Molecular phylogeny of the subgenus *Drosophila* (Diptera, Drosophilidae) with an emphasis on Neotropical species and groups: A nuclear versus mitochondrial gene approach

Lizandra J. Robe [a], Vera L.S. Valente [a], Myriam Budnik [b], Élgion L.S. Loreto [c,*]

[a] *Programa de Pós-Graduação em Genética e Biologia Molecular, Departamento de Genética, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Rio Grande do Sul, Brazil*

[b] *Instituto de Ciências Biomédicas, Faculdade de Medicina, Universidade do Chile, Santiago, Chile*

[c] *Departamento de Biologia, Universidade Federal de Santa Maria (UFSM), Camobi Km 09, CP5050, Santa Maria, Rio Grande do Sul 97105-900, Brazil*

Abstract

The genus *Drosophila* has played an essential role in many biological studies during the last 100 years but much controversy and many incompletely addressed issues still remain to be elucidated regarding the phylogeny of this genus. Because information on the Neotropical species contained in the subgenus *Drosophila* is particularly incomplete, with this taxonomic group being underrepresented in many studies, we designed a study to answer some evolutionary questions related to these species. We subjected at least 41 Drosophilidae taxa to a phylogenetic analysis using a 516-base pair (bp) fragment of the α-methyldopa (*Amd*) nuclear gene and a 672 bp fragment of the mitochondrial cytochrome oxidase subunit II (*COII*) gene both individually and in combination. We found that the subgenus *Drosophila* is paraphyletic and subdivided into two main clusters: the first containing species traditionally placed in the *virilis-repleta* radiation and the second assembling species of the *immigrans-Hirtodrosophila* radiation. Inside the first of these clusters we could detect the monophyly of both the *flavopilosa* (the sister-clade of the *annulimana* group) and the *mesophragmatica* (closely related to the *repleta* group) species groups. Concerning the *immigrans-Hirtodrosophila* lineage, *Zaprionus*, *Liodrosophila*, *Samoaia*, and *Hirtodrosophila* were the early offshoots, followed by the *immigrans*, *quinaria*, *testacea*, and *funebris* species groups. The *tripunctata* radiation appears to be a derived clade, composed of a paraphyletic *tripunctata* group, intimately interposed with members of the *cardini*, *guarani*, and *guaramunu* species groups. Overall, the *COII* gene yielded a poor phylogenetic performance when compared to the *Amd* gene, the evolutionary hypothesis of which agreed with the total evidence tree. This phenomenon can be explained by the fast saturation of transitional substitutions in *COII*, due to strong biases in both base composition and substitution patterns, as also by its great among-site rate variation heterogeneity.

Keywords: *Amd* gene; *cardini* group; *COII* gene; *Drosophila*; *flavopilosa* group; *guaramunu* group; *guarani* group; *mesophragmatica* group; Neotropical species; *tripunctata* group

1. Introduction

The family Drosophilidae (Diptera) includes about 3000 species distributed over more than 60 genera, with about half of the total number of species as yet described belonging to the genus *Drosophila* (Wheeler, 1982). According to this author, the genus *Drosophila* has a wide and scattered distribution, being composed of 15

* Corresponding author. Fax: +55 55 220 8020.

E-mail addresses: lizbiogen@pop.com.br (L.J. Robe), vera.valente@bol.com.br (V.L.S. Valente), mbudnik@med.uchile.cl (M. Budnik), elgion@base.ufsm.br (É.L.S. Loreto).

subgenera. Among these we distinguish the speciose subgenera *Drosophila* and *Sophophora*, whose systematics in most cases still remains controversial and incomplete.

The largest group of Drosophilidae inhabiting the Neotropics is the subgenus *Drosophila* (Val et al., 1981), where it includes at least 25 species groups. Malogolowkin (1953) and Throckmorton (1975) subdivided these species into two main lineages: the *virilis-repleta* section (or radiation) containing 15 species groups; and the *quinaria-tripunctata* section (also known as the *immigrans-Hirtodrosophila* radiation) containing 9 Neotropical species groups. Besides this, the subgenus *Drosophila* is also composed of the *funebris* group, basal to both radiations according to Throckmorton (1975). This author also supported the thesis that the subgenus *Drosophila* forms a paraphyletic clade from which a series of other genera and subgenera originated 'at different times and levels.' Recently, this hypothesis has been substantiated by several studies (Kwiatowski and Ayala, 1999; Remsen and DeSalle, 1998; Remsen and O'Grady, 2002; Russo et al., 1995; Tatarenkov et al., 1999, 2001; Thomas and Hunt, 1993) but it has been negated by others (DeSalle, 1992; Grimaldi, 1990; Powell and DeSalle, 1995). These uncertainties still present in *Drosophila* phylogenetic studies are compounded by the fact that Neotropical members of the subgenus *Drosophila* are poorly represented in many of these studies.

One of the subgenus *Drosophila* species groups that are underrepresented in phylogenetic studies is the *tripunctata* group, a member of the *quinaria-tripunctata* section of Malogolowkin (1953) and Throckmorton (1975). This group is composed of 64 species (Vilela and Bächli, 2000) but its internal phylogeny has rarely been tested, except by the morphological work of Frota-Pessoa (1954) and the recent molecular analysis carried out by Yotoko et al. (2003). According to Throckmorton (1975), the *tripunctata* species group constitutes a paraphyletic clade, situated in his '*tripunctata* radiation' along with other species groups such as *pallidipennis*, *rubrifrons*, *sticta*, *macroptera*, *calloptera*, *cardini*, and *guarani*. Kastritsis (1969) suggested that some members of the *guarani* group were phylogenetically closer to the *tripunctata* group than to other members of its own group, which has led to the subdivision of the former *guarani* species group into the *guarani* and *guaramunu* groups.

Conversely, the *virilis-repleta* section of the subgenus *Drosophila* has been more widely considered in molecular phylogenetic studies (Durando et al., 2000; Nurminsky et al., 1996; Tatarenkov and Ayala, 2001) but still contains a series of gaps concerning the relationships of some of its groups. For example, the Andean *mesophragmatica* species group has not as yet had its essentially conflicting inter-group relationships elucidated (Durando et al., 2000; Remsen and O'Grady, 2002; Tatarenkov and Ayala, 2001; Throckmorton, 1975). Besides that, its internal subdivision has never been studied using a molecular approach, the only phylogenetic studies to date having been based on morphological (Nacrur, 1958), cytogenetic (Brncic et al., 1971) or isozyme variations (Nair et al., 1971). Another example is the fact that even though Throckmorton (1975) questioned the exact derivation of the *flavopilosa* species group no subsequent study has been carried out to clarify this point.

Based on the considerations outlined above, this study was principally designed to elucidate the phylogenetic relationships within and between Neotropical species groups traditionally placed in the subgenus *Drosophila*, i.e., the *immigrans*, *tripunctata*, *cardini*, *guarani*, and *guaramunu* species groups included in the *quinaria-tripunctata* section and the *mesophragmatica* and *flavopilosa* species groups assembled in the *virilis-repleta* section. Besides that, we also wished to shed more light on some questions related to the monophyly of the genus and subgenus *Drosophila* and on the taxonomic status of some of their radiations. We also addressed some questions concerning the correct position of the taxa *Zaprionus*, *Samoaia*, *Liodrosophila*, *Hirtodrosophila*, and *Dorsilopha* within the family Drosophilidae. The molecular markers used in this study were the α-methyldopa (*Amd*) nuclear gene sequence and the mitochondrial cytochrome oxidase subunit II (*COII*) gene sequences, the partitions being used both separately and combined. These markers were used because they had previously proved to be informative at this level of divergence (Tatarenkov and Ayala, 2001; Tatarenkov et al., 2001; Yotoko et al., 2003). Combining a nuclear and a mitochondrial gene is also a good choice because these two types of data are unlinked and evolving under different evolutionary constraints (Lin and Danforth, 2004).

2. Materials and methods

2.1. Species

The Drosophilidae taxa used in this study (49 for *Amd*, 71 for *COII*, and 41 in the combined analysis) as well as their respective *Amd* and *COII* sequence GenBank accession numbers at National Center for Biotechnology Information (NCBI) site are given in Table 1, where the new sequences obtained in this study are underlined. The genus *Scaptodrosophila* was utilized as outgroup because there is ample evidence that this is the case with respect to the genus *Drosophila* (DeSalle, 1992; Grimaldi, 1990; Kwiatowski et al., 1994, 1997; Kwiatowski and Ayala, 1999; Pélandakis and Solignac, 1993; Russo et al., 1995; Tatarenkov et al., 1999, 2001; Throckmorton, 1975). Most of the focused groups have a Neotropical restricted distribution.

Table 1
List of *Drosophila* species and other Drosophilidae used in this study, with their GenBank accession numbers for the *Amd* and *COII* sequences

