identical to the ML trees and show high statistical support at each node (bootstrap values > 65%). The clade with the smallest statistical support is the one grouping *D. eohydei* with the *D. nigrohydei* and *D. hydei* clade, indicating that many variable positions may be shared by these species.

Fig. 2B shows the rooted ML tree obtained with the T92 Γ GC model (quartet puzzling values are not shown due to limitations in the computer application). This topology is identical to trees obtained with the homogenous models. Since its implementation requires consideration of the ancestral GC content, and therefore lacks the time-reversible characteristic, likelihood value computation depends on root position (for this reason the initial topologies T_1 and T_2 have been rooted). Interestingly, rooting shows that substitution rates have not remained constant for all the clades, since the *hydei* subgroup apparently has accumulated less change than the *mulleri* subgroup. This approach confirms that the phylogenetic signal of *Xdh* region is sufficiently high to rescue only one topology, regardless of the models and the clustering methodologies.

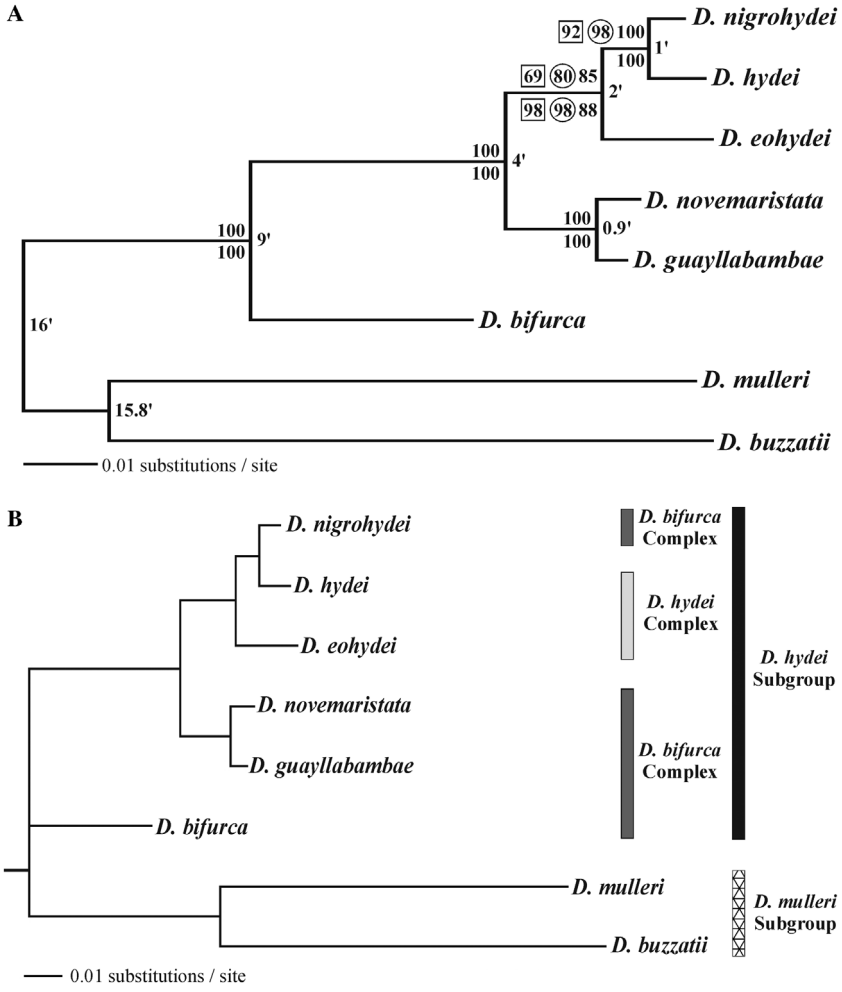

Fig. 2. *D. hydei* subgroup phylogeny. (A) Unrooted ML tree obtained from all *Xdh* sequences. Quartet puzzling support values are given at nodes. Bootstrap values obtained from trees constructed by maximum parsimony and neighbor joining distances are shown in rectangles and circles respectively, only when these differ from the quartet puzzling values. The statistical support of nodes shown under the branches correspond to the tree constructed with the GTR C model, and the values above the branches are from the tree obtained from the first two codon positions of the *Xdh* sequence with the HKY85 Γ model. The branch lengths correspond to the GTR C tree. Inferred divergence times (in million years) are indicated to the right of each clade. (B) Rooted ML tree of the *hydei* subgroup obtained from all the *Xdh* sequence with the T92 Γ GC model; the vertical bars correspond to the subgroup and complex assignment for all species.

3.3. Molecular evolution of the Xdh gene

Although all models converge to a single topology, the ML-derived parameters allowed us to understand *Xdh* evolution. In the first approach, the GTR C model gives a good description of the substitution process. Computed parameters to accommodate among-site substitution rate variation (Table 4, Section 1) show that the gamma distribution displays a "L" like shape ($\alpha = 0.39699$, $\alpha < 1$), signifying that among-site rate variation is large and therefore only a small fraction of sites are changing fast. The transition/transversion ratio ($R = 1.9502$) indicates that the deviation between these substitution types occur about four times faster than random expectation ($R = 0.5$). Substitution rates for codon positions point out that third positions change approximately four times faster than first positions, and nearly 10 times faster than second positions ($c_1{:}c_2{:}c_3 = 1{:}0.39{:}3.80$). This suggests that the rejection of stationarity could be due to the high substitution rate of the third codon positions (see Section 3.1).

