Phylogenetic Utility of Mitochondrial *COI* and Nuclear *Gpdh* Genes in *Drosophila*

Shin G. Goto and Masahito T. Kimura

GOTO, S. G., & KIMURA, M. T. (2001). Phylogenetic Utility of Mitochondrial COI and Nuclear Gpdh Genes in Drosophila. *Molecular Phylogenetics and Evolution*. 18, 404-422. https://doi.org/10.1006/mpev.2000.0893

Phylogenetic Utility of Mitochondrial *COI* and Nuclear *Gpdh* Genes in *Drosophila*

Shin G. Goto and Masahito T. Kimura

Graduate School of Environmental Earth Science, Hokkaido University, Sapporo, Hokkaido 060-0810, Japan

Phylogenetic utility of the mitochondrial *COI* (cytochrome oxidase subunit I) and nuclear *Gpdh* (glycerol3-phosphate dehydrogenase) genes was studied in the *Drosophila melanogaster* species group. The rate of substitution was higher in the *COI* gene than in the *Gpdh* gene. In addition, multiple substitutions, not only for transitional but also for transversional substitutions, occurred faster in the *COI* gene. None of the trees obtained using the *COI* gene supported the well-established monophyly of the *ananassae* subgroup. In addition, the incongruence length difference test, Templeton test, and partitioned Bremer support revealed that the trees based on the *COI* data are considerably different from those based on the *Gpdh* and the combined data set. Thus, the *COI* gene did not show good phylogenetic performance in the *melanogaster* group. The present analyses based on the *Gpdh* gene and the combined data set revealed that the *ananassae* subgroup branched off first in the *melanogaster* group followed by the *montium* subgroup and further by the *melanogaster* subgroup in contrast to the most recent phylogenetic hypothesis based on *Amy* multigenes.

INTRODUCTION

A major assumption of molecular systematics is that gene trees accurately reflect species trees. However, molecular phylogeny based on a gene can yield a gene tree that differs from the species tree in the strict sense. Therefore, it is very important to compare gene trees based on genes from different linkage groups to reconstruct the species tree.

Mitochondrial DNA sequences (or genes) have been extensively used to estimate phylogenies because of the relative technical ease for sequencing from divergent taxa and their special features, i.e., lack of introns, maternal inheritance, absence of recombination events, and haploidy (reviewed in Meyer, 1993; Avise, 1994). On the other hand, several protein-coding nuclear genes have recently been identified as having potential utility for phylogenetic analyses (Friedlander *et al.,* 1992, 1994, 1996; Palumbi and Baker, 1994; Slade *et al.,* 1994; Barrio and Ayala, 1997; Fang *et al.,* 1997; Rodríguez-Trelles *et al.,* 1999; Tatarenkov *et al.,* 1999).

The usefulness of different genes for reconstructing a phylogeny can be influenced by the substitution properties of those genes. Mitochondrial protein-coding genes generally evolve fast and attain saturation rapidly, possibly due to a deficient mismatch repair system and/or A+T-rich base composition (Brown *et al.,* 1979; DeSalle *et al.,* 1987; Tamura, 1992; Moriyama and Powell, 1997). Therefore, mitochondrial genes are not always good phylogenetic performers (Russo *et al.,* 1996; Zardoya and Meyer, 1996). However, Prychitko and Moore (1997) and Johnson and Clayton (2000) found that nuclear and mitochondrial genes revealed identical phylogenies in Aves. In addition, Murrel *et al.* (2000) found no phylogenetic incongruence between 12S rDNA and *COI* in Rhipicephalinae ticks. Moreover, O'Grandy *et al.* (1998) reported that the *COI* data are congruent to the *COII, Adh,* morphology, and total evidence tree but not to ITS1 in the *Drosophila saltans* group.

Here we compare nucleotide sequence data of mitochondrial cytochrome oxidase subunit I (*COI*) and nuclear glycerol-3-phosphate dehydrogenase (*Gpdh*) genes in the *Drosophila melanogaster* species group, subgenus *Sophophora,* genus *Drosophila,* to assess the phylogenetic utility of mitochondrial and nuclear genes. These genes have been extensively used for phylogenetic analysis in *Drosophila* (Barrio and Ayala, 1997; Gleason *et al.,* 1997; Kwiatowski *et al.,* 1997; Goto *et al.,* 1999, 2000).

Another goal of this study is to analyze the phylogenetic relationships among subgroups in the *melanogaster* species group. The *melanogaster* group is favored for many evolutionary studies (reviewed by Lemeunier *et al.,* 1986). In this group, 12 species subgroups, the *ananassae, denticulata, elegans, eugracilis, ficusphila, flavohirta, melanogaster, montium, suzukii, takahashii, rhopaloa,* and *longissima* subgroups, have been recognized (Okada, 1984; Lemeunier *et al.,* 1986; Toda, 1991).

The phylogenetic analyses for the *melanogaster* group have been extensively studied based on morphology (Bock and Wheeler, 1972; Bock, 1980), biogeography (Throckmorton, 1975; Lemeunier *et al.,* 1986) and molecular data (Ashburner *et al.,* 1984; Beverley and Wilson, 1984; Tsacas and Tscas, 1984; Kim *et al.,* 1993; Pélandakis and Solignac, 1993; Russo *et al.,* 1995; Nikolaidis and Scouras, 1996; Inomata *et al.,* 1997). However, the relationships among these subgroups are still complicated. From the study of the periphallic organs, Hsu (1949) considered that the *suzukii* subgroup was closest to the *Drosophila obscura* stem and that two lines, *melanogaster–takahashii* and *ananassae–montium,* arose from the *suzukii* subgroup. Okada (1954)

recognized three series from morphology, (1) *suzukii,* (2) *melanogaster–takahashii–ficusphila,* and (3) *ananassae–montium,* and remarked on the affinity of the *eugracilis* subgroup with the *takahashii, suzukii,* and *ficusphila* subgroups. Bock and Wheeler (1972) argued in favor of the separation of the *ananassae* and *montium* subgroups from the others early in the evolution of the species group and pointed out the resemblances between the *suzukii* and the *elegans* subgroups. Ashburner *et al.* (1984) recognized at least three lineages, *ananassae, montium,* and *elegans–eugracilis–ficusphila–melanogaster–suzukii–takahashii,* from the integration of chromosomal and morphological data. However, the phylogenetic analyses based on DNA sequence are still limited. Based on the sequence data of rDNA, Pélandakis and Solignac (1993) recognized five lineages, (1) *ananassae,* (2) *montium,* (3) *ficusphila–elegans,* (4) (*takahashii–suzukii*)*–eugracilis,* and (5) *melanogaster,* and suggested that the *ananassae* subgroup was closest to the *obscura* group (Fig. 1A). On the other hand, Inomata *et al.* (1997) recognized different lineages inferred from *Amy* (α-amylase) multigenes with very low bootstrap values (Fig. 1B). In their tree, some sequences within species even clustered with those of other species.

MATERIALS AND METHODS

Flies

In this study, 8 species were sequenced for the *Gpdh* and *COI* genes (shown with collection locality in Table 1). In addition to these species, 16 species reported in our previous studies (Goto *et al.,* 2000) and others (de Bruijn, 1983; Satta and Takahata, 1990; Reed and Gibson, 1994; Barrio and Ayala, 1997; Gleason *et al.,* 1997; Kwiatowski *et al.,* 1997) were used for the analysis. The DDBJ/GenBank/EMBL accession numbers for the sequences used in the present study are given in Table 1.

DNA Extraction, PCR Amplification, and DNA Sequencing

Multiple flies (~5 individuals) were homogenized in 100 μL of TES (0.1 M Tris–HCl, pH 8.0/0.1 M EDTA/1% SDS) and incubated at 70°C for 30 min. After addition of 14 μL of 8 M potassium acetate, the mixture was centrifuged at 10*g* for 10 min at 4°C. The supernatant was mixed with 0.5 vol of isopropanol, held for 5 min at room temperature, and precipitated by centrifugation. The pellet was rinsed with 70% ethanol, dried, and dissolved in ddH_2O. RNAs in the samples were digested with RNase A.

The primers used to amplify *Gpdh* fragments were GNL-mel and GNR-mel (Goto *et al.,* 2000; Table 2). For *D. ficusphila,* GNL and GNR primers were used (Barrio and Ayala, 1997; Table 2). PCR was performed with 100 ng of DNA, 1 U of AmpliTaq DNA polymerase (Perkin–Elmer), and a final concentration of 1.5 mM $MgCl_2$, 1x PCR buffer II as formulated by Perkin– Elmer, a 0.4 μM concentration of each primer, and 0.2 mM dNTP in a total volume of 50 μL. Amplification parameters used were 35 cycles of 30 s of denaturing at 94°C, 30 s of annealing at 57°C, and 90 s (2 and 4 min for *D. ficusphila* and *D. suzukii,* respectively) of extension at 72°C. The sequence of the *Gpdh* gene analyzed in this study was 430 bp in length (57 bp from the 3′ end of exon 3 and 373 bp from the complete exon 4).