Genus	Subgenus	Section	Group	Species	Amd Acc. Nos.	COII Acc. Nos.
Drosophila	*Drosophila*	*quinaria-tripunctata*	*guarani*	***D. ornatifrons* 1**	AY699250	AY162977[h]
				D. ornatifrons 2	—	AY162978[h]
				D. subbadia	AY699251	AY847772
				D. guaru	—	AY847763
			guaramunu	***D. griseolineata***	AY699257	AF478424[f]
				D. maculifrons 1	—	AY847766
				D. maculifrons 2	—	AY162979[h]
				D. maculifrons 3	—	AY162980[h]
			tripunctata	***D. angustibucca***[A]	AY699248	AY162983[h]
				D. angustibucca 2[A]	AY699249	—
				D. mediosignata	—	AY162985[h]
				D. paraguayensis 1	—	AY162986[h]
				D. paraguayensis 2	—	AY162987[h]
				D. cuaso 1	—	AY162984[h]
				D. cuaso 2	—	AY170441[h]
				D. metzii	—	AY162992[h]
				***D. mediopunctata* 1**	AY699254	AY162988[h]
				***D. mediopunctata* 2**	AY699255	AY162989[h]
				D. crocina = *D. mediostriata* (Vilela and Bächli, 1990)	AY699252	AY847759
				D. mediostriata	AY699253	AY847767
				D. paramediostriata 1	—	AY162995[h]
				D. paramediostriata 2	—	AY162996[h]
				***D. bandeirantorum* 1**	AY699256	AY162990[h]
				D. bandeirantorum 2	—	AY162991[h]
				***D. tripunctata* 1**	AF293709[a]	AF478432[f]
				D. tripunctata 2	—	AF519343[g]
				D. mediopicta	—	AY847768
				D. medioimpressa	—	AY162994[h]
			cardini	*D. cardini* 1	—	AY162974[h]
				D. cardini 2	—	AF519319[g]
				D. cardinoides	AY699258	AY162975[h]
				D. neocardini	AY699260	AY847770
				D. polymorpha	AY699259	AY162976[h]
			pallidipennis	*D. pallidipennis* 1	—	AY162981[h]
				D. pallidipennis 2	—	AY162982[h]
			calloptera	*D. calloptera*	—	AF478419[f]
			testacea	***D. putrida* 1**	AF293723[a]	AF478431[f]
				D. putrida 2	—	AF519335[g]
			quinaria	*D. quinaria*	—	AF478428[f]
				D. phalerata	AF293721[a]	AF147115[c]
			immigrans	***D. immigrans* 1**	AY699261	AF478424[f]
				***D. immigrans* 2**	AF293713[a]	AY162993[h]
				D. immigrans 3	—	AF519324[g]
			funebris	***D. funebris***	AF293709[a]	AF478422[f]
		virilis-repleta	*mesophragmatica*	***D. gasici*AR**	AY699239	AY847760
				***D. gasici*CA**	AY699240	AY847761
				***D. gasici*CO**	AY699241	AY847762
				D. brncici	AY699238	AY847757
				D. mesophragmatica	AY699242	AY847769
				D. gaucha	AF324955[b]	AF478423[f]
				D. pavani	AY699245	AY847771
				D. viracochi	AY699244	AY847773
			repleta	***D. hydei***	AF293712[a]	AF478429[f]
				D. eohydei	—	AF145889[d]
				D. nigrohydei	—	AF145890[d]
				D. buzzatii	AF324947[b]	AF146169[b]
				D. borborema	—	AF146171[b]
				D. serido	—	AF146173[b]
				D. mercatorum	AF324957[b]	—

(*continued on next page*)

Table 1 (*continued*)

Genus	Subgenus	Section	Group	Species	Amd Acc. Nos.	COII Acc. Nos.
			dreyfusi	***D. camargoi***	AF324951[b]	AF478421[f]
			canalinea	*D. canalinea*	AF324952[b]	—
			flavopilosa	***D. cestri***	AY699246	AY847758
				D. incompta	AY699247	AY847764
			annulimana	***D. aracatacas***[B]	AF324949[b]	AY847756
			virilis	*D. virilis*	AF293729[a]	—
			robusta	*D. robusta*	AF293724[a]	—
			nannoptera	***D. nannoptera***	AF324959[b]	AF478425[f]
	Sophophora		*melanogaster*	***D. simulans* 1**	AY699262	AF474082[e]
				D. simulans 2	AF293726[a]	—
				D. erecta	AF293708[a]	—
				D. yakuba	—	X03240[a]
			obscura	*D. pseudoobscura*	AF293722[a]	—
	Dorsilopha			***D. busckii***	AF293707[a]	AF519347[g]
Hirtodrosophila				***H. pictiventris***	AF293711[a]	AF478434[f]
Liodrosophila				***L. aerea***	AF293715[a]	AF478435[f]
Samoaia				***S. leonensis***	AF293725[a]	AF478438[f]
Zaprionus				***Z. indianus***	AY699263	AY847774
				Z. tuberculatus	AF293731[a]	AF478440[f]
Scaptodrosophila				***S. latifasciaeformis***	AY699264	AY847765

(1) The classification proposed by Grimaldi (1990) was followed in this paper. (2) More than one *Drosophila* population was sampled for some of the species listed. This is illustrated by the three *D. gasici* lineages, where population AR was from Arica (Chile), population CA was from Cochabamba (Bolivia), and population CO was from Bogota (Colombia). For the other species the different populations are demarcated by different terminal numbers. (3) Only the species listed in boldface were included on the simultaneous analysis. (4) The sequences newly obtained in this study are underlined.

[A] Species recently redefined as *D. nappae* by Vilela et al. (2004).

[B] For *COII* the *annulimana* species group was represented by *D. annulimana*.

[a] *Amd* sequences from Tatarenkov et al. (2001) and *COII* sequences from Clary et al. (1982).

[b] *Amd* sequences from Tatarenkov and Ayala (2001) and *COII* sequences from Spicer (1995).

[c] *COII* sequences from Spicer and Jaenike (1996).

[d] *COII* sequences from Spicer and Pitnick (1996).

[e] *COII* sequences from O'Grady and Kidwell (2002).

[f] *COII* sequences from Remsen and O'Grady (2002).

[g] *COII* sequences from Perlam et al. (2003).

[h] *COII* sequences from Yotoko et al. (2003).

2.2. DNA manipulation

The total DNA of each of the isolineages established was extracted according to the protocol described by Jowett (1986), with some modifications. The degenerate *Amd-un2* and *Amd-bw* primers (Tatarenkov et al., 2001) were used to amplify a variable length fragment of the *Amd* gene (Fig. 1A) while the TL2-J-3037 and TK-N-3785 primers (Simon et al., 1994) were used to amplify a 684-base pair (bp) fragment of the *COII* gene (Fig. 1B). In each case the final components of the 50 µl reaction mixture were: 5 µl reaction buffer, 2.5 mM $MgCl_2$, 0.25 mM of each dNTP, 20 pmol of each primer, 1 U of *Taq* Polymerase (all from Invitrogen), and 100 ng DNA. The *Amd* amplification parameters varied according to the group under study, following local optimization criteria, while the *COII* amplification parameters were more or less constant: denaturation at 94 °C for 7 min, followed by 35 cycles of denaturation for 1 min at 94 °C, annealing for 1 min at 55 °C, and extension for 75 s at 72 °C, after which the reaction mixture was additionally held at 72 °C for 5 min to complete the extension. The *Amd* amplicons obtained

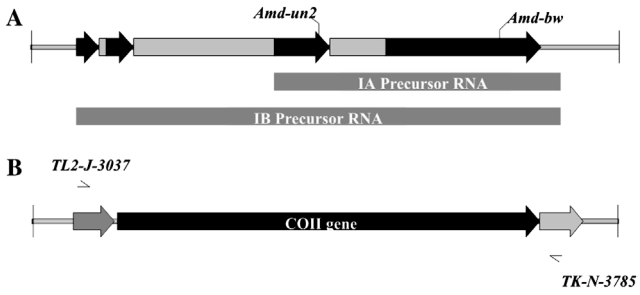

Fig. 1. Genomic maps for the *Amd* and *COII* genes. The arrows denote exons and the clear gray rectangles represent the introns. (A) The *Amd-un2* and *Amd-bw* primers were used to amplify a region of the *Amd* gene encompassing 92 bp (of 471 bp) from the third exon, 991 bp (of 1316 bp) from the fourth exon, and a variable length intron. This gene gives rise to two different precursor RNAs (indicated by the dark gray rectangles below its genomic region) which code for the two *Amd* isoforms. Note that for the B isoform the third exon is spliced off (Adams et al., 2000). (B) The TL2-J-3037 primer anneals to the tRNA-Leu coding region while the TK-N-3785 primer anneals to the tRNA-Lys coding regions (both indicated by the gray arrows), these two primers being used to amplify the entire 684 bp coding region of the *COII* gene.

were separated on 0.8% agarose gel and fragments of the expected size were excised and purified using the Qiaquick DNA-Gel Extraction System (Qiagen), according to the supplier's specifications. The more specific *COII* amplicons were directly purified by incubation at 37 °C for 30 min with exonuclease I and shrimp alkaline phosphatase (SAP) (both from Amersham) followed by a 15 min inactivation step at 80 °C. DNA sequencing was performed directly using the purified amplicons and a MegaBACE 500 automatic sequencer. The dideoxy chain-termination reaction was implemented using the DYEnamic ET (Amersham) kit and the *Amd-bw* primer for *Amd* and the TL2-J-3034 or TK-N-3785 primers for *COII*.

2.3. Phylogenetic analysis

The nucleotide sequences had their identity initially confirmed using the BlastN program available at the NCBI site. Inspection and correction of automated sequences were subsequently accomplished using the Staden Package Gap 4 program (Staden, 1996), wherefrom a confidence consensus sequence was obtained for each taxon. These sequences were aligned using the ClustalX 1.81 program (Jeanmougin et al., 1998) according to the system default parameters. This alignment was subjected to a saturation test using the DAMBE program (Xia and Xie, 2001) and to a permutation test probability (PTP) performed on PAUP 4.0b10 (Swofford, 2003).

The individual and combined phylogenetic analyses were executed using four main methods: (1) neighbor joining (NJ) (Saitou and Nei, 1987) using MEGA 2.1 (Kumar et al., 2001); (2) maximum parsimony (MP) using an heuristic search with 100 replicates of random addition of taxa performed using the equal weight criteria as contained in PAUP 4.0b10; (3) maximum likelihood (ML) using PAUP 4.0b10 following the model proposed by the Akaike information criterion (AIC) test (Akaike, 1974) executed using the ModelTest program (Posada and Crandall, 1998); (4) Bayesian analysis performed using the MrBayes 3.0b4 (Huelsenbeck and Ronquist, 2001) according to the model proposed by the ModelTest, with the evaluation of at least 1,000,000 generations and a 'burn-in' region of 2000 trees. The confidence values of each clade for the first three analyses were measured by the bootstrap test (Felsenstein, 1985) with 1000 replications (except for the ML analysis, where only 500 replications were considered). For the Bayesian analysis the posterior probability of each clade on the 50% majority rule consensus tree was calculated (Hall, 2001). Alternative topologies were compared using the Templeton test (Templeton, 1983) performed under the parsimony criteria and Shimodaira–Hasegawa test (Shimodaira and Hasegawa, 1999) performed under the likelihood criteria.

To test for congruence between the two data partitions, 1000 replicates of the partition homogeneity (PHT) test (essentially the incongruence length difference (ILD) test of Farris et al., 1994, 1995), as contained in PAUP 4.0b10, were applied. The partition Bremer support (PBS) scores method (Baker et al., 1998; Bremer, 1988, 1994) was also used to measure the amount of support provided by *Amd* and *COII* to each node on the total evidence phylogeny. These PBS values were calculated using PAUP 4.0b10 and TreeRot.v2 (Sorenson, 1999) and then standardized by dividing the total Bremer support of each gene by the minimum number of steps for that gene (Durando et al., 2000).

The Bayesian framework and the ModelTest results were also used for investigating and characterizing the substitution patterns in our two datasets.