When the nucleotide substitution pattern was characterized for only the first and second codon positions (second approach, 1390 pb in total), the best description of the process is provided by the HKY85 Γ model. The parameters employed to accommodate these variations (Table 4, Section 2) show that the gamma distribution maintains its "L" like shape ($\alpha = 0.26269$, $\alpha < 1$), demonstrating that among-site rate variation is large even when third codon positions are excluded. The transition/transversion ratio ($R = 2.20374$) and specific codon position substitution rates ($c_1{:}c_2 = 1{:}0.42$) remain similar to the predicted values when the complete sequence was considered.

When we choose the ML implementation of the Not—Homogenous model of Galtier and Gouy (1998), (third approach, T92 Γ GC model) the estimated parameters (Table 4, Section 3) show that the gamma distribution maintains its "L" shape ($\alpha = 0.3782$, $\alpha < 1$), but the transition/transversion ratio ($R = 4.51652$) is greater than those in the other two approaches, indicating that transitions happen nine times as fast as random expectation ($R = 0.5$). The values for both initial topologies are very similar in all three different approaches, which validates the use of T_1 as a working hypothesis.

4. Discussion

4.1. Model selection and parameter estimations

Since in our work the violation of the stationarity assumption implies that among-sequences nucleotide composition differences could affect the phylogenetic reconstruction (see Section 2.3), we decided to overcome this problem by employing three different models of the nucleotide substitution process: two of which modelling homogeneous processes that differ in the included codon positions, and one being a heterogeneous model that assumes differences in the among-sequence nucleotide composition. We preferred to work with different models because it allowed us to know whether the obtained phylogenetic reconstruction was affected by nucleotide composition or not. The use of a single model would have disregarded important information supplied by the specific substitution parameters (derived of substitution processes). Our results showed that the stationarity assumption violation in the *Xdh* sequence could be due to differences in site-specific substitution rates and to sequence length (2085 pb), and not to a heterogeneous nucleotide substitution. In general, parameter estimations demonstrated that among-site rate variation is large ($\alpha < 1$), third codon positions have a high substitution rate and transitions happen between four to nine times as fast as random expectation. The inference similarities among the three different model concerning parameter computation and tree reconstruction, gives us confidence that the inferred phylogeny used all the *Xdh* phylogenetic signal correctly.