The primers used to amplify *COI* were F- and R-COI (Gleason *et al.,* 1997; Goto *et al.,* 1999, 2000; Table 2). PCR components were the same as those used for the amplification of *Gpdh* except the primers. Amplification parameters used were 35 cycles of 30 s of denaturing at 94°C, 30 s of annealing at 50°C, and 30 s of extension at 72°C. The sequence of the *COI* gene analyzed in this study was 407 bp in length and the first base corresponded to position 2205 in the *D. melanogaster* mtDNA sequence (the GenBank accession number is U37541).

The amplified fragments were purified using a QIAquick Gel Extraction Kit (QIAGEN). The sequences were obtained from an ABI 373A automated sequencer (PE Applied Biosystems) with a DNA sequencing kit (Dye Terminator Cycle Sequencing Ready Reaction; PE Applied Biosystems) according to the suppliers' instructions.

Cycle sequencing was performed using five oligonucleotides (GNL-mel, L4BN, R4M, Gpdh-F, and Gpdh-R; Barrio and Ayala, 1997; Goto *et al.,* 1999, 2000) and two oligonucleotides (F-COI and R-COI) for *Gpdh* and *COI* fragments, respectively (Table 2). In the sequencing results of *Gpdh,* distinct polymorphisms within single species were not detected.

Phylogenetic Inference

For the phylogenetic analysis, we used the maximum-parsimony (MP; Swofford and Olsen, 1990) and neighbor-joining (NJ; Saitou and Nei, 1987) methods with PAUP 4.0b4a (Swofford, 2000) and the maximum likelihood (ML; Felsenstein, 1981) method with PUZZLE 4.0.2 (Strimmer and von Haeseler, 1996). The ML analyses were performed using the HKY algorithm (Hasegawa *et al.,* 1995) with a discrete approximation to the Γ-distribution, and the transition/transversion ratio (Ts/Tv), the fraction of invariable sites (θ), and the shape parameter (α) were estimated according to the model. In addition, the mixed model of rate heterogeneity (1 invariable rate + 10 Γ-distributed rates) was executed for PUZZLE 4.0.2 (Strimmer and von Haeseler, 1996).

Since the total base composition does not reflect the base composition of the positions that are free to vary (see Results), the base composition used in the ML calculations was based only on the base frequencies of the variable positions. For the NJ analyses, we used the K80 (Kimura, 1980), TN93 (Tamura and Nei, 1983), and Log Det/Paralinear (Lake, 1994; Lockhart *et al.,* 1994) models. In addition, we adopted the parameters (Ts/Tv, θ, and α) estimated from the Γ-discrete ML analysis for the NJ analysis (Spicer, 1995; Sullivan *et al.,* 1996). Sullivan *et al.* (1996) reported that parsimony-based estimates of Ts/Tv ratio are severe underestimates.

The statistical confidence of a particular cluster of sequences in the NJ and MP trees was evaluated by the bootstrap test (Felsenstein, 1985). In addition, we assessed the level of confidence with the decay indexes (Bremer, 1988) for the MP tree. Moreover, to assess the confidence we used the quartet-puzzling scores (Strimmer and von Haeseler, 1996; Strimmer *et al.,* 1997) for the ML trees.

The incongruence length difference test (ILD; Mickevich and Farris, 1981; Farris *et al.,* 1994, 1995) was executed to assess the incongruence among the *COI, Gpdh,* and combined data set with PAUP 4.0b4a (Swofford, 2000). One thousand permutations were executed to generate a null distribution of tree length differences.

The partition Bremer support (PBS; Bremer, 1988, 1994; Baker and DeSalle, 1997) was also executed to measure the amount of support provided by *COI* and *Gpdh* to each node on the total evidence phylogeny. PBS shows the contribution of each partition to the decay index of every node on the total evidence tree (Baker and DeSalle, 1997). All partition lengths for any given node will always sum to the decay index for that node on the total evidence tree. The PBS values were calculated with PAUP 4.0b4a (Swofford, 2000) and TreeRot.v2 (Sorenson, 1999).

Finally, the Templeton test (Templeton, 1983) was used to evaluate the obtained phylogenetic trees.

In all the analyses, *D. bifasciata* was used as an outgroup.

RESULTS

Base Composition and Bias

Table 3 shows the percentage and absolute numbers of variable and phylogenetically informative sites when considering both ingroup taxa and when including the outgroup (*D. bifasciata*). As can be seen in Table 3, most of the positions that have changed are third-position sites, which is expected since substitutions in most first- and second-codon positions result in nonsynonymous changes. The *Gpdh* and *COI* are not significantly more variable than each other as determined by χ^2 test. This includes the comparisons not only when just considering the ingroup taxa ($P = 0.849$), but also when considering the outgroup taxa ($P = 0.400$). In addition, the same holds true not only in the ingroup comparison ($P = 0.852$) but also in the comparison considering the outgroup taxa ($P = 0.933$) when only the phylogenetically informative characters are considered.

Table 4 shows base composition and base composition bias. The base composition bias is calculated according to Irwin *et al.* (1991) and ranges in value between 0 and 1, with zero indicating no bias and one indicating complete bias. The base compositions are significantly biased (χ^2 test, $P < 0.01$) in both *Gpdh* and *COI.* The bias was considerable in *COI;* the fragments of 21 *Drosophila* species had a high proportion (68.9–71.4%) of A+T, especially in third-codon positions (89.6–97.8%) and at fourfold degenerate sites (90.3–100%) (data not shown). For phylogenetic reconstruction, the variable and phylogenetically informative positions will be important. As can be seen in Table 4, noticeable differences exist in the amount of bias between all positions and variable or phylogenetically informative sites.

Divergence in Gpdh Gene

Figure 2A shows the transition to transversion ratio (Ts/Tv) based on *p*-distance in *Gpdh.* Transitions were generally higher than transversions in the comparisons between species belonging to the same subgroup and between those belonging to different subgroups (Ts/Tv > 1).

The *p*-distances of transitions and transversions were clearly larger in between-subgroups comparisons than in within-subgroup comparisons (Figs. 3A and 3B). However, the range of the distances between species groups (*D. bifasciata*) considerably overlapped with the range of the distances between species subgroups not only in transitions but also in transversions.

Divergence in COI Gene

In contrast to the *Gpdh,* the scatter plots of the *p*-distances between all substitutions and transitions or transversions were considerably overlapped between the within- and the between-subgroup(s) comparisons (Figs. 3C and 3D).

In addition, the Ts/Tv ratios decrease dramatically with the increase of all substitutions (Fig. 2B). The ratio ranged from 0.15 to 7.01. The highest ratios are between closely related species, i.e., *D. biauraria–D. triauraria.* The lower ratios are for comparisons between species belonging to the different subgroups. This was reflected in the following phenomenon: for transversional divergences within the subgroup, the mean Ts/Tv ratio was 2.07, while for the divergences between the subgroups, the ratio was 0.68. Thus, a strong bias for transitional substitutions is present in the pairs of closely related species, with a loss of this bias for those of more distantly related species. This phenomenon has been explained by the fast saturation of transitional substitutions due to the strong biases in both base composition and substitution patterns. Therefore, for the phylogenetical distant comparisons, transitional substitutions of *COI* likely provide little phylogenetic information.

Comparison of Gpdh and COI Genes

Figure 4 compares the divergence between the *COI* and the *Gpdh* genes. The divergence in the *COI* gene based on all substitutions increased with the divergence in the *Gpdh* gene at the initial stage (Fig. 4A). The rate of divergence was 1.6 times higher in the *COI* gene than in the *Gpdh* gene. However, the divergence in the *COI* gene reached a plateau when it exceeded 8–9%. Similarly the divergence in *COI* based on transversions or substitutions only at first- and secondcodon positions reached a plateau when it was plotted against the divergence in the *Gpdh* (Figs. 4B and 4C). Thus, multiple substitutions in the *COI* gene are apparent not only for transitions but also for transversions and substitutions at first- and second-codon positions.

Phylogenetic Analyses Based on Gpdh

Figure 5A shows the strict consensus MP tree based on the *Gpdh.* In this tree, the *ananassae* subgroup branched off first followed by the *montium, elegans, ficusphila, melanogaster, suzukii,* and *takahashii* subgroups. However, the probabilities are rather low in almost all nodes at the subgroup levels, except the relationships between the *suzukii* and the *takahashii* subgroups (bootstrap = 91%, decay index = 5). The positions of the *melanogaster* and *ficusphila* subgroups are especially unreliable. In addition, the resolution of the relationships within subgroups is not so high. The monophyly of the *ananassae, elegans,* and *takahashii* subgroups was supported by high probabilities.

The ML analysis shows a limited resolution but almost all nodes show high quartet-puzzling scores (Fig. 5B). This tree confirmed that the *ananassae* subgroup branched off first followed by the *montium* subgroup. In contrast to the MP tree, the *melanogaster* subgroup branched off next. The remaining subgroups formed a cluster as a polychotomy. In the *montium* subgroup, *D. bocki* and *D. watanabei* formed a cluster.

The NJ tree with the K80 (Kimura, 1980) model is nearly identical to the ML tree (Fig. 6A); the *ananassae* subgroup branched off first followed by the *montium* subgroup, and the *melanogaster* subgroup branched off next. In the remaining clade, the *elegans* subgroup branched off first followed by the *ficusphila, suzukii,* and *takahashii* subgroups. However, the position of *D. ficusphila* shows a rather low probability. In the *takahashii* subgroup, the relationships among species are resolved well and *D. prostipennis* branched off first, but the positions of the remaining species are supported by low probabilities. In the *montium* subgroup, *D. watanabei* branched off first and *D. bocki* was next, but its probability was unreliable.