3. Results and discussion

3.1. Individual analysis of the Amd gene

The multiple alignment of the 49 sequences of the *Amd* gene displayed 264 (of 516 bp) variable sites, of which 237 were parsimoniously informative. The average transition/transversion ratio was 1.3, i.e., more than 2.5 times the value expected considering a random substitution distribution. The general frequencies of the four nucleotides T, C, A, and G were, respectively, 25.2, 24.3, 20.8, and 29.8. This slight bias toward the G + C content increased when only the third codon positions were considered, reaching a threshold of 75.4 in *Drosophila griseolineata*. A saturation test (Fig. 2A), performed in order to evaluate the phylogenetic information presented by the sequences (Schneider, 2003), revealed that in most cases both transitions and transversions remained informative. But a portion of the curve for transitions appears to become partially saturated above the 16% divergence value, indicating the occurrence of multiple substitutions and homoplasy in this mutation class. As the transitions always remained higher than transversions, instead of simply ruling out those substitutions, we decided to use more complex evolutionary models which could account for multiple substitutions and homoplasy as well as for the transition/transversion ratio and the G + C content.

The Tamura 3-Parameter (Tamura, 1992) model was chosen for the NJ analysis. In this case, the distances obtained varied from 0.00 (for same species populations) to 0.345 (between *D. pseudoobscura* and *S. leonensis*) with a mean of 0.227. As in most comparisons the assigned values did not exceed the 0.30 threshold (suggested by Russo (2001) as the limit above which the differences are excessively underestimated), we chose not to use the amino acid sequences. The general time reversible (GTR) model using the ratio of invariable sites (I)

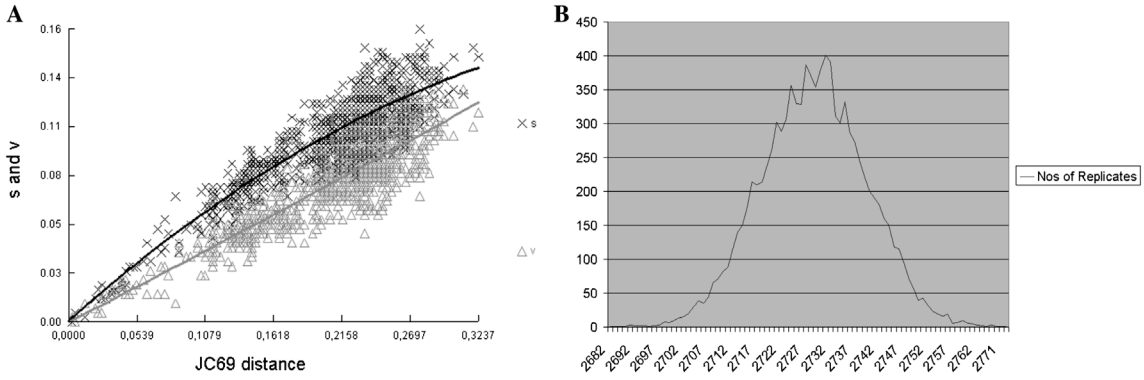

Fig. 2. Graphs showing: (A) *Amd* transition (s) and transversion (v) rates plotted against the JC distances (Jukes and Cantor, 1969) for different Drosophilidae species; (B) distribution of the scores of the randomly generated maximum parsimony (MP) trees constructed using *Amd* sequences. The score of the most parsimonious *Amd* tree was 1730.

and the gamma distribution of the variable sites (G) was the model indicated by the AIC test for likelihood analyses. For the MP analysis the length of the best tree (1730) laid well outside the scores presented by the permutated data (Fig. 2B) in the PTP test ($p = 0.0001$), attesting to the strong *Amd* phylogenetic signal. This tree presented a consistency index (ci) of 0.287 and a retention index (ri) of 0.582.

Bayesian analysis constructed 1,000,000 trees of which 6696 different trees were finally considered in the construction of the consensus tree (Fig. 3) as well as in the estimation of the posterior probabilities of each clade. The phylogenies obtained by the other methods yielded essentially similar results, since both the Templeton and the Shimodaira–Hasegawa tests detected no statistically significant differences between any of them. It was particularly interesting to see that our Bayesian topology performed better than the ML one under the likelihood criterion of the Shimodaira–Hasegawa test, although this difference was not statistically significant ($p = 0.572$). As compared with the 44 clades of the Bayesian tree the ML tree was 91% similar (40 out of 44 clades), the MP tree 82% similar (36 out of 44 clades), and the NJ tree 79% similar (35 out of 44 clades) (data not shown). So it appears that although some clusterings differ among the different trees (particularly those of more inclusive levels, especially within the *immigrans-Hirtodrosophila* radiation) there are only a few divergent clades among them. This reiteration among different inference methods can be taken as an indication of the strong internal consistency and congruency of the data.

When *Scaptodrosophila latifasciaeformis* was used as the outgroup the subgenus *Sophophora* appeared at the base of the genus *Drosophila* phylogeny, where it was preceded only by the subgenus *Dorsilopha* (Fig. 3), here represented by *D. busckii*, which composed a basal polytomy with *S. latifasciaeformis*. This early offshoot of the subgenus *Sophophora* agrees with a series of previous studies (Kwiatowski and Ayala, 1999; Remsen and O'Grady, 2002; Russo et al., 1995; Tatarenkov et al., 1999, 2001; Throckmorton, 1975) but contrasts with the findings of Grimaldi (1990) and DeSalle (1992). On the other hand, the basal position of *Dorsilopha* was unexpected because most studies have placed this subgenus among the derived clades inside the genus *Drosophila* (Grimaldi, 1990; Kwiatowski and Ayala, 1999; Tatarenkov et al., 1999, 2001; Throckmorton, 1975).

The subgenus *Drosophila* was the next to branch off and appeared to be phylogenetically closer to other Drosophilidae genera (*Zaprionus*, *Samoaia*, *Hirtodrosophila*, and *Liodrosophila*) than to the subgenus *Sophophora*. The subgenus *Drosophila* was also paraphyletically defined with respect to these taxa, which at least partially agrees with other published work (Kwiatowski and Ayala, 1999; Tatarenkov et al., 1999, 2001; Throckmorton, 1975; Yotoko et al., 2003). In contrast with these, Remsen and O'Grady (2002) presented *Hirtodrosophila*, *Samoaia*, and *Liodrosophila* as basal clades within Drosophilidae, a result that is also supported by Grimaldi's analysis (Grimaldi, 1990), where *Zaprionus* also appeared as sister to a monophyletic *Drosophila* genus.

The subgenus *Drosophila* was further subdivided into two main clades, one composed of members of the *virilis-repleta* radiation with a constitution essentially similar to that presented by Throckmorton (1975) and another composed of the traditional *immigrans-Hirtodrosophila* radiation of Throckmorton (1975) plus the species *L. aerea* and *D. funebris*. Nevertheless only the *virilis-repleta* cluster presented good support values (1.00).

As regards the *virilis-repleta* radiation, the confirmation of the *mesophragmatica* group monophyly is important given its high confidence (support value = 1.00). This group was internally subdivided into three main clades, one consisting of *D. gasici*, *D. brncici*, and *D. mesophragmatica*, another composed of the sibling species *D. pavani* and *D. gaucha* and a third basal

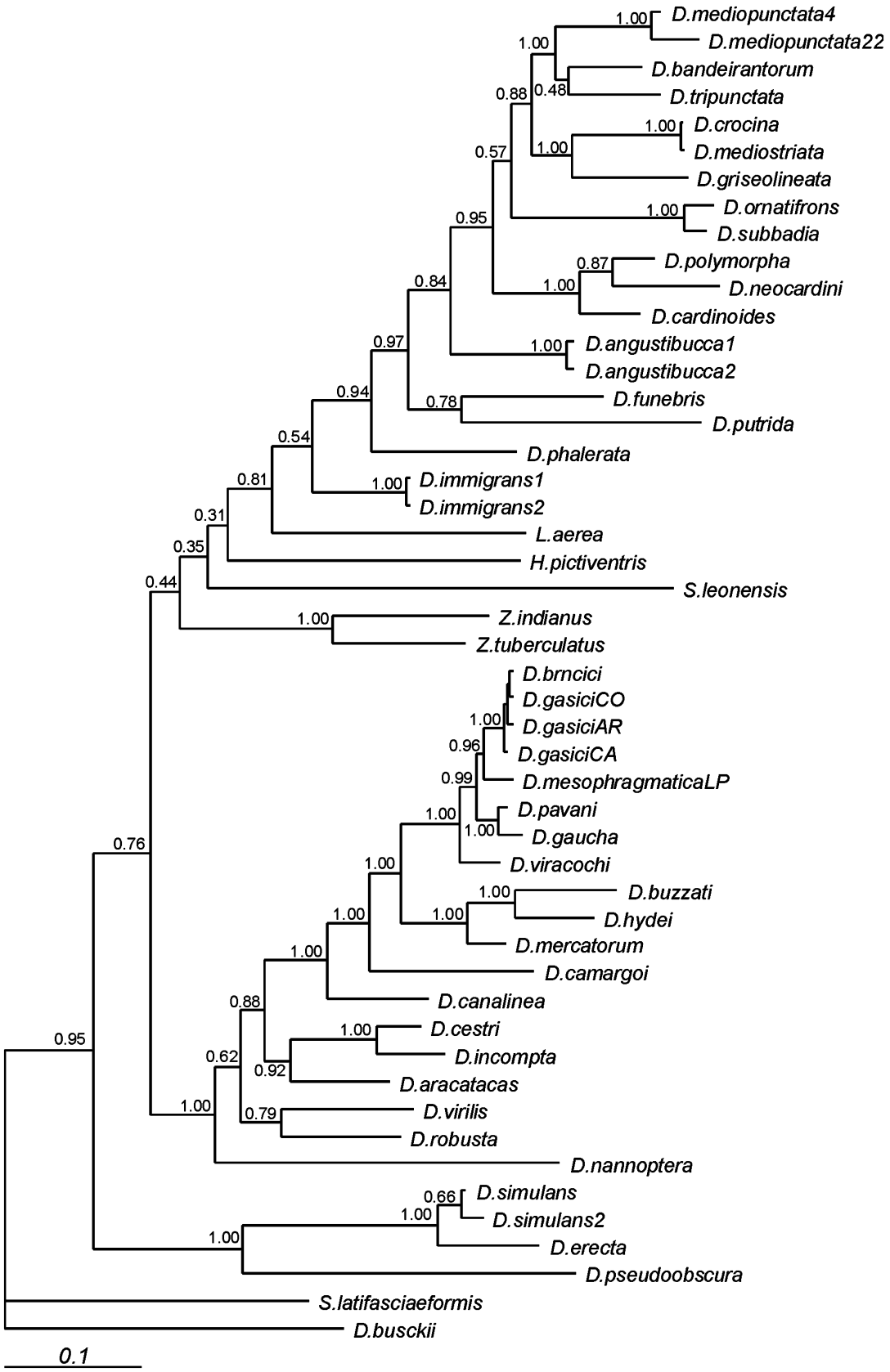

Fig. 3. The *Amd* consensus phylogeny obtained using Bayesian analysis and the GTR + I + G model as proposed by the AIC test for the 49 nucleotide sequences. The posterior probability of each clade is indicated beside its respective internal branch.