4.2. Molecular vs morphological and chromosomal phylogenies

All models generate the same phylogenetic reconstruction (Figs. 2A and B), supporting the monophyletic origin of the *hydei* subgroup. This congruence resolves present doubts on the taxonomic positioning of some species like *D. novemaristata* and *D. guayllabambae*, since their subgroup assignment was not based on derived characters (Spicer and Pitnick, 1996). Interestingly, the resulting molecular phylogeny disagrees in part with present relationships based on species morphology and chromosomal rearrangements (Wasserman, 1962, 1992; Vilela, 1983), questioning the subgroup division into the *hydei* and *bifurca* complexes, since both appear as paraphyletic groups. This is not surprising in the *bifurca* complex, because no character has ever been reported to robustly support its monophyletic origin. In particular, the low morphological and chromosomal differentiation of the *bifurca* complex species only indicates that these lineages are the most primitive within the subgroup. However, the clustering of *D. nigrohydei* (*bifurca* complex) and *D. hydei* (*hydei* complex) as sister species is difficult to interpret in view of their chromosomal and morphological constitution. The monophyletic status of the *hydei* complex was based upon the presence of the fixed paracentric inversion 2z, assuming this is a unique derived character. This contradicts the inclusion of *D. nigrohydei*, closely related to *D. hydei*, into the *hydei* complex, since *D. nigrohydei* shows primitive morphologic characteristics (similar to the other species of the *bifurca* complex) and the 2z inversion has never been found in any population of this species (Wasserman, 1962, 1992).

The ancestral morphology of *D. nigrohydei* may suggest that morphological evolution is independent of lineage differentiation. However, this affirmation seems to be incorrect, since *D. eohydei* and *D. hydei* share very similar characteristics in their reproductive morphology, most probably derived from a common ancestor and not by convergence. Likewise, it is unlikely that *D. nigrohydei* reverted toward ancestral forms of reproductive morphology, since the evolution of reproductive characters depends on complex gene interactions—under great selective pressures—that make homoplasic reversion very unlikely (Zeng et al., 2000).

As for chromosomal evolution, the inclusion of *D. nigrohydei* within the *hydei* complex is still more questionable. According to Wasserman (1992), one of the advantages of employing shared paracentric inversions to establish phylogenetic relationships is that these appear to be unique events. This is founded on the fact that the probability of parallel cytological evolution or convergence is minute, since it would require that the rate of origin and fixation of inversions that share exactly the same breakage points be relatively high, a fact that has never been proved. The tree could only be explained by a prolonged retention of polymorphism, where the 2z inversion would have been fixed independently in *D. hydei* and *D. eohydei*, and not in *D. nigrohydei*. This is contradictory, since the fixation of an inversion depends on adaptative value and chance, and therefore maintenance of polymorphism during a long span of time would require a strong selective pressure in favor of the polymorphic state. In accordance with Wasserman (1992), inversions can have a positive effect

on fitness, but generally do not maintain the polymorphism over prolonged evolutionary time.

4.3. Gene trees vs species trees in the hydei subgroup evolution

It is not possible to eliminate the possibility that topologies obtained in this study show the evolutionary story of a gene, and do not necessarily depict the true species evolutionary tree. According to Pamilo and Nei (1988), gene trees can differ from species trees, especially when moderate allelic polymorphism exists within a species and when divergence times are short. Ting et al. (2000) point out that discordance between molecular and species phylogenies may be due to the genome being a mosaic of fragments, each one with its own genealogy. Therefore, it is necessary to consider factors capable of obscuring the phylogenetic inference, such as the retention of an ancestral polymorphism or introgression by secondary contacts.