The analysis for the data having base compositional bias can yield an underestimation of the true amount of divergence (Saccone *et al.,* 1989, 1993; Collins *et al.,* 1994). Therefore, we used the TN93 distance (Tamura and Nei, 1993) and Log Det/Paralinear procedure (Lockhart *et al.,* 1994; Lake, 1994) to compensate for the base composition bias. The TN93 distance produced the same topology and almost similar bootstrap values with the K80 model (data not shown). The Log Det/Paralinear procedure also produced a nearly identical tree with similar bootstrap values except that *D. ficusphila* branched off earlier than the *elegans* subgroup (Fig. 6B). However, the position of the *elegans* subgroup shows unreliable probability (23%).

Jin and Nei (1990) reported that the difficulty with distance estimation is whether the data meet the assumptions of the model. When the phylogeny is estimated with modification for the Γ-distribution, the answer depends on the value of the shape parameter (α) that is employed. Even if the parameters estimated from the ML analysis (Table 5) are adopted for the NJ analysis, the obtained tree is nearly identical to other NJ trees, except that *D. ficusphila* and the *elegans* subgroup formed a cluster with very low probability and *D. trilutea* branched off first followed by *D. lutescens* in the *takahashii* subgroup (Fig. 6C).

Phylogenetic Analyses Based on COI

Figures 7A and 7B show the strict consensus MP trees based on *COI* using all substitutions and only transversions, respectively. In both trees, the relationships at the basal level were poorly resolved, and probabilities are worse. In addition, the monophyly of the *ananassae* subgroup was not supported; i.e., *D. bipectinata* branched off early and *D. ananassae* formed a cluster with the

species in the *montium* subgroup (*D. bocki* or *D. watanabei*). Moreover, in the tree using only transversion, *D. backi* branched off independently from the *montium* subgroup. In the *takahashii* subgroup, *D. lutescens* branched off first when all substitutions were considered, but *D. prostipennis* did so when only transversions were considered.

The same holds true in the ML tree (Fig. 7C). *D. bipectinata* does not form a cluster with *D. ananassae,* and each subgroup branched off independently except that the *suzukii* and *takahashii* subgroups formed a cluster. In contrast to other trees, this ML analysis revealed the close relationships of *D. trilutea–D. takahashii* and *D. lutescens–D. prostipennis* in the *takahashii* subgroup with a very short branch.

On the other hand, the NJ tree with K80 model using all substitutions resolved the relationship among subgroups well with very low reliabilities (Fig. 8A). The *elegans* subgroup branched off first followed by the *montium* subgroup. In the remaining clade, the *ficusphila, melanogaster, suzukii,* and *takahashii* subgroups branched off in that order. However, the monophyly of the *ananassae* subgroup was still not supported. In the *montium* subgroup, *D. constricta* was positioned in a different lineage from *D. trapezifrons.* The relationships among the species belonging to the *takahashii* subgroup are unique. The NJ analysis with the Log Det/Paralinear procedure and TN93 model using all substitutions also produced the same topological trees (data not shown).

On the other hand, the K80 and TN93 models using only transversions recognized two main lineages with unreliable probability: *montium–*(*ficusphila–elegans*) and *melanogaster–*(*suzukii–takahashii*) (Fig. 8B). This tree also does not support the monophyly of the *ananassae* subgroup. In contrast to the NJ analysis using all substitutions, *D. constricta* positioned in the same lineage as *D. trapezifrons.* In the *takahashii* subgroup, the branching pattern of species was different from the previous trees in the present study.

The NJ analysis with the ML-estimated parameters (Table 5) recognized different lineages with very low reliabilities: *suzukii–*(*melanogaster–takahashii*) and *ficusphila–*(*elegans–montium*) (Fig. 8C). In this tree, the monophyly of the *ananassae* subgroup was still not supported, and *D. ananassae* formed a cluster with *D. watanabei* with a rather higher bootstrap value (75%). In addition, *D. biauraria* did not form a cluster with *D. triauraria* in contrast to the other trees.

Simultaneous Analyses

To test the incongruence between *COI* and *Gpdh,* the ILD (Mickevich and Farris, 1981; Farris *et al.,* 1994, 1995) test was performed. The *COI* data show significant incongruence with the *Gpdh* (ILD = 13, $P = 0.001$) and combined data set (ILD = 14, $P = 0.025$). On the other hand, *Gpdh* does not show such incongruence with the combined data set (ILD = 2, $P = 0.996$). Figure 9 shows the MP tree based on the combined data set. The relationships among subgroups are identical to the MP tree derived from *Gpdh,* but the bootstrap values and decay indexes are rather higher (Fig. 9 and Table 6). In the *takahashii* subgroup, the relationships among species are unique.

The PBS (Bremer, 1988, 1994; Baker and DeSalle, 1997) was also executed to measure the amount of support provided by *COI* and *Gpdh* to each node on the total evidence phylogeny (Table 6). In all nodes, the *Gpdh* supports the total evidence tree positively. In addition, the PBS values of *Gpdh* are generally higher than those of *COI.* On the other hand, the *COI* data do not support the monophyly of the *ananassae* subgroup (node 1 in Fig. 9) and the most basal node (node 2 in Fig. 9). The PBS values of *COI* are higher than those of *Gpdh* only in some within-subgroup comparisons.

The ML analysis of the combined data set resolved better than other ML analyses of *Gpdh* and *COI* (Fig. 10). This tree is somewhat different from the MP analysis based on the combined data set: i.e., *ficusphila–* (*suzukii–takahashii*), *elegans* and *melanogaster* formed a cluster as a polychotomy. This tree also supports the early separation of the *ananassae* and *montium* subgroups from the others.

The NJ analyses with the K80 (Fig. 11), TN93, and Log Det/Paralinear (data not shown) model produced the same topological trees. The relationships among subgroups are identical to that of the NJ tree based on the *Gpdh* with the K80 and TN93 distances (Fig. 6A) with rather higher bootstrap values except the position of the *elegans* subgroup (bootstrap = 48%). The relationships among the species belonging to the *takahashii* subgroup are the same as the NJ analysis with the K80 distance using only transversions of *COI* (Fig. 8B). In the *montium* subgroup, the relationships among species are identical to the MP tree based on the combined data set.

The NJ analysis with the ML-estimated parameters (Table 5) produced a unique topology with rather high probabilities (Fig. 12). This tree also confirmed that the *ananassae* subgroup branched off first followed by the *montium* subgroup. In contrast to the other trees, the *elegans, melanogaster, ficusphila, suzukii,* and *takahashii* subgroups branched off in that order in the remaining clade. However, the position of the *melanogaster* subgroup was supported worse. The relationships among species belonging to the *takahashii* subgroup are the same as those from the NJ analysis based on the *COI* with K80 model using all substitutions (Fig. 8A). In the *montium* subgroup, the relationships among species are the same as all of the trees based on the combined data set.

Templeton Test

Finally, the Templeton parsimony test (Templeton, 1983) was performed on the phylogenetic hypotheses inferred from all the different analyses (15 trees from Figs. 5–12). The results of this test (Table 7) revealed that a tree shown in Fig. 8A (from the NJ analysis with the K80, TN93, and Log Det/Paralinear models using all substitutions of *COI*) is best for the *COI* data but is rejected by the *Gpdh* and combined data set. All of the trees based on the *COI* data are significantly rejected by the *Gpdh* and combined data sets. On the other hand, a tree shown in Fig. 9 (from the MP analysis based on the combined data set) is best for the *Gpdh* and the combined data set and is not rejected by the *COI* data. In addition, this tree produced highest rescaled consistency index (Table 7). Moreover, this tree was supported as the best tree by the Kishino– Hasegawa (Kishino and Hasegawa, 1989) test (data not shown).

DISCUSSION

Phylogenetic Utility of Gpdh and COI

It has been reported that the rate of nucleotide substitution is higher in mitochondrial genes than in nuclear genes in *Drosophila* (DeSalle *et al.,* 1987; Satta *et al.,* 1987; Tamura, 1992; Moriyama and Powell, 1997) as well as in mammals (Brown *et al.,* 1982; Miyata *et al.,* 1982). The present study also revealed that the rate of substitution was 1.6 times higher in the *COI* gene than in the *Gpdh* gene. This rate contrasts with the result from mammals, in which the mtDNA evolves ≥ 10 times faster than single-copy nuclear DNA (Brown *et al.,* 1982). However, it agrees with the results from Satta *et al.* (1987) in *Drosophila* mtDNA.

Mitochondrial genes have also been suggested to reach the state of multiple substitution faster than nuclear genes, especially for transitional substitutions, perhaps because of the higher substitution rates and A+T-rich composition (Nei, 1987; Moriyama and Powell, 1997). The present study also revealed that a strong bias for transitional substitutions is present in pairs of closely related species, with a loss of this bias in those of more distantly related species (Fig. 2B). The higher transitional substitution rate in *COI* would diminish the phylogenetic signals because multiple substitutions erase history and create homoplasy (Baker and DeSalle, 1997). Therefore, for the phylogenetically distant comparisons, transitional substitutions of the *COI* likely provide little phylogenetic information. Thus, Gleason *et al.* (1997) recommended the use of only transversions for phylogenetic analysis using the *COI* gene. In addition, DeSalle *et al.* (1987) used all substitutions for close relatives but only transversions for more distant branch points in their study on ND1, 2, and 5 genes in Hawaiian *Drosophila.* Beckenbach *et al.* (1993) reported that biologically meaningful groups separate out much more clearly when transitions are ignored than when all substitutions are included in the analyses of *COII* in the *obscura* species group.