and monotypic clade containing *D. viracochi*. Such grouping is widely consistent with the subgroups obtained by Brncic et al. (1971) through cytogenetic analysis, as is also the case for the paraphyly of *D. gasici*, which was probably due to high ancestral polymorphism. The *repleta* group was the *mesophragmatica* group sister-clade, which partially agrees with some previous published work (Durando et al., 2000; Tatarenkov and Ayala, 2001) but which is at variance with the findings of Throckmorton (1975) and Remsen and O'Grady (2002). To this *mesophragmatica*/*repleta* clade it follows the basal branching of *D. camargoi* (*dreyfusi* group) and *D. canalinea* (*canalinea* group), defining a monophyletic *repleta* radiation as previously also detected by Tatarenkov and Ayala (2001). Unexpectedly, the clade formed by the *flavopilosa* species group and *D. aracatacas* (*annulimana* species group) clustered as the sister-clade of the *repleta* radiation. Throckmorton (1975) placed these groups as basal to the other *virilis-repleta* radiation members but they occupy a more derived position in this work. *D. nannoptera* (*nannoptera* group) and a clade made up of *D. virilis* (*virilis* group) and *D. robusta* (*robusta* group) were actually the first split inside the *virilis-repleta* radiation, although the first of these branching-points was only weakly supported (support value = 0.62).

Within the *immigrans-Hirtodrosophila* radiation the genus *Zaprionus* (support value = 0.35) as well as the species *S. leonensis* (0.31), *H. pictiventris* (0.81), and *L. aerea* (0.54) constitute the early offshoot, although the values in parentheses show that the statistical support for this was low. The next to branch off was *D. immigrans* (*immigrans* group), another taxon indicated by Throckmorton as composing the basis of this radiation, which is followed by *D. phalerata* (*quinaria* group) and by the clade joining *D. putrida* (*testacea* group) and *D. funebris* (*funebris* group). This last group, together with *Liodrosophila*, was proposed by Throckmorton (1975) as being basal as compared to the *virilis-repleta* and *immigrans-Hirtodrosophila* radiations, but this was not confirmed either by our work or by other published studies (Remsen and O'Grady, 2002; Tatarenkov et al., 2001; Yotoko et al., 2003), especially concerning the position of *D. funebris*. Remaining inside this major clade representing the *immigrans-Hirtodrosophila* radiation there was a monophyletic *tripunctata* radiation composed of a *tripunctata* group entirely paraphyletic in respect to the *cardini*, *guarani*, and *guaramunu* species groups (support value = 0.84). This paraphyly of the *tripunctata* group has previously been noted by several other workers (Frota-Pessoa, 1954; Throckmorton, 1975; Yotoko et al., 2003), although Yotoko et al. (2003) pointed out that the *tripunctata* radiation was also paraphyletic concerning the *funebris* and *testacea* species groups, i.e., the clades which make up the *tripunctata* radiation sister-group in our present study.

It was possible to subdivide the *tripunctata* radiation cluster into five main clades which we list here in an ascending order beginning with the most primitive: monotypic clade 1 containing *D. angustibucca* a member of *tripunctata* group subgroup I (Frota-Pessoa, 1954); clade 2 presenting a monophyletic *cardini* group whose internal composition agrees with the clusters proposed for the *cardini* subgroup by Heed and Russel (1971); clade 3 containing the two *guarani* group species, *D. ornatifrons* and *D. subbadia*; clade 4 made up of the *guaramunu* group species *D. griseolineata* along with *D. mediostriata* (synonym *D. crocina*, according to Vilela and Bächli, 1990) which is a member of subgroup III of the *tripunctata* group; clade 5 containing three members of the *tripunctata* group, *D. mediopunctata* (subgroup II), *D. bandeirantorum* (subgroup III), and *D. tripunctata* (subgroup IV). The relationships between the five clusters described above present a good overall resolution for the Bayesian tree, although for the remainder analysis performed the bootstrap values obtained for them were generally very low (results not shown). As can be seen by the arrangement of the subgroups and groups above, it was not generally possible to detect similarities between the morphological groupings obtained by Frota-Pessoa (1954) and the internal composition of the *tripunctata* radiation cluster presented here.

3.2. Individual analysis of the COII gene

For the *COII* gene the multiple alignment of the 71 sequences showed 330 (of 672 bp) variable sites of which 241 were parsimoniously informative. The medium transition/transversion ratio was 1.0 but this increased to 6.5 when only first codon positions were considered. The frequencies of the four nucleotides T, C, A, and G were also significantly biased, presenting an average of 39.9, 14.2, 32.4, and 13.5, respectively. This bias toward A + T content increased even more when only the third codon positions were maintained reaching an average of 90.6. The saturation test (Fig. 4A) revealed that, in contrast to our results for the *Amd* gene, the transition/transversion ratios for *COII* decreased significantly with the increasing of the distance, a clear sign of fast saturation of transitional substitutions due to the strong biases in both base composition and substitution patterns. However, because we wanted to solve some important questions at the intra-group level where this mutation class remains informative we decided to adopt an equal weight strategy inside a complex evolutionary model.

The Tamura–Nei model (Tamura and Nei, 1993) with a gamma distribution was chosen here for the NJ analysis. In this case, all the distances obtained varied from 0.00 (for populations of the same species but also between *D. brncici*, *D. gasiciAR*, and *D. gasiciCO*) to 0.247 (between *D. nannoptera* and *D. mediopunctata*), with a mean of 0.15. The AIC test indicated that the

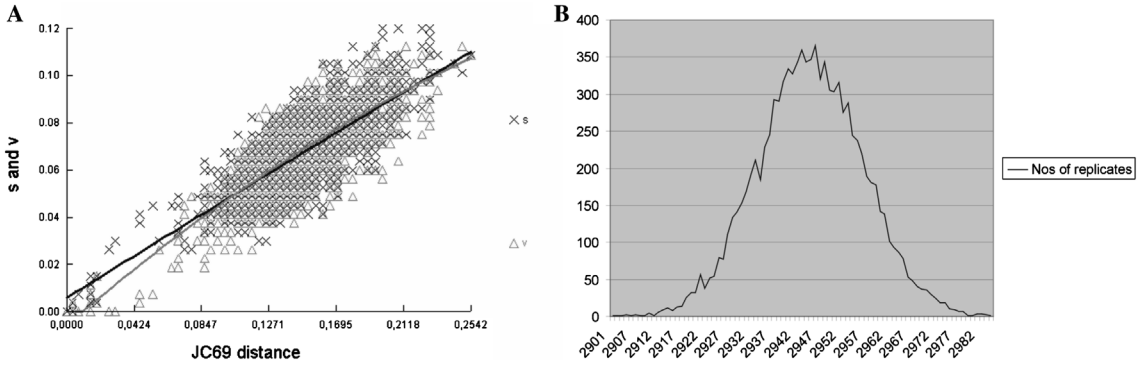

Fig. 4. Graphs showing: (A) *COII* transition (s) and transversion (v) rates plotted against the JC distances (Jukes and Cantor, 1969) for different Drosophilidae species; (B) distribution of the scores of the randomly generated MP trees constructed using *COII* sequences. The score of the most parsimonious *COII* tree was 1937.

GTR + I + G model was again the best for the likelihood analyses. The MP analysis produced a best tree whose length (1937) laid outside the scores presented by the permutated data (Fig. 4B) in the PTP test ($p = 0.0001$), which also attested to the strong *COII* phylogenetic signal. This tree presented a ci of 0.256 and a ri of 0.568.

The Bayesian analysis for COII evaluated 2,000,000 generations from which 15,243 different trees were used to construct the 50% majority rule consensus tree (Fig. 5). The phylogeny constructed by the ML method was 87.5% similar to the Bayesian tree, but only 59% of the Bayesian tree clades were also supported by the MP analysis (38 out of 64) and 64% by the NJ analysis (41 out of 64). The Templeton test performed under a parsimony criterion regarded the MP hypothesis as significantly better than the others, while the Shimodaira–Hasegawa test performed under a likelihood criterion detected no significant differences between the trees. In general, those clusters involving inter-species relationships among members of the *immigrans-tripunctata* radiation were the major source of this inconsistency, these clades also being only weakly supported by the Bayesian posterior probabilities.

As the *COII* gene is a rapidly evolving mitochondrial locus it produced some unexpected results at the species group-level analysis (Fig. 5). This is the case, for example, for the polyphyletic *repleta* group and for the "floating" placement of *D. nannoptera*, *D. simulans*, and *D. yakuba*, probably led by the phenomenon of long branch attraction, given the high saturation presented by this gene. The *COII* gene was also unable to detect the monophyly of either the *flavopilosa* group or the *repleta/virilis-repleta* radiations, although it presented a monophyletic *mesophragmatica* group, whose sister-group relationship with the *annulimana* group was not previously reported. The *mesophragmatica* group internal branching pattern was different from that obtained by *Amd* in that *D. viracochi* clustered with *D. brncici*, *D. gasici*, and *D. mesophragmatica* instead of being the early offshoot inside its group. This result also disagrees with the studies of Nacrur (1958) and Brncic et al. (1971) but it is very similar to the relationships presented by Nair et al. (1971) using isozyme variation analysis, except that our study branched *D. mesophragmatica* earlier than *D. viracochi*.