In our case, it is probable that the molecular grouping of *D. nigrohydei* within the *hydei* complex species is unveiling other phenomena occurring during the divergence process. In this sense, the crossability observed among the species of the *hydei* subgroup supports the phylogeny that we obtained in this work. Thus, our results are concordant with the capability of hybridization between *D. nigrohydei* males and *D. eohydei* females, and between *D. hydei* males and *D. nigrohydei* females, meaning that *D. nigrohydei* is more closely related to the *hydei* complex than it was previously thought (Wasserman, 1992). If these species have maintained their potential to interbreed up to this day, it is not difficult to believe that during the early stages of divergence, gene flow had been high and had allowed introgression from the *hydei* complex species into *D. nigrohydei*. Thus, the phylogenetic signal of *Xdh* in *D. nigrohydei* could be influenced by recombination events with genetic material from closely related species and/or by retention of ancestral polymorphisms. Spicer and Pitnick (1996) previously discussed the possibility that the *bifurca* complex does not have a monophyletic origin based on their reconstruction of the *hydei* subgroup phylogeny using mitochondrial markers. In that study, the parsimony analyses demonstrated that it was only necessary to increase the tree size by a few steps to form a group between *D. nigrohydei* and the clade composed by *D. hydei* and *D. eohydei*. This would agree with our results, suggesting introgression, and may indicate that *D. nigrohydei* diverged earlier. In spite of the questionable position of *D. nigrohydei*, the high quartet puzzling and bootstrap values observed in almost all nodes supports the phylogeny we obtained.

4.4. Evolutionary phylogeography

Vilela (1983) suggests that the *hydei* subgroup may have originated in an ancestral population of Western North America located in a Mexican transitional zone (Patterson and Stone, 1952, pp. 510–523). This scenario would imply that different migration events have influenced species evolution. The first species to diverge was *D. bifurca*, remaining a stable and independent lineage until the present, with low morphological and cytological differentiation. The sister lineage would have differentiated into two well-defined phyletic lines. One would give rise to the clade that contains *D. novemaristata* and *D. guayllabambae*, in an early migration event toward the Andean equatorial zone. In a second migration, the other lineage would have invaded Central America and the north of South America, giving rise to *D. eohydei*, *D. neohydei*, *D. hydei* and—maybe before them—*D. nigrohydei*. However, the current knowledge of the ecology and distribution of all these species is still very limited, and this historical view is only a working hypothesis.

4.5. Is the molecular clock ticking at a constant rate?

Another important aspect worth evaluating is whether these sequences fit into the molecular clock hypothesis. According to Takezaki et al. (1995), the nucleotide substitution rates are never identical for all lineages, although it is expected that in closely related species the heterogeneity could be almost imperceptible. We tested this hypothesis by means of the likelihood ratio test (Felsenstein, 1981), comparing the likelihoods of trees constrained by the molecular clock assumption against trees where this restriction is relaxed. The results showed that the molecular clock hypothesis is not accepted ($-2\log\Lambda = 17.62318$, 6 df, $P = 0.00725$) under the GTR C model for the complete sequence when all species were considered. However, when the species of the *mulleri* subgroup (outgroup) were excluded from analysis, the molecular clock fits reasonably to the data ($-2\log\Lambda = 7.80806$, 4 df, $P = 0.09887$). This may indicate that these outgroup species evolve faster than the *hydei* subgroup species (see Section 3.2 and Fig. 2B). On the other hand, when all species and only the first two codon positions were considered, and the HKY85 Γ model was implemented, the molecular clock shows a reasonably good fit to the data ($-2\log\Lambda = 10.70272$, 6 df, $P = 0.09801036$).