However, Moriyama and Powell (1997) found that transversional synonymous substitutions as well as transitional substitutions in several mitochondrial genes including *COI* seem to reach saturation between the *melanogaster* and the *obscura* groups. In the present study, the range of transversional divergences in the *COI* gene for between-subgroup comparisons considerably overlapped with the range of divergences for within-subgroup comparisons (Fig. 3D). In addition, comparison of the *COI* and *Gpdh* genes suggested that transversions or substitutions only at the first and second-codon positions in the *COI* gene are saturated even between species subgroups of the *melanogaster* group (Figs. 4B and 4C). Thus, the transversional and first- and second-codon positional substitutions would likely provide limited phylogenetic information.

In this study, the phylogenetic trees based on the *COI* gene are considerably inconsistent with those based on the *Gpdh* gene, the combined data set, and previous phylogenetic hypotheses, even if only transversions were adopted. In particular well-established monophyly of the *ananassae* subgroup (Pélandakis *et al.,* 1991; Pélandakis and Solignac, 1993; Inomata *et al.,* 1997) was not supported by all of the trees based on the *COI* data. In addition, the PBS values of the *COI* were generally lower than those of the *Gpdh* in the total evidence tree. Moreover, the Templeton test also revealed that all of the trees based on the *COI* by various methods are significantly rejected by the *Gpdh* and combined data sets (Table 7). Furthermore, the ILD test reveals a significant difference between *COI* and *Gpdh* or combined data sets.

One argument against combining different phylogenetic characters is that a large data set may "swamp" the phylogenetic signal of a smaller data set (Hillis, 1987; Miyamoto and Fitch, 1995). In our study, this did not occur. The *Gpdh* data set (430 total characters, 134 variables, and 94 informative sites) was not significantly different from the *COI* set (407 total characters, 116 variables, and 88 informative sites), yet the *Gpdh* data provided 69.7% of the total PBS values in the total evidence tree (Table 6).

The basal node (node 2) in Fig. 9 was negatively supported with the PBS values of the *COI.* The PBS values of the *COI* were higher than those of the *Gpdh* only in the comparison within subgroups. In addition, the bootstrap values of *COI* are rather lower in between-subgroup comparisons than in within-subgroup comparisons. DeSalle *et al.* (1987), Beckenbach *et al.* (1993), and Gleason *et al.* (1997) recommended the use of a mitochondrial gene for closely related species in *Drosophila.* Moreover, Gleason

et al. (1998) reported that the phylogenetic tree derived from the *COI* was incongruent with the one from the nuclear genes in the *D. willistoni* group. These results might suggest that the phylogenetic utility of the *COI* gene is rather limited for analyses within species subgroups in *Drosophila.* To examine this possibility, we constrained the monophyly of each subgroup (nodes 1, 6, 9, 12, and 13 in Fig. 9) and calculated the subsequent ILD. In this measurement, the ILD values does not drop and *COI* still shows significant incongruence with *Gpdh* (ILD = 13, $P = 0.003$) and the combined data set (ILD = 14, $P = 0.031$). In addition, even if the node(s) showing negative value(s) of PBS (nodes 1 and/or 2 in Fig. 9) was (were) constrained, *COI* still shows significant incongruence with *Gpdh* and the combined data set: i.e., the ILD does not drop in the comparison to *Gpdh* (node 2 constrained: ILD = 13, $P = 0.002$; nodes 1 and 2 constrained: ILD = 13, $P = 0.004$), and it does increase in the comparison to the combined data set (ILD = 19, $P = 0.033$; ILD = 15, $P = 0.029$). These results might indicate that the problem with *COI* is for comparisons within the subgroup as well as for comparisons among distantly related species and/or among subgroups.

However, Baker and DeSalle (1997) and O'Grandy *et al.* (1998) encountered an asymmetrical pattern of the ILD using several data sets. For instance, the ILD shows several data to be incongruent with some other partitions, but each of these partitions is congruent with at least one other partition. Therefore, no partition is in conflict with all other partitions. They concluded that if a partition was homogeneous when compared to at least one other partition, it should be included in the total-evidence analysis.

The incongruence of the *COI* with the *Gpdh* and combined data set would be explained to some extent by the multiple substitutions not only in transitions but also in transversions. One more explanation might be offered by rate heterogeneity. It has been shown that extreme nucleotide substitution rate heterogeneity can present difficulties for all methods of phylogeny reconstruction (Sullivan *et al.,* 1995). The analyses presented in Table 5 suggest that the estimated rates are different between *COI* and *Gpdh.* Lower α and higher ρ in the *COI* suggest the extreme rate heterogeneity (Yang, 1996).

In contrast to the results in the present study, O'Grandy *et al.* (1998) found that the *COI* is consistent with the *COII* (mitochondrial), *Adh* (nuclear), morphology, and the total evidence tree (*COI* + *COII* + *Adh* + ITS1 + morphology) but not with ITS1 in the *D. saltans* group, as determined by the ILD test. The phylogenetic utility of the *COI* might be different between the groups even in *Drosophila.*

In the *Gpdh* gene, it seems that saturation does not occur at least at the species subgroup level. However, the range of divergences between species groups considerably overlapped with the range of divergences between species subgroups, suggesting that saturation occurs to some extent at the level of the species group. Therefore, it would be recommended to use additional criteria (i.e., different genes with lower saturation rates or amino acid substitution) to resolve phylogenetic relationships among species groups, subgenera, and/or genera.

However, we note here that a recent study by Yang (1998) shows that the bias, commonly attributed in the literature to saturation, may have been exaggerated. Simulation analyses reveal that saturation occurs only at a much higher level of sequence divergence than has previously been suggested. In addition, Yang (1998) has pointed out that, by some current criteria, many data sets would be declared as saturated, even before enough substitutions have accumulated to be informative. According to this, a problem much more serious than saturation might have occurred at least in *COI* in the *melanogaster* group.

Phylogenetic Relationships

First we note that none of the trees in the present study completely match the previous hypotheses inferred from rDNA (Pélandakis and Solignac, 1993; Fig. 1A) and *Amy* multigenes (Inomata *et al.,* 1997; Fig. 1B).

The Templeton (1983) test revealed that a tree shown in Fig. 8A (the NJ analysis based on *COI* with the K80, TN93, and Log Det/Paralinear models using all substitutions) is best for the *COI* data but is rejected by the *Gpdh* and combined data set. On the other hand, a tree shown in Fig. 9 (the MP analysis based on the combined data set) is best for *Gpdh* and the combined data set. This tree is not rejected by the *COI.* In addition, rescaled consistency index and the Kishino–Hasegawa test in *Gpdh* and the combined data set also support this tree as the best.

In the tree shown in Fig. 9, the *ananassae* subgroup branched off first followed by the *montium* subgroup. The position of the *montium* subgroup is supported by rather low probabilities in the tree shown in Fig. 9, but is supported well in the ML trees of the *Gpdh* and combined data set. This branching order is consistent with the rDNA analysis (Pélandakis and Solignac, 1993; Fig. 1A) and with all of the analyses based on the *Gpdh* and the combined data set in the present study. In addition, Bock and Wheeler (1972) and Ashburner *et al.* (1984) suggested on the basis of morphological and chromosomal data that the *ananassae* and *montium* subgroups separated from the others early in the evolution of the group. On the other hand, in the *Amy* multigene analysis (Inomata *et al.,* 1997; Fig. 1B), the *ananassae* subgroup branched off early in the *melanogaster* group, but the *montium* subgroup is closely related to the *takahashii* subgroup with an unreliable bootstrap value (26%).

The relationships of Australasian and Oriental subgroups (the *elegans* and *ficusphila* subgroups) are still obscure. It is reported that the *ficusphila* subgroup is close to the *melanogaster, suzukii,* and *takahashii* subgroups (Ashburner *et al.,* 1984). In addition, Okada (1954) recognized the close relation of the *melanogaster, takahashii,* and *ficusphila* subgroup. Moreover, the resemblance

of the *elegans* and *suzukii* subgroups has been reported (Bock and Wheeler, 1972). In the best tree supported by the *Gpdh* and combined data set (Fig. 9), next to the *montium* subgroup, the *elegans* subgroup branched off followed by the *ficusphila* subgroup. However, the position of *D. ficusphila* was poorly supported. This relationship was supported by several trees (MP and NJ analyses in *Gpdh*), but the probabilities are still not so high. On the other hand, the previous molecular studies suggested that the *elegans* and *ficusphila* subgroups formed a cluster and branched off next to the *montium* subgroup (Pélandakis and Solignac, 1993) or to the *ananassae* and *eugracilis* subgroups (Inomata *et al.,* 1997) with unreliable bootstrap values (Fig. 1). These relationships did not match any trees in the present study. Moreover, Inomata *et al.* (1997) reported that one of the copies of the *Amy* genes in *D. elegans* formed a cluster with *D. takahashii* genes and another was closely related to *D. ficusphila* genes.