The most derived part of the *COII* tree again consisted of representatives of the *immigrans-Hirtodrosophila* radiation plus *D. funebris* and *L. aerea*, whose most general constitution was similar to that presented by *Amd*. The branching pattern among the subgenus *Dorsilopha* and the genera *Hirtodrosophila*, *Samoaia*, *Liodrosophila*, and *Zaprionus* remained obscure, given its weak confidence values, but these taxa appeared to be the early offshoot inside that major clade, leaving, again, the genus and subgenus *Drosophila* entirely paraphyletic. After these splitting, all the remaining species belonged to the subgenus *Drosophila* (support value = 0.54) but their branching pattern was also somewhat confusing. We found that the *funebris* and *quinaria* groups split off first, leaving a clade containing species of the *tripunctata* radiation plus *D. nannoptera*, *D. immigrans* (*immigrans* group), and *D. putrida* (*testacea* group) with a very weak confidence value (0.21). Thus, in contrast to the results produced with the *Amd* gene, the *tripunctata* radiation was here paraphyletic concerning these three species, a result that agrees with the findings of Yotoko et al. (2003).

Inside this major cluster that composes the *immigrans-Hirtodrosophila* radiation there were six main clades, but their external relationships do not have any statistical support. These clades are listed below in an ascendant order, beginning with the most primitive one:

- Clade 1 grouped *D. immigrans*, *D. pallidipennis*, and *D. nannoptera* (support value = 0.28) presenting them as basal inside the *tripunctata* radiation. This composes the sister-clade of a paraphyletic *tripunctata* group, presented on the remaining clades.

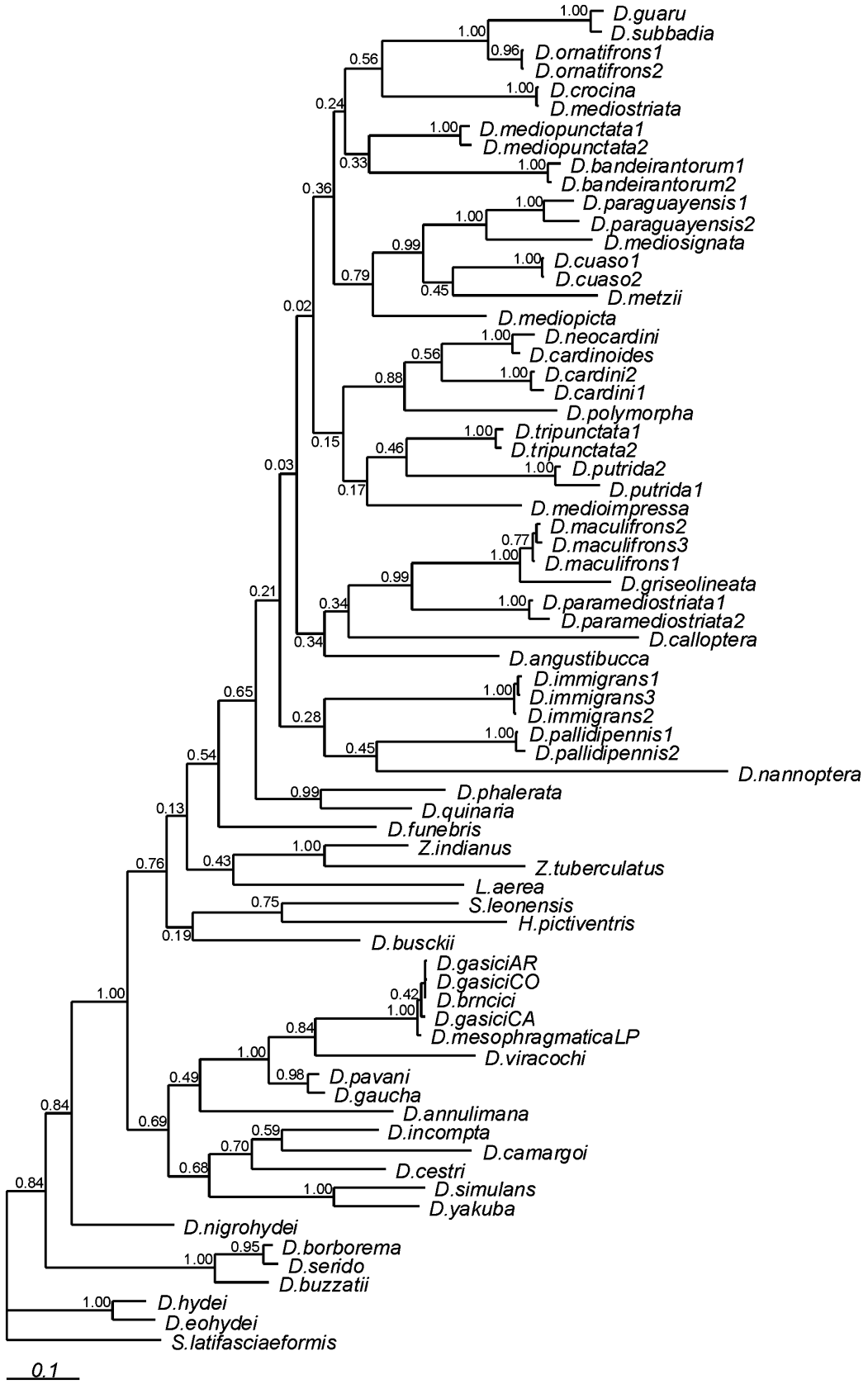

Fig. 5. The *COII* consensus phylogeny obtained using Bayesian analysis and the GTR + I + G model, as proposed by the AIC test for the 71 nucleotide sequences. The posterior probability of each clade is indicated beside its respective internal branch.

- Clade 2, partially congruent with Yotoko et al. (2003), grouped a monophyletic *guaramunu* group with *D. paramediostriata*, a member of subgroup III of the *tripunctata* group (Frota-Pessoa, 1954) (support value = 0.99) and these with *D. calloptera* (*calloptera* group) (support value = 0.34). *D. angustibucca* (*tripunctata* group, subgroup I) constituted the early offshoot inside this clade (support value = 0.34).
- Clade 3 presented the cluster of *D. putrida* (*testacea* group), *D. tripunctata* (*tripunctata* group, subgroup IV), and *D. medioimpressa* (*tripunctata* group, subgroup II) (support value = 0.17) as the sister-clade of a monophyletic *cardini* group (support value = 0.88), with a confidence value of 0.15. The *cardini* group had internal relationships which entirely conflicted with that presented by Heed and Russel (1971) and by our individual analysis of the *Amd* gene because *COII* clustered *D. neocardini* with *D. cardinoides* and *D. cardini*, leaving *D. polymorpha* on a separate branch (probability of 0.56).
- Clade 4 showed better support values joining only species of the *tripunctata* group. It presented *D. mediopicta* (subgroup III) as the sister of a cluster previously reported by Yotoko et al. (2003) that grouped *D. paraguayensis* and *D. mediosignata* (subgroup II) with *D. cuaso* (subgroup II) and *D. metzii* (subgroup IV). Bächli et al. (2000) described *D. paraguayensis*, *D. mediosignata*, and *D. cuaso* as three cryptic species.
- Clade 5 was also produced by the *Amd* analysis and clustered two representatives of the *tripunctata* group, *D. mediopunctata* (subgroup II) and *D. bandeirantorum* (subgroup III), but with no statistical support (support value = 0.33).
- Clade 6, again with a low support value (0.56), produced a new result by presenting *D. mediostriata* (*tripunctata* group, subgroup III) as the sister species of a monophyletic *guarani* group which clustered *D. guaru* with *D. subbadia* and these with *D. ornatifrons* (support value = 1.00).

3.3. Simultaneous analysis of Amd and COII

The PHT test performed to evaluate the congruence between the *Amd* and *COII* data partitions detected a significant conflict between these two datasets, as indicated by the *p* value obtained (0.001). Even so we decided to combine the two partitions, since that previous analyses made by Remsen and DeSalle (1998) and Remsen and O'Grady (2002) showed that interesting results can be produced from simultaneous analysis even if the constituent data partitions display heterogeneity.

For this new dataset we included only the 41 taxa for which the two gene sequences were available. The total number of characters analyzed was 1188 bp (516 from *Amd* and 672 from *COII*). The Tamura–Nei model with a gamma distribution was again chosen for the NJ analysis and the GTR + I + G model was again indicated by the AIC test for use in the likelihood analyses.

Bayesian analysis evaluated 2,000,000 generations from which 4111 different trees were considered in the production of the consensus tree (Fig. 6). The low number of different trees obtained is a good indicator that the 'top of the hill' was reached. Bayesian analysis again performed better than the ML one under the likelihood criterion of the Shimodaira–Hasegawa test, although the difference was not statistically significant ($p = 0.683$). The phylogenies achieved by the other methods were also overall very similar to the Bayesian topology, since neither the Templeton test nor the Shimodaira–Hasegawa test detected significant differences between them. Table 2 presents the confidence values for the Bayesian clades numbered in Fig. 6 under the different phylogenetic reconstruction methods. In general, although many of the clusters were not statistically significant (particularly those of more inclusive level) there were only a few divergent clades among the different trees, most of which again referred to the *immigrans-Hirtodrosophila* inter-group relationships. Nevertheless, the reinforcement of results obtained by different inference methods can be taken as an indication of the strong internal consistency and congruency of the data overall.

As with the *Amd* analysis, following the separation of *S. latifasciaeformis* and *D. busckii*, *D. simulans* emerged as the sister-clade of all the remaining Drosophilidae species (Fig. 6). This early offshoot of the subgenus *Sophophora* left a paraphyletic subgenus *Drosophila*, phylogenetically closer to the genera *Hirtodrosophila*, *Samoaia*, *Zaprionus*, and *Liodrosophila* than to the subgenus *Sophophora*. Our total evidence tree therefore clearly showed a well-supported (probability of 0.97) paraphyletic subgenus *Drosophila* in accordance with the *Amd* data and several studies by other authors (Kwiatowski and Ayala, 1999; Remsen and O'Grady, 2002; Russo et al., 1995; Tatarenkov et al., 1999, 2001; Thomas and Hunt, 1993; Throckmorton, 1975; Yotoko et al., 2003).

There was also strong support for the *immigrans-Hirtodrosophila* (support value = 0.85) and *virilis-repleta* (support value = 0.96) radiations within the subgenus *Drosophila*, although the first of these clusters did not display the same composition as that presented by Throckmorton (1975) because of the inclusion of *D. funebris* and *L. aerea*. This last result was also noted by Tatarenkov et al. (2001) concerning the two species and by Remsen and O'Grady (2002) and Yotoko et al. (2003) concerning only *D. funebris*. The *virilis-repleta* radiation branching pattern was extremely similar to that obtained with the *Amd* gene individual analysis.