In principle, it is erroneous to calibrate the molecular clock to estimate divergence times when differences in the nucleotide substitution rates among lineages exist. However, even though substitution rates are uniform, if the nucleotide sites show little variation, i.e., they are highly constrained as in first and second positions of our study, an increasing deviation of divergence time estimates impedes a reliable calibration of the clock (Nei and Kumar, 2000). Considering that the substitution process has been constant for the *hydei* subgroup, we have obtained a rough approximation of its time of origin. As a reference point, we adopt the divergence time between

D. hydei and *D. mulleri*, estimated to be approximately 16 million years ago (inferred from the *Adh* gene, in Russo et al., 1995). In this way, and using computed lengths for each branch of the tree constructed under the molecular clock hypothesis, we found than the *hydei* subgroup originated about 9 million years ago. This estimate agrees with Spicer and Pitnick (1996), where the subgroup origin was dated between 9 and 21 million years.

The divergence times for each node of the tree are shown in Fig. 2A. The estimates we present here should be interpreted in light of the fact that the correct tree position for *D. nigrohydei* is uncertain.

4.6. Final remarks

Based on the results of this work, we assert that the division of the *hydei* subgroup into two complexes is not natural, since they are clearly paraphyletic. Based on this reasoning, and the fact that the *hydei* subgroup consists of a low number of species, one could suppress the complexes and maintain the subgroup as a unique monophyletic unit. Alternatively, one could maintain the present systematic classification on the basis of other available criteria (e.g., cytology, morphology, and crossability), and await future studies that will make it possible to conclusively clarify the phylogenetic relationships among all the species of the subgroup.

Acknowledgments

We are indebted to V. Rafael and D. Lachaise for providing us with *Drosophila* stocks. We want to thank F. Rodríguez-Trelles for his help with the phylogenetic reconstruction methodology, and M. Peiró for her technical assistance. We also thank Saraha Schaack for her helpful comments in the English style revision. This research has been supported by Grants BOS 2000-0295-C02 and BOS 2003-05904-C02 to A.F. from Ministerio de Ciencia y Tecnología (MCYT), Spain. T.M. was supported by a MUTIS—MAE fellowship from the Agencia Española de Cooperación Internacional (AECI) and the Ministerio de Asuntos Exteriores (MAE), Spain.

References

Bock, I.R., 1984. Interspecific hybridization in the genus *Drosophila*. Evol. Biol. 18, 41–70.

Comeron, J.M., Aguadé, M., 1996. Synonymous substitutions in the *Xdh* gene of *Drosophila*: heterogeneous distribution along the coding region. Genetics 144 (3), 1053–1062.

Cunninghan, C.W., Zhu, H., Hillis, D.M., 1998. Best-fit maximum-likelihood models for phylogenetic inference: empirical tests with known phylogenies. Evolution 52 (4), 978–987.

Dobzhansky, T.H., Sturtevant, A.H., 1938. Inversions in the chromosomes of *Drosophila pseudoobscura*. Genetics 23 (1), 28–64.

Durando, C.M., Baker, R.H., Etges, W.J., Heed, W.B., Wasserman, M., DeSalle, R., 2000. Phylogenetic analysis of the *repleta* species group of the genus *Drosophila* using multiple sources of characters. Mol. Phylogenet. Evol. 16 (2), 296–307.

Felsenstein, J., 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17 (6), 368–376.

Felsenstein, J., 2001. PHYLIP: Phylogeny Inference Package. Version 3.6a2 Distributed by the author, Department of Genetics, University of Washington, Seattle, USA.

Galtier, N., Gouy, M., 1998. Inferring pattern and process: maximum-likelihood implementation of nonhomogeneous model of DNA sequence evolution for phylogenetic analysis. Mol. Biol. Evol. 15 (7), 871–879.

Gregg, T.G., 1978. Crossing over and nondisjunction in interspecific hybrids between *Drosophila hydei* and *neohydei*. Genetics 88 (Suppl.), s33–s34.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41, 95–98.

Hasegawa, M., Kishino, H., Yano, T., 1985. Dating of the human–ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22 (2), 160–174.

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

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

Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16 (2), 111–120.

Kumar, S., 1996. Patterns of nucleotide substitution in mitochondrial protein coding genes of vertebrates. Genetics 143 (1), 537–548.