Next to the *elegans* and/or *ficusphila* subgroup(s), the African subgroup (the *melanogaster* subgroup) branched off and formed a cluster with the Asian tropic or subtropic subgroup, the *suzukii* and *takahashii* subgroup, in the present analysis. Pélandakis and Solignac (1993) and Ashburner *et al.* (1984) also reached the same conclusion. The close relation of the *melanogaster* subgroup to the *takahashii* subgroup is also suggested by allozyme data (Tsacas and Tscas, 1984).

The phylogenetic relationships in the *takahashii* subgroup are still ambiguous; i.e., the best tree (Fig. 9) supported by the *Gpdh* and combined data set revealed that *D. lutescens* branched off first and then *D. takahashii,* but the position of the latter is supported by low probability. In addition, almost all trees based on different data and various methods in the present study produced different relationships. It would be expected that the averages of the *p*-distances in the comparison within the *takahashii* subgroup (0.025 in the *COI* and *Gpdh*) are lower than those in the comparison within other subgroups (0.072 and 0.037 in the *COI* and *Gpdh,* respectively). In addition, Inaba *et al.* (1993) and Parkash *et al.* (1994) reported that the phylogenetic trees in this subgroup, based on restriction analyses and protein differences, do not always coincide. Moreover, the species belonging to the *takahashii* subgroup are very similar with respect to their external morphology, including the male genitalia (Lemeunier *et al.,* 1986), and produce interspecific hybrids (Kimura, 1982; Inaba *et al.,* 1993).

In the *montium* subgroup, the relationships of *D. biauraria–D. triauraria* (the *auraria* lineage) and *D. rufa–D. lacteicornis* (the *rufa* lineage) are recognized with high probabilities. The previous studies also supported these relationships (Ohnishi *et al.,* 1983; Ohnishi and Watanabe, 1984; Kim *et al.,* 1993). In addition, the *auraria–rufa* lineage formed a cluster with the *D. trapezifrons* and *D. constricta* clade. On the other hand, *D. bocki* (the *kikkawai* species complex) and *D. watanabei* (the *jambulina* species complex) formed a cluster. Kim *et al.* (1993) also supported the relationships among these species complexes in the *montium* subgroup.

ACKNOWLEDGMENTS

We are very grateful to Dr. R. DeSalle from the American Museum of Natural History and anonymous reviewers for their invaluable suggestions. We also thank Dr. D. L. Denlinger from Ohio State University for critically reading the manuscript. This study was supported by a Grand-in-Aid for Overseas Scientific Survey from the Ministry of Education, Science, Sports, and Culture of Japan (No. 11691161) and Research Fellowships of the Japan Society for the Promotion of Science for Young Scientists.

REFERENCES

Ashburner, M., Bodmer, M., and Lemeunier, F. (1984). On the evolutionary relationships of *Drosophila melanogaster. Dev. Genet.* **4:** 295–312.

Avise, J. C. (1994). "Molecular Markers, Natural History, and Evolution," Chapman and Hall, New York.

Baker, R. H., and DeSalle, R. (1997). Multiple sources of character information and the phylogeny of Hawaiian Drosophilids. *Syst. Biol.* **46:** 654–673.

Barrio, E., and Ayala, F. J. (1997). Evolution of *Drosophila obscura* species group inferred from the *Gpdh* and *Sod* genes. *Mol. Phylogenet. Evol.* **7:** 79–93.

Beckenbach, A. T., Wei, Y. W., and Liu, H. (1993). Relationships in the *Drosophila obscura* species group, inferred from mitochondrial cytochrome oxidase II sequences. *Mol. Biol. Evol.* **10:** 619–634.

Beverley, S. M., and Wilson, A. C. (1984). Molecular evolution in *Drosophila* and the higher Diptera. II. A time scale for fly evolution. *J. Mol. Evol.* **21:** 1–13.

Bock, I. R. (1980). Current status of the *Drosophila melanogaster* species group (Diptera). *Syst. Entomol.* **5:** 341–356.

Bock, I. R., and Wheeler, M. R. (1972). The *Drosophila melanogaster* species group. *Univ. Texas Publ.* **7213:** 1–102.

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.

Brown, W. M., George, M., Jr., and Wilson, A. C. (1979). Rapid evolution of animal mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* **76:** 1967–1971.

Brown, W. M., Prager, E. M., Wang, A., and Wilson, A. C. (1982). Mitochondrial DNA sequences of primates: Tempo and mode of evolution. *J. Mol. Evol.* **18:** 225–239.

Collins, T. M., Wimberger, P. H., and Naylor, G. J. P. (1994). Compositional bias, character-state bias, and character-state reconstruction using parsimony. *Syst. Biol.* **43:** 482–496.

de Bruijn, M. H. L. (1983). *Drosophila melanogaster* mitochondrial DNA, a novel organization and genetic code. *Nature* **304:** 234–241.

DeSalle, R., Freedman, T., Prager, E. M., and Wilson, A. C. (1987). Tempo and mode of sequence evolution in mitochondrial DNA of Hawaiian *Drosophila. J. Mol. Evol.* **26:** 157–164.

Fang, Q. Q., Cho, S., Regier, J. C., Mitter, C., Matthews, M., Poole, R. W., Friedlander, T. P., and Zhao, S. (1997). A new nuclear gene for insect phylogenetics: *dopa decarboxylase* is informative of relationships within Heliothinae (Lepidoptera: Noctuidae). *Syst. Biol.* **46:** 269–283.

Farris, J. S., Ka¨llersjo¨, M., Kluge, A., and Bult, C. (1994). Testing significance of congruence. *Cladistics* **10:** 315–320.

Farris, J. S., Ka¨llersjo¨, M., Kluge, A., and Bult, C. (1995). Constructing a significance test for incongruence. *Syst. Biol.* **44:** 570–572.

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: A maximum likelihood approach. *J. Mol. Evol.* **17:** 368–374.

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

Friedlander, T. P., Regier, J. C., and Mitter, C. (1992). Nuclear gene sequences for higher-level phylogenetic analysis: 14 promising candidates. *Syst. Biol.* **41:** 483–489.

Friedlander, T. P., Regier, J. C., and Mitter, C. (1994). Phylogenetic information content of five nuclear gene sequences in animals: Initial assessment of character sets from concordance and divergence studies. *Syst. Biol.* **43:** 511–525.

Friedlander, T. P., Regier, J. C., Mitter, C., and Wagner, D. L. (1996). A nuclear gene for higher level phylogenetics: Phosphoenolpyruvate carboxykinase tracks Mesozoic-age divergences within Lepidoptera (Insecta). *Mol. Biol. Evol.* **13:** 594–604.

Gleason, J. M., Caccone, A., Moriyama, E. N., White, K. P., and Powell, J. R. (1997). Mitochondrial DNA phylogenies for the *Drosophila obscura* group. *Evolution* **51:** 433–440.

Gleason, J. M., Griffith, E. C., and Powell, J. R. (1998). A molecular phylogeny of the *Drosophila willistoni* group: Conflicts between species concepts? *Evolution* **52:** 1093–1103.

Goto, S. G., Kitamura, H. W., and Kimura, M. T. (2000). Phylogenetic relationships and climatic adaptations in the *Drosophila takahashii* and *montium* species subgroups. *Mol. Phylogenet. Evol.* **15:** 147–156.

Goto, S. G., Yoshida, T., Beppu, K., and Kimura, M. T. (1999). Evolution of overwintering strategies in Eurasian species of the *Drosophila obscura* species group. *Biol. J. Linnean Soc.* **68:** 429– 441.

Gu, X., Fu, Y.-X., and Li, W.-H. (1995). Maximum likelihood estimation of the heterogeneity of substitution rate among nucleotide sites. *Mol. Biol. Evol.* **12:** 546–557.

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

Hillis, D. M. (1987). Molecular versus morphological approaches to systematics. *Annu. Rev. Ecol. Syst.* **18:** 23–42.

Hinton, C. W. (1984). Morphogenetically specific mutability in *Drosophila ananassae. Genetics* **106:** 631–653.

Hsu, T. C. (1949). The external genital apparatus of male Drosophilidae in relation to systematics. *Univ. Texas Publ.* **4920:** 80–142.

Inaba, A., Fukatami, A., and Aotsuka, T. (1993). Phylogenetic relationship among species of the *Drosophila takahashii* species subgroup. *Zool. Sci. (Suppl)* **10:** 178.

Inomata, N., Tachida, H., and Yamazaki, T. (1997). Molecular evolution of the *Amy* multigenes in the subgenus *Sophophora* of *Drosophila. Mol. Biol. Evol.* **14:** 942–950.

Irwin, D. M., Kocher, T. D., and Wilson, A. C. (1991). Evolution of the cytochrome *b* gene of mammals. *J. Mol. Evol.* **32:** 128–144.

Jin, L., and Nei, M. (1990). Limitations of the evolutionary parsimony method of phylogenetic analysis. *Mol. Biol. Evol.* **7:** 82–102.

Johnson, K. P., and Clayton, D. H. (2000). Nuclear and mitochondrial genes contain similar phylogenetic signal for pigeons and doves (Aves: Columbiformes). *Mol. Phylogenet. Evol.* **14:** 141–151.