In the major *immigrans-tripunctata* radiation cluster, a clade joining the two *Zaprionus* species with *L. aerea*

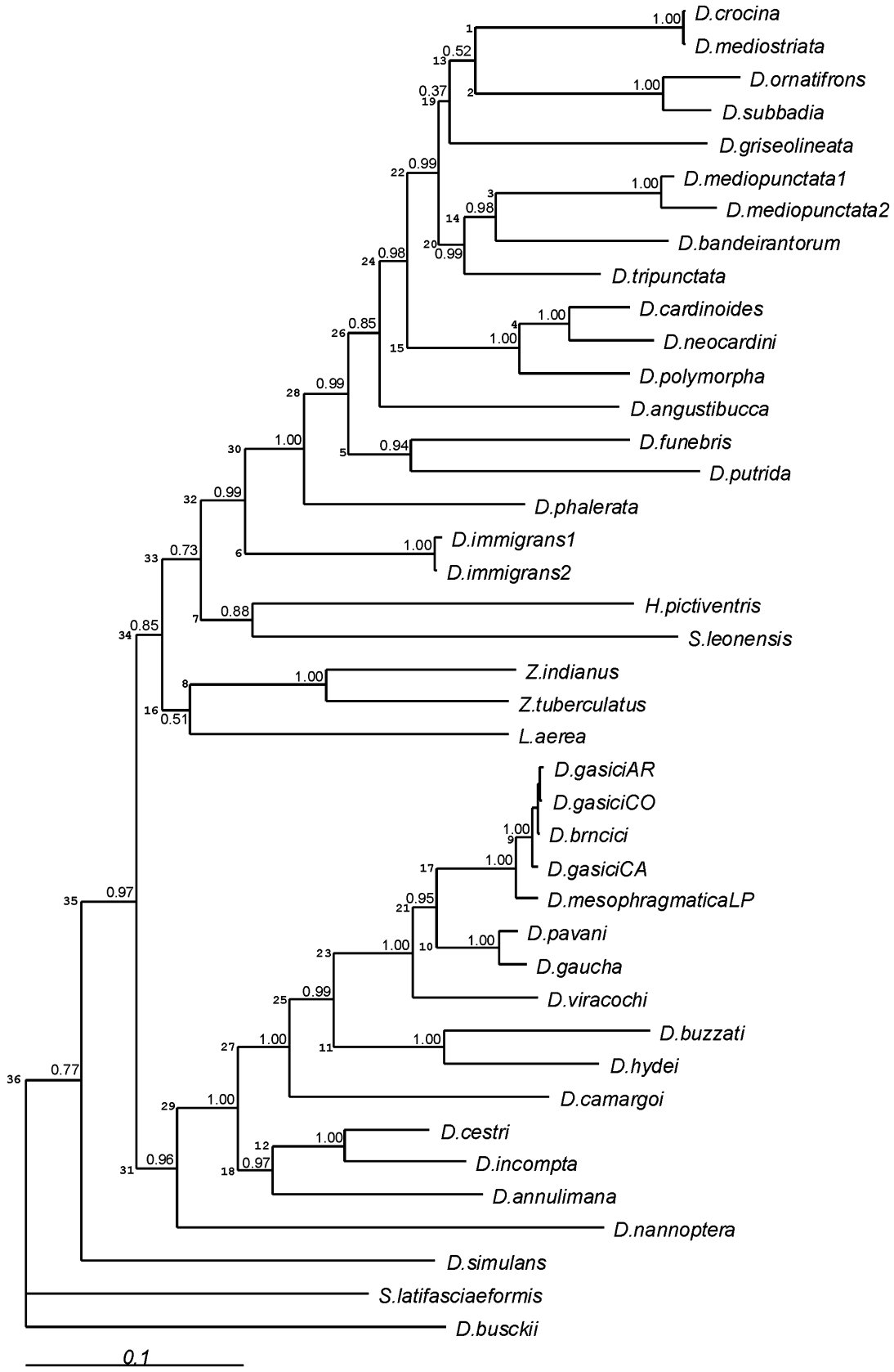

Fig. 6. Total evidence consensus phylogeny obtained using Bayesian analysis and the GTR + I + G model, as proposed by the AIC test for the 41 nucleotide sequences. The posterior probability of each clade is indicated beside its respective internal branch. At the front of the internal branches there is an arbitrarily defined number representing the clade that follows and which can be used to interpret Tables 2 and 3.

Table 2
Bootstrap values presented for each of the clades shown in Fig. 6 by the trees constructed using different phylogenetic reconstruction methods

Clade	Phylogenetic reconstruction method		
	NJ	MP	ML
1	100	100	100
2	100	100	100
3	100	100	100
4	77	91	93
5	42	33	—
6	100	100	100
7	—	—	27
8	100	98	99
9	99	98	100
10	100	100	99
11	98	96	98
12	100	97	98
13	—	—	—
14	38	38	49
15	99	99	99
16	—	—	—
17	100	100	100
18	58	61	53
19	—	—	—
20	40	17	29
21	—	52	71
22	11	—	18
23	100	100	99
24	—	—	16
25	70	48	65
26	36	39	22
27	50	40	63
28	27	39	48
29	83	50	71
30	50	57	55
31	67	26	45
32	37	35	44
33	—	—	—
34	—	—	11
35	—	—	18
36	32	37	36

constituted an early split (support value = 0.73), followed by the clade grouping *H. pictiventris* and *S. leonensis* (support value = 0.99). These species also constituted the early offshoot of the *immigrans-Hirtodrosophila* radiation in our individual *Amd* and *COII* analyses. Similar results have also been reported by Throckmorton (1975), by Remsen and DeSalle (1998) nuclear combined analysis for *Zaprionus* and *Hirtodrosophila*, by Tatarenkov et al. (2001) for *Liodrosophila*, *Zaprionus*, and *Samoaia*, and by Yotoko et al. (2003) for *Zaprionus* only. Other studies, however, have pointed to these genera as being outgroups for all *Drosophila* subgenera other than *Sophophora* (Russo et al., 1995; Tatarenkov et al., 1999) and even as the outgroups to the genus *Drosophila* (Grimaldi, 1990; DeSalle, 1992 concerning *Zaprionus* and *Hirtodrosophila*; Remsen and O'Grady, 2002 concerning *Samoaia* and *Hirtodrosophila*; Yotoko et al., 2003 concerning *Hirtodrosophila*). Therefore, more studies are necessary in order to reach a more reliable phylogeny.

After the separation of these taxa, all the remaining species belong to the subgenus *Drosophila*. In this combined analysis, the basal split order was essentially similar to that of the individual *Amd* analysis, with *D. immigrans* branching off first followed by *D. phalerata* and by the *D. putrida*/*D. funebris* clade. In this analysis the *tripunctata* radiation was also monophyletic and had a good confidence value (0.85).

The *tripunctata* radiation cluster produced by our total evidence data was subdivided into four main clades, whose differences with those clades obtained individually by *Amd* referred to the internal composition of the *cardini* group, which more closely resembled that obtained with *COII* analysis, and the placement of the *guarani* clade inside the clustering of *D. griseolineata* and *D. mediostriata*, although with a low support value (0.37). This *D. mediostriata*/*guarani*/*D. griseolineata* cluster can be assigned to the mixture of the *COII* gene information joining *D. mediostriata* with the *guarani* group and the *Amd* gene signal grouping *D. mediostriata* with *D. griseolineata*. The branching order and inter-clade relationships nevertheless remained essentially similar to that found for *Amd* individually.

To assess the relative contribution of the two gene regions to our simultaneous analysis (SA) hypothesis the PBS scores for each gene on each node were calculated. This test only confirmed the pattern described above, i.e., the *Amd* gene provided almost twice the *COII* support and contributed much more to our simultaneous analysis (Table 3). The breakdown of the PBS values also indicated substantial conflict between these two partitions, because 11 of the 36 nodes resolved in our SA consensus tree had conflicting PBS values. This pattern may have been the result of either differences in internal homoplasy within each gene or conflicting signals between the genes.

The different MP and Bayesian topologies inferred using the three distinct datasets (reduced by the number of superposed taxa) were also statistically tested among each other by the Templeton and Shimodaira–Hasegawa tests. Overall, the *Amd* gene data did not reject the SA trees while the *COII* topologies were statistically worse than the former under all tests performed. Otherwise, the *Amd* and SA trees were significantly rejected by the *COII* partition. Under the combined dataset only the *COII* topologies were significantly rejected by both tests, while the *Amd* topologies were not statistically different from the SA topologies. Therefore, the phylogenetic trees based on the *COII* gene were again very inconsistent with those based on the *Amd* gene and on the combined dataset, but, overall, the *Amd* and SA topologies appear not to have been significantly different.

Table 3
Partitioned Bremer support (PBS) scores across the simultaneous analysis Bayesian tree for each gene partition in each of the clades presented in Fig. 6

Clade	*Amd*	*COII*
1	20.2	24.8
2	15	12
3	12	18
4	−2	10
5	0	3
6	14.2	31.8
7	0	0
8	15.7	3.3
9	6	0
10	5	11
11	15.7	−1.7
12	14	−4
13	0	0
14	1	−1
15	8	3
16	0	0
17	2	15
18	1.6	4.4
19	0	0
20	1	−1
21	1.6	0.4
22	0	0
23	8	4
24	0	0
25	11	−7
26	3.3	1.7
27	11	−7
28	−2	7
29	19	−13
30	7	0
31	0	0
32	−3	8
33	0	0
34	0	0
35	0	0
36	5.7	−0.7
Total PBS	191	122
Min steps	497	496
Total PBS/**Min steps**[a]	**0.38**	**0.24**

[a] PBS values summed across the tree and standardized by the minimum number of steps for each partition.

3.4. Amd and COII molecular evolution patterns

Given the detected conflict between the *Amd* and *COII* partitions, it is important to identify the source of this disagreement. By using the Modeltest results and performing Bayesian analysis using the GTR + site-specific rates (SSR) model it was possible to quantitatively compare the substitution patterns presented by the nuclear *Amd* and the mitochondrial *COII* genes. It is well known that nuclear genes generally evolve more slowly than mitochondrial genes, making them better markers for deep divergences (Simon et al., 1994). However, when we consider our data alone this pattern is not seen because the relative rate presented by *Amd* (1.11) was somewhat greater than that presented by *COII* (0.91). A better understanding of rate variation can be obtained by looking at the rates among codon positions within genes, from which it can be seen that the first, second, and third codon positions of the *Amd* gene evolved faster than the *COII* codon positions (Fig. 7A). Indeed, this pattern could already be expected because the median distance values obtained for *Amd* (0.227) were greater than those for *COII* (0.15). It should not be forgotten, however, that these results could be a by-product of the saturation presented by our mitochondrial gene.