Latorre, A., Moya, A., Ayala, F.J., 1986. Evolution of mitochondrial DNA in *Drosophila subobscura*. Proc. Natl. Acad. Sci. USA 83 (22), 8649–8653.

Lockhart, P.J., Steel, M.A., Hendy, M.D., Penny, D., 1994. Recovering evolutionary trees under a more realistic model of sequence evolution. Mol. Biol. Evol. 11 (4), 605–612.

Marín, I., Ruiz, A., Pla, C., Fontdevila, A., 1993. Reproductive relationships among ten species of the *Drosophila repleta* group from South America and the West Indies. Evolution 47 (5), 1616–1624.

Morán, T., Mafla, A.B., 2000. Metaphase karyotypes of *Drosophila guayllabambae* and *Drosophila novemaristata* (*repleta* group, *hydei* subgroup). D.I.S. 83, 100–102.

Nei, M., Kumar, S., 2000. Molecular Evolution and Phylogenetics. Oxford University Press, New York, USA.

Page, R.D.M., Holmes, E.C., 1998. Molecular Evolution: A Phylogenetic Approach. Blackwell, United Kingdom.

Pamilo, P., Nei, M., 1988. Relationships between gene trees and species trees. Mol. Biol. Evol. 5 (5), 568–583.

Patterson, J.T., Stone, W.S., 1952. Evolution in the Genus *Drosophila*. Macmillan, New York.

Piñol, J., Francino, O., Fontdevila, A., Cabré, O., 1988. Rapid isolation of *Drosophila* high molecular weight DNA to obtain genomic libraries. Nucleic Acids Res. 16 (6), 2736.

Rafael, V., Arcos, G., 1988. *Drosophila guayllabambae* n.sp., un nuevo miembro del grupo *repleta*, subgrupo *hydei* (Diptera, Drosophilidae). Evol. Biol. 2, 167–176.

Rafael, V., Arcos, G., 1989. Subgrupo *inca*, un nuevo subgrupo del grupo *repleta*, con descripción de *Drosophila huancavilcae* n.sp (Díptera, Drosophilidae). Evol. Biol. 3, 233–243.

Rice, W.R., 1989. Analyzing tables of statistical tests. Evolution 43 (1), 223–225.

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 (1), 33–52.

Riley, M.A., Kaplan, S.R., Veuille, M., 1992. Nucleotide polymorphism at the *Xanthine dehydrogenase* locus in *Drosophila pseudoobscura*. Mol. Biol. Evol. 9 (1), 56–69.

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

Rodríguez-Trelles, F., Tarrío, R., Ayala, F.J., 1999. Molecular evolution and phylogeny of the *Drosophila saltans* species group inferred from the *Xdh* gene. Mol. Phylogenet. Evol. 13 (1), 110–121.

Rodríguez-Trelles, F., Alarcón, L., Fontdevila, A., 2000. Molecular evolution and phylogeny of the *buzzatii* complex (*Drosophila repleta* group): a maximum-likelihood approach. Mol. Biol. Evol. 17 (7), 1112–1122.

Russo, C.A.M., Takezaki, N., Nei, M., 1995. Molecular phylogeny and divergence times of Drosophilid species. Mol. Biol. Evol. 12 (3), 391–404.

Rzhetsky, A., Nei, M., 1995. Test of the applicability of several substitution models for DNA sequence data. Mol. Biol. Evol. 12 (1), 131–151.

Schäfer, U., 1976. Hybrid breakdown in *Drosophila hydei* X *D. neohydei* hybrids. Genetics 83 (Suppl.), s67–s68.

Spicer, G., Pitnick, S., 1996. Molecular systematics of the *Drosophila hydei* subgroup as inferred from mitochondrial DNA sequences. J. Mol. Evol. 43 (3), 281–286.

Strimmer, K., von Haeseler, A., 1996. Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. Mol. Biol. Evol. 13 (7), 964–969.