Kim, B. K., Aotsuka, T., and Kitagawa, O. (1993). Evolutionary genetics of the *Drosophila montium* subgroup. II. Mitochondrial DNA variation. *Zool. Sci.* **10:** 991–996.

Kimura, M. T. (1982). Inheritance of cold hardiness and sugar contents in two closely related species, *Drosophila takahashii* and *Drosophila lutescens. Jpn. J. Genet.* **57:** 575–580.

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

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

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

Lake, J. A. (1994). Reconstructing evolutionary trees from DNA and protein sequences: Paralinear distances. *Proc. Natl. Acad. Sci. USA* **91:** 1455–1459.

Lemeunier, F., David, J. R., Tsacas, L., and Ashburner, M. (1986). The *melanogaster* species group. *In* "The Genetics and Biology of *Drosophila*" (M. Ashburner, H. L. Carson, and J. N. Thompson, Jr., Eds.), Vol. 3e, pp. 147–256, Academic Press, London.

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

Meyer, A. (1993). Evolution of mitochondrial DNA in fishes. *In* "Molecular Biology Frontiers, Biochemistry and Molecular Biology of Fishes" (P. W. Hochachka and T. P. Mommsen, Eds.), Vol. 2, pp. 1–38. Elsevier, Amsterdam.

Mickevitch, M. F., and Farris, J. S. (1981). The implication of congruence in *Menidia. Syst. Zool.* **30:** 351–370.

Miyamoto, M. M., and Fitch, W. M. (1995). Testing species phylogenies and phylogenetic methods with congruence. *Syst. Biol.* **44:** 64–76.

Miyata, T., Hayashida, H., Kikuno, R., Hasegawa, M., Kobayashi, M., and Koike, K. (1982). Molecular clock of silent substitution: At least six-fold preponderance of silent changes in mitochondrial genes over those in nuclear genes. *J. Mol. Evol.* **19:** 28–35.

Moriyama, E. N., and Powell, J. R. (1997). Synonymous substitution rates in *Drosophila:* Mitochondrial versus nuclear genes. *J. Mol. Evol.* **45:** 378–391.

Murrell, A., Campbell, N. J. H., and Barker, S. C. (2000). Phylogenetic analyses of the Rhipicephaline ticks indicate that the genus *Rhipicephalus* is paraphyletic. *Mol. Phylogenet. Evol.* **16:** 1–7.

Nei, M. (1987). "Molecular Evolutionary Genetics," Columbia Univ. Press, New York.

Nikolaidis, N., and Scouras, Z. G. (1996). The *Drosophila montium* subgroup species. Phylogenetic relationships based on mitochondrial DNA analysis. *Genome* **39:** 874–883.

O'Grandy, P. M., Clark, J. B., and Kidwell, M. G. (1998). Phylogeny of the *Drosophila saltans* species group based on combined analysis of nuclear and mitochondrial DNA sequences. *Mol. Biol. Evol.* **15:** 656–664.

Ohnishi, S., Kim, K.-W., and Watanabe, T. K. (1983). Biochemical phylogeny of *Drosophila montium* species subgroup. *Jpn. J. Genet.* **58:** 141–151.

Ohnishi, S., and Watanabe, T. K. (1984). Systematics of the *Drosophila montium* species subgroup: a biochemical approach. *Zool. Sci.* **1:** 801–807.

Okada, T. (1954). Comparative morphology of drosophilid flies. I. Phallic organs of the *melanogaster* group. *Kontyuˆ* **22:** 36–46.

Okada, T. (1984). New or little known species of *Drosophila* (*Lordiphosa*) with taximetrical analyses (Diptera, Drosophilidae). *Kontyuˆ* **52:** 565–575.

Palumbi, S. R., and Baker, C. S. (1994). Contrasting population structure from nuclear intron sequences and mtDNA of humpback whales. *Mol. Biol. Evol.* **11:** 426–435.

Parkash, R., Jyoutsna, J., and Vandna, V. (1994). Allozyme phylogeny of five species of *takahashii* species subgroup of *Drosophila. Korean J. Genet.* **16:** 187–196.

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

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

Prychitko, T. M., and Moore, W. S. (1997). The utility of DNA sequence of an intron from the b-fibrinogen gene in phylogenetic analysis of wood peckers (Aves: Picidae). *Mol. Phylogenet. Evol.* **8:** 193–204.

Reed, D. S., and Gibson, J. B. (1994). Molecular heterogeneity of naturally occurring *sn*-glycerol-3-phosphate dehydrogenase lowactivity variants in *Drosophila melanogaster. Biochem. Genet.* **32:** 161–179.

Rodrı´guez-Trelles, F., Tarrı´o, R., and Ayala, F. J. (1999). Molecular evolution and phylogeny of the *Drosophila saltans* species group inferred from the *Xdh* gene. *Mol. Phylogenet. Evol.* **13:** 110–121.

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

Russo, C. A. M., Takezaki, N., and Nei, M. (1996). Efficiencies of different genes and different tree-building methods in recovering a known vertebrate phylogeny. *Mol. Biol. Evol.* **13:** 525–536.

Saccone, C., Lanave, C., and Pesole, G. (1993). Time and biosequences. *J. Mol. Evol.* **37:** 154–159.

Saccone, C., Pesole, G., and Preparata, G. (1989). DNA microenvironments and the molecular clock. *J. Mol. Evol.* **29:** 407–411.

Saitou, N., and Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4:** 406–425.

Satta, Y., Ishiwa, H., and Chigusa, S. I. (1987). Analysis of nucleotide substitutions of mitochondrial DNAs in *Drosophila melanogaster* and its sibling species. *Mol. Biol. Evol.* **4:** 638–650.

Satta, Y., and Takahata, N. (1990). Evolution of *Drosophila* mitochondrial DNA and the history of the *melanogaster* subgroup. *Proc. Natl. Acad. Sci. USA* **87:** 9558–9562.

Slade, R. W., Moritz, C., and Heideman, A. (1994). Multiple nucleargene phylogenies: Application to pinnipeds and comparison with a mitochondrial DNA gene phylogeny. *Mol. Biol. Evol.* **11:** 341–356.

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:** 749–759.

Strimmer, K., Goldman, N., and von Haeseler, A. (1997). Bayesian probabilities and quartet puzzling. *Mol. Biol. Evol.* **14:** 210–211.

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

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

Sullivan, J., Holsinger, K. E., and Simon, C. (1996). The effect of topology on estimates of among-site rate variation. *J. Mol. Evol.* **42:** 308–312.

Swofford, D. L. (2000). PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, MA.

Swofford, D. L., and Olsen, G. J. (1990). Phylogeny reconstruction. *In* "Molecular Systematics" (D. M. Hills and C. Moritz, Eds.), pp. 411–501, Sinauer Associates, Sunderland, MA.

Tamura, K. (1992). The rate and pattern of nucleotide substitution in *Drosophila* mitochondrial DNA. *Mol. Biol. Evol.* **9:** 814–825.

Tamura, K., and 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:** 512–526.

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

Templeton, A. R. (1983). Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and apes. *Evolution* **37:** 221–244.

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

Toda, M. J. (1991). Drosophilidae (Diptera) in Myanmar (Burma) VII. The *Drosophila melanogaster* species-group, excepting the *D. montium* species-subgroup. *Oriental Insects* **25:** 69–94.

Tsacas, S. C., and Tscas, L. (1984). A phenetic tree of eighteen species of the *melanogaster* group of *Drosophila* using allozyme data as compared with classification based on other criteria. *Genetica* **64:** 139–144.

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

Yang, Z. (1998). On the best evolutionary rate for phylogenetic analysis. *Syst. Biol.* **47:** 125–133.

Zardoya, R., and Meyer, A. (1996). Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates. *Mol. Biol. Evol.* **13:** 933–942.

TABLE 1

Experimental Species, Collection Localities, and Accession Numbers for *Gpdh* and *COI* Genes

Subgenus Species group Species subgroup Species	Collection locality	Accession No.		
		Gpdh[a]		*COI*
Sophophora				
obscura group				
D. bifasciata		U47883		U51611
melanogaster group *melanogaster* subgroup				
D. melanogaster Meigen		X61223		U37541
D. simulans Sturtevant		AF085163		M57909
D. teissieri Tsacas		U47809		U51618
takahashii subgroup				
D. lutescens Okada	Previous study[b]	AB027276	AB027281	AB027267
D. trilutea Bock and Wheeler	Previous study[b]	AB027270	AB027286	AB027261
D. prostipennis Lin	Previous study[b]	AB027275	AB027282	AB027266
D. takahashii Sturtevant	Previous study[b]	AB027273	AB027284	AB027264
montium subgroup				
D. biauraria Bock and Wheeler	Previous study[b]	AB027278	AB027279	AB027259
D. triaurarira Bock and Wheeler	Previous study[b]	AB027271	AB027287	AB027262
D. rufa Kikkawa and Peng	Previous study[b]	AB027274	AB027283	AB027265
D. trapezifrons Okada	Previous study[b]	AB027272	AB027288	AB027263
D. watanabei Gupta	Previous study[b]	AB027269	AB027285	AB027260
D. constricta Chen, Shao and Fan	Previous study[b]	AB027277	AB027280	AB027268
D. lacteicornis Okada	Iriomote, Japan	AB032134	AB032142	AB032126
D. bocki Baimai	Iriomote, Japan	AB032135	AB032143	AB032127
suzukii subgroup				
D. suzukii (Matsumura)	Tokyo, Japan	AB032136	AB032144	AB032128
elegans subgroup				
D. gunungcola Sultana, Kimura and Toda	Sukarami, Indonesia	AB032137	AB032145	AB032129
D. elegans Bock and Wheeler	Hongkong, China	AB032138	AB032146	AB032130
ananassae subgroup				
D. bipectinata Duda	Iriomote, Japan	AB032139	AB032147	AB032131
D. ananassae Doleschall	Old laboratory strain[c]	AB032140	AB032148	AB032132
ficusphila subgroup				
D. ficusphila Kikkawa and Peng	Iriomote, Japan	AB032141	AB032149	AB032133

[a] Two accession numbers for *Gpdh* indicate exons 3 and 4, respectively.