It was previously seen that the *COII* gene showed greater base compositional bias than the *Amd* gene, but other patterns of nucleotide substitution, such as the Q matrix of transformation, may also be important. The instantaneous rate matrix for *COII* was more asymmetrical relative to that presented by *Amd* and was also more skewed towards one type of change over another (Fig. 7B), although in both cases there was a higher overall rate of transitions. Another contrasting point that can be seen in Fig. 7B is that while the *Amd* nuclear gene showed very similar overall rates of transversions, the *COII* mitochondrial dataset had a rate of CG transversions which was almost 14 times higher than the rate of AT transversions. The obvious consequence of this highly skewed transformation rate matrix is greater levels of homoplasy, which are not easily corrected for.

Another parameter that often differs between nuclear and mitochondrial genes is the shape of the gamma distribution, as given by the α value, describing the among-site rate variation. As can be seen from Fig. 7C, the *Amd* gene showed a more homogeneous pattern of among-site rate variation as compared to the *COII* gene, which had lower α values. The *Amd* gene also presented a higher proportion of invariable sites (Pi values), a parameter which is positively correlated with α in coding regions according to Lin and Danforth (2004). These authors also showed a positive correlation between α and ci, the consistency index, which suggest that data partitions with more heterogeneous substitution rates show a higher level of homoplasy. This correlation was also detected by us, with the *Amd* gene having a higher ci value (0.287) than the *COII* gene (0.256).

The combination of all these factors may together explain the overall poor performance of the mitochondrial *COII* gene relative to the nuclear *Amd* gene, given that all these properties should lead to high levels of homoplasy and, therefore, saturation. The choice of complex models that account for these biased substitution patterns appeared not to solve this problem.

4. Conclusion

Our study reinforces the work previously published by Goto and Kimura (2001) and Lin and Danforth (2004) in

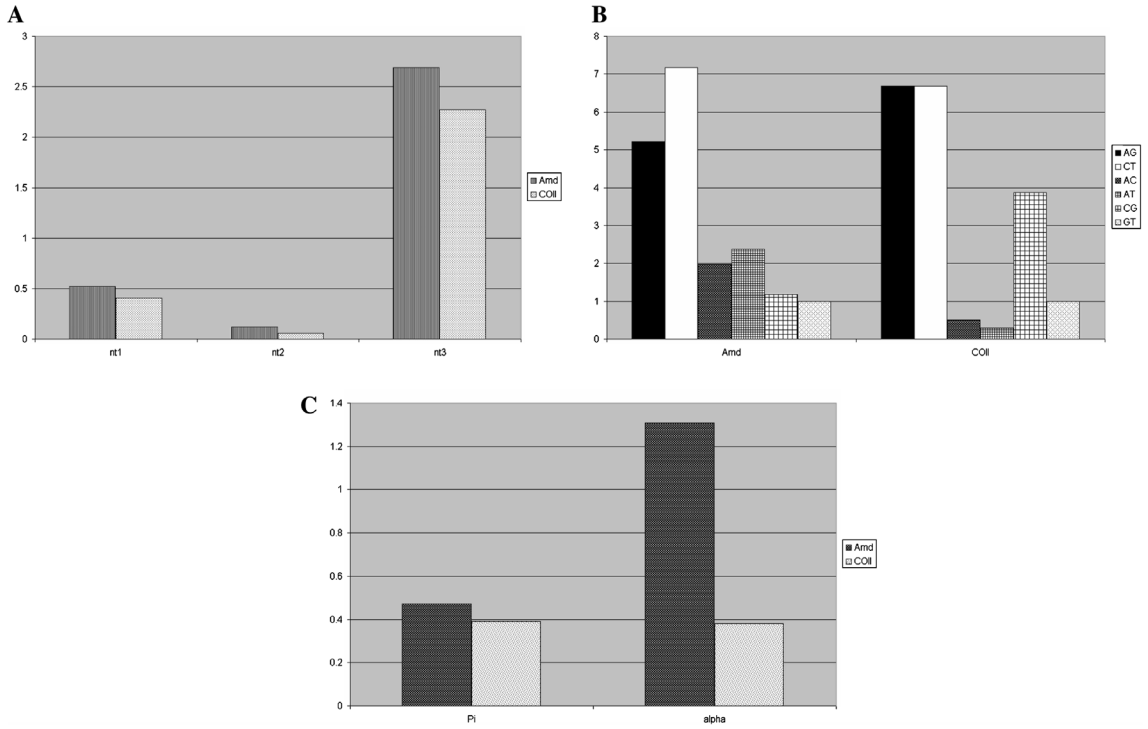

Fig. 7. The *Amd* and *COII* nucleotide substitution patterns: (A) relative rates of nucleotide substitution based on the GTR + SSR model, with sites partitioned by gene and by codon position; (B) *Amd* and *COII* transformation rate matrices obtained with AIC test performed on ModelTest; (C) *Amd* and *COII* gamma shape and proportion of invariable sites (Pi) obtained with AIC Test.

that it shows that nuclear genes have a slight advantage over mitochondrial genes when used for phylogenetic analyses, especially these of more inclusive levels. The *Amd* gene had higher consistency index (ci) values and almost twice the partition Bremer support (PBS) scores as compared to the *COII* gene. The partition homogeneity test (PHT) revealed significant incongruence between *Amd* and *COII* and the Templeton and the Shimodaira–Hasegawa tests indicated that the maximum parsimony (MP) and the Bayesian trees (the two types of trees tested) constructed using the *COII* gene were significantly rejected by the *Amd* and the combined dataset. In addition, the *COII* support values were rather lower than the *Amd* support values, especially for between-group comparisons. The poor phylogenetic performance of the *COII* gene can be attributed to its strong A/T bias associated with its highly asymmetrical rate matrices and high among-site rate heterogeneity. These properties appear to have resulted in the *COII* mitochondrial gene having high levels of homoplasy which are not easily corrected for by the use of complex evolutionary models. Lin and Danforth (2004) state that the properties that characterize genes suitable for phylogenetic studies appear to be the opposite to those which characterize the *COII* gene, i.e., genes better suited for phylogenetic analysis have a more even base composition, higher α values, and a less skewed transformation rate matrix. Otherwise, although some saturation occurs in the *Amd* gene, at least as far as transitions are concerned, the phylogenetic utility of this gene can be recovered by the use of complex evolutionary models.

In our study, the phylogenetic trees based on the *COII* gene were significantly different from those based on the *Amd* gene and the combined dataset and we consider our total evidence tree (TET) (Fig. 6) as the best phylogenetic hypothesis to now available for these species. Except for minor differences, our TET was essentially similar to the phylogenetic scheme produced by Throckmorton (1975). In this sense, it proposes a paraphyletic genus *Drosophila* concerning the genera *Hirtodrosphila*, *Liodrosophila*, *Zaprionus*, and *Samoaia* and a paraphyletic subgenus *Drosophila* concerning these taxa but not the subgenus *Dorsilopha* which constitutes a basal polytomy with *S. latifasciaeformis*. The subgenus *Drosophila* also constitutes a derived split within the genus *Drosophila*, while the subgenus *Sophophora* is the early offshoot. The major division of species in the subgenus *Drosophila* species occurs between the *virilis-repleta* and *immigrans-Hirtodrosophila* radiations proposed by Throckmorton (1975), but in our study the last radiation yielded a somewhat different composition because it included *Liodrosophila* and *D. funebris* as basal species within the *immigrans-Hirtodrosophila* radiation while Throckmorton (1975) assigned them as basal species within the subgenus *Drosophila*. The paraphyly of the *tripunctata* group pointed out by various authors (Frota-Pessoa, 1954; Throckmorton, 1975; Yotoko et al., 2003) was confirmed by us, although we found that the

tripunctata radiation was monophyletic. Even so it was not possible to find more similarities between the morphological subdivision of Frota-Pessoa (1954) and the internal composition of our *tripunctata* radiation. This was also the most inconsistent region of our trees, with a low confidence level and incongruent results in regard to the different phylogenetic reconstruction methods and to the different datasets. A possible reason for this is the fact that the internal branches concerning the *tripuncata* radiation internal relationships were overall short. All these patterns support the scenario originally presented by Throckmorton (1975) and also stated by Yotoko et al. (2003) in which these taxa emerged by means of fast and multiple substitution episodes. If these results come to be confirmed, a large set of combined and informative datasets need to be accumulated in order to reach a more reliable phylogeny for these species.

Finally, it is worth reiterating that, taxonomically, the family Drosophilidae remains an agglomeration of merophyletic taxa (merophyly designates the absence of monophyly according to Bernardi (1981)), beginning with the genus *Drosophila* and its most diverse subgenus, the subgenus *Drosophila*. As our results show, genera emerging from subgenera are a normal feature of the taxonomy of these taxa. Nevertheless, despite the accumulation of evidence, such as that presented here, few effective attempts have been made to review the nomenclature of the family and reconcile its taxonomy with recent systematic studies.

Acknowledgments

We are grateful to the following people: Marco Silva Gottschalk (M.Sc.) for kind collaboration on the identification of the species cited in this paper; Liliana Essi (M.Sc.) and Paulo Marcos Pinto (M.Sc.) for the valuable help in difficult moments; Marcos de Oliveira Carvalho for assistance with bioinformatics; Adriana Ludwig for providing the *flavopilosa* group DNA; Dr. Robert Thomas and Mr. Félix Nonnemacher for revision of English. We also thank the Brazilian agencies CNPq for providing grants and fellowships, and CNPq, FAPERGS and PROPESQ-UFRGS research funding.

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Automat. Contr. 19, 716–723.

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N., Sutton, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X., Brandon, R.C., Rogers, Y.H., Blazej, R.G., Champe, M., Pfeiffer, B.D., Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor, G.L., Abril, J.F., Agbayani, A., An, H.J., Andrews-Pfannkoch, C., Baldwin, D., Ballew, R.M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y., Benos, P.V., Berman, B.P., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M.R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K.C., Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B., Davies, P., de Pablos, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M., Dodson, K., Doup, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C., Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferriera, S., Fleischmann, W., Fosler, C., Gabrielian, A.E., Garg, N.S., Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z., Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J., Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J., Wei, M.H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z., Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C., McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Mobarry, C., Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L., Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K., Nusskern, D.R., Pacleb, J.M., Palazzolo, M., Pittman, G.S., Pan, S., Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K., Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I., Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C., Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R., Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wassarman, D.A., Weinstock, G.M., Weissenbach, J., Williams, S.M., Woodage, T., Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F., Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H., Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M., Venter, J.C., 2000. The genome sequence of *Drosophila melanogaster*. Science 287, 2185–2195.