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

Takezaki, N., Rzhetsky, A., Nei, M., 1995. Phylogenetic test of the molecular clock and linearized trees. Mol. Biol. Evol. 12 (5), 823–833.

Tamura, K., 1992. Estimation of the number of nucleotide substitutions when there are strong transition–transversion and G + C-content biases. Mol. Biol. Evol. 9 (4), 678–687.

Tamura, K., Nei, M., 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and Chimpanzees. Mol. Biol. Evol. 10 (3), 512–526.

Tarrío, R., Rodríguez-Trelles, F., Ayala, F.J., 1998. New *Drosophila* introns originate by duplication. Proc. Natl. Acad. Sci. USA 95 (4), 1658–1662.

Tarrío, R., Rodríguez-Trelles, F., Ayala, F.J., 2001. Shared nucleotide composition biases among species and their impact on phylogenetic reconstructions of the Drosophilidae. Mol. Biol. Evol. 18 (8), 1464–1473.

Tatarenkov, A., Ayala, F.J., 2001. Phylogenetic relationships among species groups of the *virilis–repleta* radiation of *Drosophila*. Mol. Phylogenet. Evol. 21 (2), 327–331.

Tavaré, S., 1986. Some probabilistic and statistical problems on the analysis of DNA sequences. In: Muir, R.M. (Ed.), Lectures in Mathematics in the Life Sciences, vol. 17. American Mathematical Society, Providence, USA, pp. 57–86.

Throckmorton, L.H., 1975. The phylogeny, ecology and geography of *Drosophila*. In: King, R.C. (Ed.), Handbook of Genetics, vol. 3. Plenum Press, New York, USA, pp. 421–469.

Ting, C.T., Tsaur, S.C., Wu, C.I., 2000. The phylogeny of closely related species as revealed by the genealogy of a speciation gene, *Odysseus*. Proc. Natl. Acad. Sci. USA 97 (10), 5313–5316.

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

Wasserman, M., 1962. Cytological studies of the *repleta* group of the genus *Drosophila*. IV. The hydei subgroup. Univ. Texas Publ. 6205, 73–84.

Wasserman, M., 1982. Cytological evolution in the *Drosophila repleta* species group. In: Barker, J.S.F., Starmer, W.T. (Eds.), Ecological Genetics and Evolution. Academic Press, New York, USA, pp. 49–64.

Wasserman, M., 1992. Cytological evolution of the *Drosophila repleta* species group. In: Krimbas, C.B., Powell, J.R. (Eds.), *Drosophila* Inversion Polymorphism. CRC Press, Boca Raton, USA, pp. 455–552.

Wharton, L., 1944. Interspecific hybridization of the *repleta* group. Univ. Texas Pub. 4445, 175–193.

Wheeler, M.R., 1949. Taxonomic studies on the Drosophilidae. Univ. Texas Pub. 4920, 157–195.

Yang, Z., 1994. Estimating the pattern of nucleotide substitution. J. Mol. Evol. 39 (1), 105–111.

Yang, Z., Lauder, I.J., Lin, H.J., 1995. Molecular evolution of the hepatitis B virus genome. J. Mol. Evol. 41 (5), 587–596.

Yang, Z., 1996a. Among-site rate variation and its impact on phylogenetic analyses. Trends Ecol. Evol. 11 (9), 367–372.

Yang, Z., 1996b. Maximum-Likelihood models for combined analyses of multiple sequence data. J. Mol. Evol. 42 (5), 587–596.

Yang, Z., 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. 13(5), 555–556. Available from: <http://abacus.gene.ucl.ac.uk/software/paml.html>.

Zeng, Z.B., Liu, J., Stam, L.F., Kao, C.H., Mercer, J.M., Laurie, C.C., 2000. Genetic architecture of a morphological shape difference between two *Drosophila* species. Genetics 154 (1), 299–310.