[b] Goto *et al.* (2000).

[c] *ca;px* strain in Hinton (1984).

TABLE 2

Primers Used for the Amplification and Sequencing of *Gpdh* and *COI*

Primers	Sequence (5' > 3')
For *Gpdh*	
GNL-mel	GTG GTG CCC CAC CAG TTC AT
GNR-mel	GGC TTG AGC TGA TTT GTG CA
L4BN	CCA TGY GCT GTC TTG ATG GG
R4M	ACA GCC GCC TTG GTG TTG TCG CC
Gpdh-F	TCA AGC TCG GCG ACA ACA
Gpdh-R	CCC ATC AAC ACG GCG CAT GG
GNL	CCC GAC CTG GTT GAG GCT AGC CAA GAA TGC
GNR	ACA TAT GCT CAG GGT GCT AGC GTA TGC A
For *COI*	
F-COI	CCA GCT GGA GGA GGA GAT CC
R-COI	CCA GTA AAT AAT GGG TAT CAG TG

TABLE 3

Percentage of Variable and Phylogenetically Informative Sites of *Gpdh* and *COI*

	Gpdh		*COI*	
	Ingroup	Including outgroup[a]	Ingroup	Including outgroup[a]
Total positions	430		407	
Variable	28.6 (123)	31.2 (134)	28.0 (114)	28.5 (116)
First	2.8 (12)	3.0 (13)	3.4 (14)	3.7 (15)
Second	0.0 (0)	0.0 (0)	0.2 (1)	0.2 (1)
Third	25.8 (111)	28.1 (121)	24.3 (99)	24.6 (100)
Informative	21.2 (91)	21.9 (94)	20.6 (84)	21.6 (88)
First	1.2 (5)	1.6 (7)	2.5 (10)	2.5 (10)
Second	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)
Third	20.0 (86)	20.2 (87)	18.2 (74)	19.2 (78)

Note. Values in parentheses indicate the number of positions. [a] The outgroup comparison includes *D. bifasciata* sequence.

TABLE 4

Base Composition and Bias of All, Variable, and Phylogenetically Informative Sites in *Gpdh* and *COI*

	All		Variable		Informative	
	Mean	SD	Mean	SD	Mean	SD
			Gpdh			
A	0.201	0.012	0.070	0.039	0.083	0.049
C	0.271	0.023	0.481	0.073	0.451	0.092
G	0.309	0.010	0.275	0.031	0.237	0.041
T	0.219	0.020	0.174	0.066	0.229	0.084
Bias[a]	0.106*		0.341*		0.268*	
			COI			
A	0.302	0.008	0.369	0.027	0.355	0.033
C	0.134	0.007	0.072	0.023	0.090	0.029
G	0.163	0.005	0.020	0.017	0.021	0.019
T	0.402	0.010	0.538	0.035	0.534	0.044
Bias[a]	0.271*		0.543*		0.519*	

[a] The bias is calculated using the formula of Irwin *et al.* (1991). *χ^2 test ($P < 0.01$).

TABLE 5

The Parameters Estimated from the Γ-Discrete ML Analyses

	Gpdh	*COI*	TE[a]
α_b	2.21	0.54	1.94
Ts/Tv[c]	3.69	29.69	1.44
θ_d	0.63	0.66	0.63
ρ_e	0.75	0.88	0.76

[a] Total evidence tree. [b] Shape parameter. [c] Ratio of transition to transversion. [d] Fraction of invariable sites. [e] Total rate heterogeneity (Gu *et al.,* 1995).

TABLE 6

Results of Partitioned Bremer Support (PBS) Analysis

Node[a]	*Gpdh*	*COI*	TE[b]
1	8.0	-2.0	6.0
2	5.0	-2.0	3.0
3	4.0	2.0	6.0
4	1.5	0.5	2.0
5	1.0	1.0	2.0
6	3.0	4.0	7.0
7	4.0	2.0	6.0
8	7.0	2.0	9.0
9	11.0	3.0	14.0
10	0.0	1.0	1.0
11	0.5	0.5	1.0
12	12.0	3.0	15.0
13	4.7	3.3	8.0
14	1.5	2.5	4.0
15	3.0	2.0	5.0
16	1.0	0.0	1.0
17	2.0	3.0	5.0
18	3.0	2.0	5.0
19	1.0	4.0	5.0
Total	73.2 (69.7%)	31.8 (30.3%)	105.0

[a] See Fig. 9. [b] Total evidence tree.

TABLE 7

The Length and Results of the Templeton Parsimony Test for Each of the Phylogenetic Hypotheses Inferred from All Different Analyses

	Tree	Length	Difference	SD	RC[a]	Significantly worse than best tree?
Gpdh	Fig. 5A	322	3	1.7	0.424	No
	Fig. 5B	334	15	4.7	0.394	Yes
	Fig. 6A	323	4	4.5	0.421	No
	Fig. 6B	322	3	4.4	0.424	No
	Fig. 6C	323	4	4.0	0.421	No
	Fig. 7A	378	59	8.5	0.302	Yes
	Fig. 7B	389	70	10.4	0.283	Yes
	Fig. 7C	380	61	9.0	0.299	Yes
	Fig. 8A	370	51	7.9	0.317	Yes
	Fig. 8B	357	38	6.7	0.343	Yes
	Fig. 8C	383	64	9.9	0.293	Yes
	Fig. 9	319	—	—	0.431	Best
	Fig. 10	327	8	3.1	0.411	Yes
	Fig. 11	323	4	3.7	0.421	No
	Fig. 12	324	5	2.6	0.419	No
COI	Fig. 5A	332	17	6.8	0.245	Yes
	Fig. 5B	339	24	7.0	0.231	Yes
	Fig. 6A	334	19	6.9	0.241	Yes
	Fig. 6B	335	20	7.0	0.239	Yes
	Fig. 6C	332	17	7.0	0.245	Yes
	Fig. 7A	317	2	3.5	0.276	No
	Fig. 7B	341	26	7.2	0.227	Yes
	Fig. 7C	333	18	5.9	0.243	Yes
	Fig. 8A	315	—	—	0.280	Best
	Fig. 8B	318	3	6.1	0.274	No
	Fig. 8C	322	7	5.7	0.265	No
	Fig. 9	325	10	5.8	0.259	No
	Fig. 10	333	18	6.3	0.243	Yes

	Fig. 11	331	16	6.3	0.247	Yes
	Fig. 12	328	13	5.9	0.253	Yes
Gpdh + COI	Fig. 5A	654	10	3.1	0.334	Yes
	Fig. 5B	673	29	6.5	0.313	Yes
	Fig. 6A	657	13	6.1	0.331	Yes
	Fig. 6B	657	13	5.9	0.331	Yes
	Fig. 6C	655	11	5.9	0.333	No
	Fig. 7A	695	51	10.3	0.291	Yes
	Fig. 7B	730	86	12.1	0.257	Yes
	Fig. 7C	713	69	10.7	0.273	Yes
	Fig. 8A	685	41	10.1	0.301	Yes
	Fig. 8B	675	31	9.2	0.311	Yes
	Fig. 8C	705	61	11.9	0.281	Yes
	Fig. 9	644	—	—	0.346	Best
	Fig. 10	660	16	4.7	0.327	Yes
	Fig. 11	654	10	5.1	0.334	Yes
	Fig. 12	652	8	3.7	0.336	Yes

[a] Rescaled consistency index.