Bächli, G., Vilela, C.R., Ratcov, V., 2000. Morphological differences among *Drosophila paraguayensis* Duda 1927 and its close relatives (Diptera, Drosophilidae). Bull. Soc. Entomol. Suisse 73, 67–92.

Baker, R.H., Yu, X., DeSalle, R., 1998. Assessing the relative contribution of molecular and morphological characters in simultaneous analysis trees. Mol. Phylogenet. Evol. 9, 427–436.

Bernardi, N., 1981. Phylogenetic relationships, monophyletic group and related concepts. Rev. Bras. Entomol. 25 (4), 323–326.

Bremer, K., 1988. The limits of amino acid sequence data in angiosperm phylogenetics reconstruction. Evolution 42, 795–803.

Bremer, K., 1994. Branch support and tree stability. Cladistics 10, 295–304.

Brncic, D., Nair, P.S., Wheeler, M.R., 1971. Cytotaxonomic relationships within the *mesophragmatica* species group of *Drosophila*. Univ. Texas Publ. 7103, 1–16.

Clary, D.O., Goddard, J.M., Martin, S.C., Fauron, C.M., Wolstenholme, D.R., 1982. *Drosophila* mitochondrial DNA: a novel gene order. Nucleic Acids Res. 10 (21), 6619–6637.

DeSalle, R., 1992. The phylogenetic relationships of flies in the family Drosophilidae deduced from mtDNA sequences. Mol. Phylogenet. Evol. 1, 31–40.

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, 296–307.

Farris, J.S., Källersjö, M., Kluge, A., Bult, C., 1994. Testing significance of congruence. Cladistics 10, 315–320.

Farris, J.S., Källersjö, M., Kluge, A., Bult, C., 1995. Constructing a significance test for incongruence. Syst. Biol. 44, 570–572.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approaching using bootstrap. Evolution 39, 783–791.

Frota-Pessoa, O., 1954. Revision of the *tripunctata* group of *Drosophila* with description of fifteen new species. Arq. Mus. Paranaense 10, 253–304.

Goto, S.G., Kimura, M.T., 2001. Phylogenetic utility of mitochondrial *COI* and nuclear *Gpdh* genes in *Drosophila*. Mol. Phylogenet. Evol. 18 (3), 404–422.

Grimaldi, D.A., 1990. A phylogenetic, revised classification of the genera in the Drosophilidae (Diptera). Bull. Am. Mus. Natl. Hist. 197, 1–139.

Hall, B.G., 2001. Phylogenetic Trees Made Easy—a How to Manual for Molecular Biologists, second ed. Sinauer Associates, Massachusetts.

Heed, W.B., Russell, J.S., 1971. Phylogeny and population structure in island and continental species on the *cardini* group of *Drosophila* studied by inversion analysis. Univ. Texas Publ. 7103, 91–130.

Huelsenbeck, J.P., Ronquist, F., 2001. MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754–755.

Jeanmougin, F., Thompson, J.D., Gouy, M., Higgins, D.G., Gibson, T.J., 1998. Multiple sequence alignment with ClustalX. Trends Biochem. Sci. 23, 403–405.

Jowett, T., 1986. Preparation of nucleic acids. In: Roberts, D.B. (Ed.), *Drosophila*: a Practical Approach. IRL Press, Oxford, pp. 275–276.

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

Kastritsis, C.D., 1969. The chromosomes of some species of the *guarani* group of *Drosophila*. J. Hered. 60, 50–57.

Kumar, S., Tamura, K., Jakobsen, I.B., Nei, M. 2001. MEGA: Molecular Evolutionary Genetics Analysis Software. Version 2. Pennsylvania State University, University Park, and Arizona State University, Tempe.

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

Kwiatowski, J., Krawczyk, M., Jaworsky, M., Sharecky, D., Ayala, F.J., 1997. Erratic evolution of glycerol-3-phosphate dehydrogenase in *Drosophila*, *Chymomyza* and *Ceratitis*. J. Mol. Evol. 44, 9–22.

Kwiatowski, J., Ayala, F.J., 1999. Phylogeny of *Drosophila* and related genera: conflict between molecular and anatomical analyses. Mol. Phylogenet. Evol. 13, 319–328.

Lin, C.-P., Danforth, B.N., 2004. How do insect nuclear and mitochondrial gene substitution patterns differ? Insights from Bayesian analysis of combined datasets. Mol. Phylogenet. Evol. 30, 686–702.

Malogolowkin, C., 1953. Sobre a genitália dos drosofilídeos. IV. A genitália masculina no subgênero *Drosophila* (Diptera, Drosophilidae). Rev. Bras. Biol. 13, 245–264.

Nacrur, J., 1958. Genitália masculina de *Drosophila* do grupo *mesophragmatica* (Diptera). Rev. Bras. Biol. 18, 243–249.

Nair, P.S., Brncic, D., Kojima, K.-I., 1971. II. Isozyme variations and evolutionary relationships in the *mesophragmatica* species group of *Drosophila*. Stud. Genet. VI, 17–28.

Nurminsky, D.I., Moriyama, E.N., Lozovskaya, E.R., Hartl, D.L., 1996. Molecular phylogeny and genome evolution in the *Drosophila virilis* species group: duplications of the *alcohol dehydrogenase* gene. Mol. Biol. Evol. 13, 132–149.

O'Grady, P.M., Kidwell, M.G., 2002. Phylogeny of the subgenus *Sophophora* (Diptera: Drosophilidae) based on combined analysis of nuclear and mitochondrial sequences. Mol. Phylogenet. Evol. 22 (3), 443–453.

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

Perlam, S.J., Spicer, G.S., Shoemaker, D.D., Jaenike, J., 2003. Associations between mycophagous *Drosophila* and their *Howardula* nematode parasites: a worldwide phylogenetic shuffle. Mol. Ecol. 12 (1), 237–249.

Posada, C., Crandall, K.A., 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14, 817–818.

Powell, J.R., DeSalle, R., 1995. *Drosophila* molecular phylogenies and their uses. In: Hecht, M.K. (Ed.), Evolutionary Biology. Plenum, New York, pp. 87–139.

Remsen, J., DeSalle, R., 1998. Character congruence of multiple data partitions and the origin of the Hawaiian Drosophilidae. Mol. Phylogenet. Evol. 9 (2), 225–235.

Remsen, J., O'Grady, P., 2002. Phylogeny of Drosophilinae (Diptera: Drosophilidae), with comments on combined analysis and character support. Mol. Phylogenet. Evol. 24, 249–264.

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

Russo, C.A.M.de, 2001. Como escolher genes para problemas filogenéticos específicos. In: Matioli, S.R. (Ed.), Biologia molecular e evolução. Holos, Ribeirão Preto.

Saitou, N., Nei, M., 1987. The Neighbor-Joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 9, 406–425.

Schneider, H., 2003. Métodos de análise filogenética—Um guia prático. Holos and SBG, Ribeirão Preto.

Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16, 1114–1116.

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., Flook, P., 1994. Evolution, weighting and phylogenetic utility of mitochondrial gene-sequences and a compilation of conserved polymerase chain reaction primers. Ann. Entomol. Soc. Am. 87, 651–701.

Sorenson, M.D. 1999. TreeRot, version 2. Boston University, Boston, MA.

Spicer, G.S., 1995. Phylogenetic utility of the mitochondrial cytochrome oxidase gene: molecular evolution of the *Drosophila buzzatii* species complex. J. Mol. Evol. 41 (6), 749–759.

Spicer, G.S., Jaenike, J., 1996. Phylogenetic analysis of breeding site use and amanitin tolerance within the *Drosophila quinaria* species group. Evolution 50, 2328–2337.

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

Staden, R., 1996. The Staden sequence analysis package. Mol. Biotechnol. 5, 233–241.

Swofford, D.L. 2003. PAUP: phylogenetic analysis using parsimony (and other methods). Version 4. Sinauer Associates, Massachusetts.

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

Tamura, K., Nei, M., 1993. Estimation of the number of base nucleotide substitution in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. 10, 512–526.

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

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

Tatarenkov, A., Zurovcová, M., Ayala, F.J., 2001. *Ddc* and *Amd* sequences resolve phylogenetic relationships of *Drosophila*. Mol. Phylogenet. Evol. 20, 321–325.

Templeton, A.R., 1983. Convergent evolution and non-parametric inferences from restriction fragment and DNA sequence data. In: Weir, B. (Ed.), Statistical Analysis of DNA Sequence Data. Marcel Dekker, New York, pp. 151–179.

Thomas, R.H., Hunt, J.A., 1993. Phylogenetic relationships in *Drosophila*: a conflict between molecular and morphological data. Mol. Biol. Evol. 10 (2), 362–374.

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

Val, F.C., Marques, M.D., Vilela, C.R., 1981. Drosophilidae of Neotropical region. In: Ashburner, M., Carson, H.L., Thompson, J.N.

(Eds.), The Genetics and Biology of Drosophila. Academic Press, Orlando, pp. 123–168.

Vilela, C.R., Bächli, G., 1990. Taxonomic studies on Neotropical species of seven genera of (Drosophilidae Diptera). Bull. Soc. Entomol. Suisse 63 (Suppl.), 1–332.

Vilela, C.R., Bächli, G., 2000. Five new species of Neotropical *Drosophila* (Diptera, Drosophilidae). Bull. Soc. Entomol. Suisse 73, 49–65.

Vilela, C.R., Valente, V.L.S., Basso-da-Silva, L., 2004. *Drosophila angustibucca* Duda *sensu* Frota-Pessoa is an undescribed species (Diptera, Drosophilidae). Rev. Bras. Entomol. 48, 233–238.

Xia, X., Xie, Z., 2001. DAMBE: data analysis in molecular biology and evolution. J. Hered. 92, 371–373.

Yotoko, K.S.C., Medeiros, H.F., Solferini, V.N., Klaczko, L.B., 2003. A molecular study of the systematics of the *Drosophila tripunctata* group and the *tripunctata radiation*. Mol. Phylogenet. Evol. 28, 614–619.

Wheeler, M.R., 1982. The Drosophilidae: a taxonomic overview. In: Ashburner, M., Carson, H.L., Thompson, J.N. (Eds.), The Genetics and Biology of Drosophila. Academic Press, Orlando, pp. 1–97.