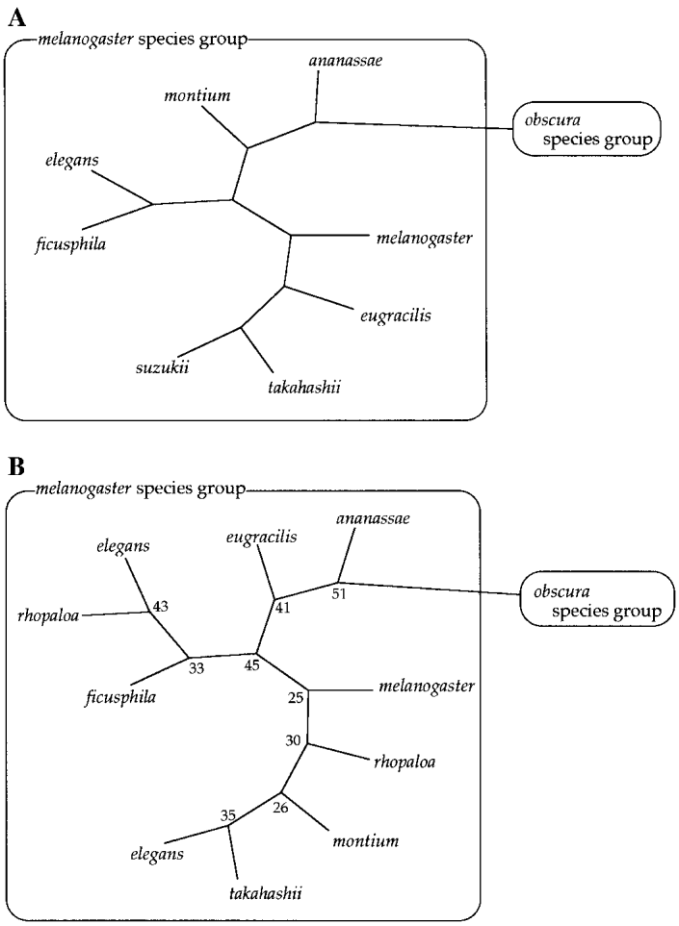

FIG. 1. A diagram of the phylogenetic hypotheses inferred from rDNA (Pélandakis and Solignac, 1993; A) and from *Amy* multigenes (Inomata *et al.,* 1997; B). These trees are unrooted. Some sequences within species even clustered with those of other species in *Amy.* Bootstrap values (percentage of 1000 pseudoreplicates) are shown at each node of the tree from *Amy* (B). Pélandakis and Solignac (1993) did not present the probabilities.

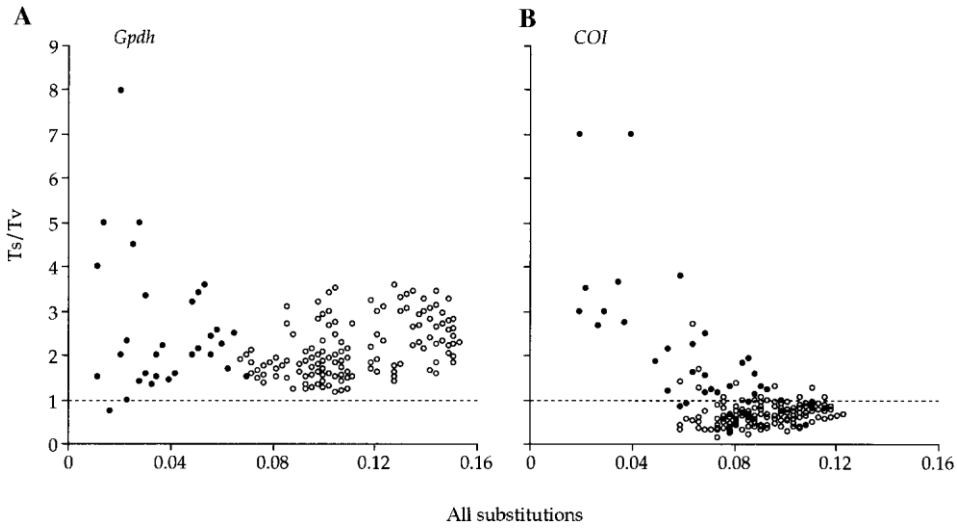

FIG. 2. Scatter plots of *p*-distances between transition to transversion ratios (Ts/Tv) versus all substitutions in the *Gpdh* (A) and *COI* (B) genes. (●) Comparison within the subgroup; (○) comparison between subgroups.

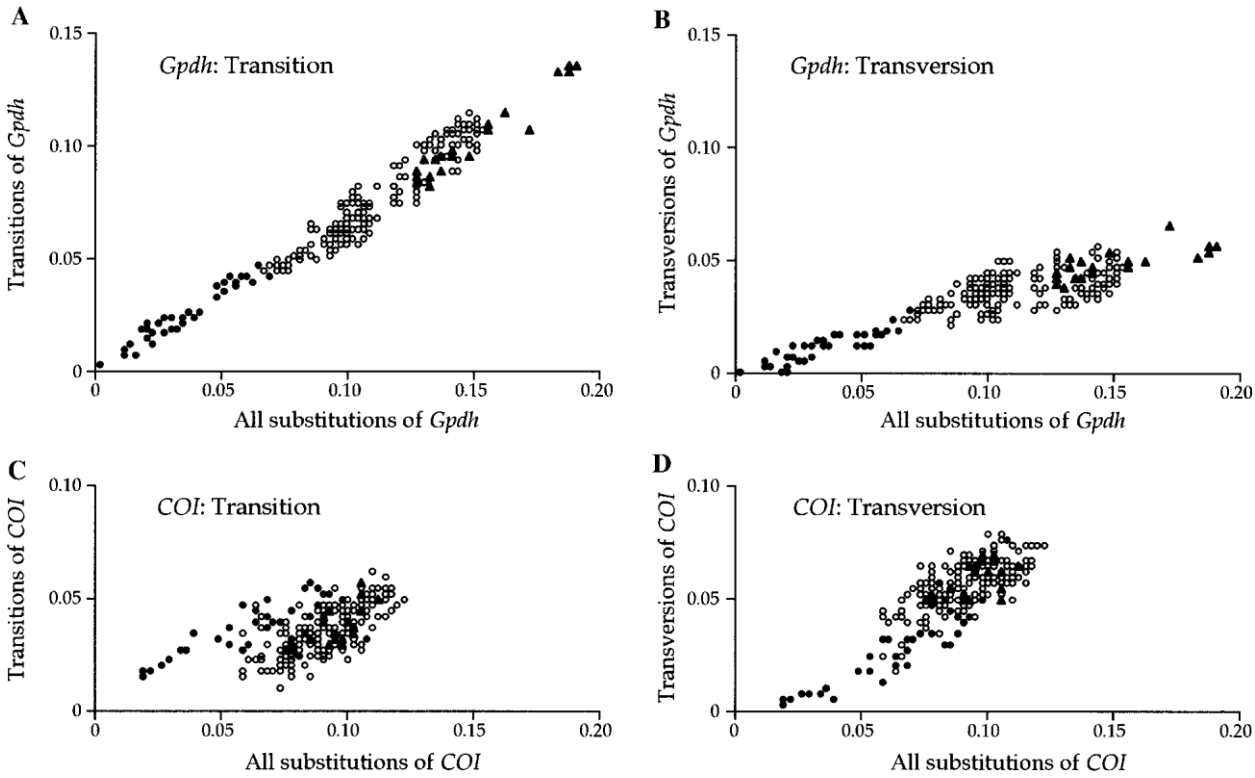

FIG. 3. Scatter plots of *p*-distances between transitions (A and C) or transversions (B and D) versus all substitutions for all pairwise comparisons of the *Gpdh* (A and B) and *COI* (C and D) genes. (●) Comparison within the subgroup; (○) comparison between subgroups. The comparisons between species belonging to the *melanogaster* group and *D. bifasciata* (the *obscura* group) are also indicated (▲).

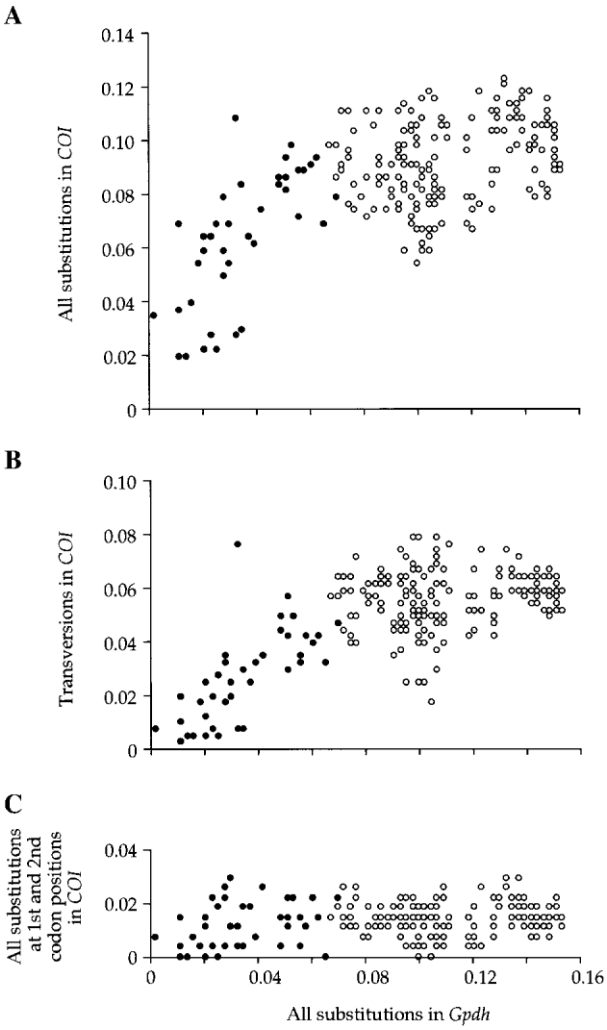

FIG. 4. Scatter plots of *p*-distances between all (A), transversional (B), and first- plus second-codon positional substitutions (C) of *COI* versus all substitutions of *Gpdh.* (●) Comparison within the subgroup; (○) comparison between subgroups.

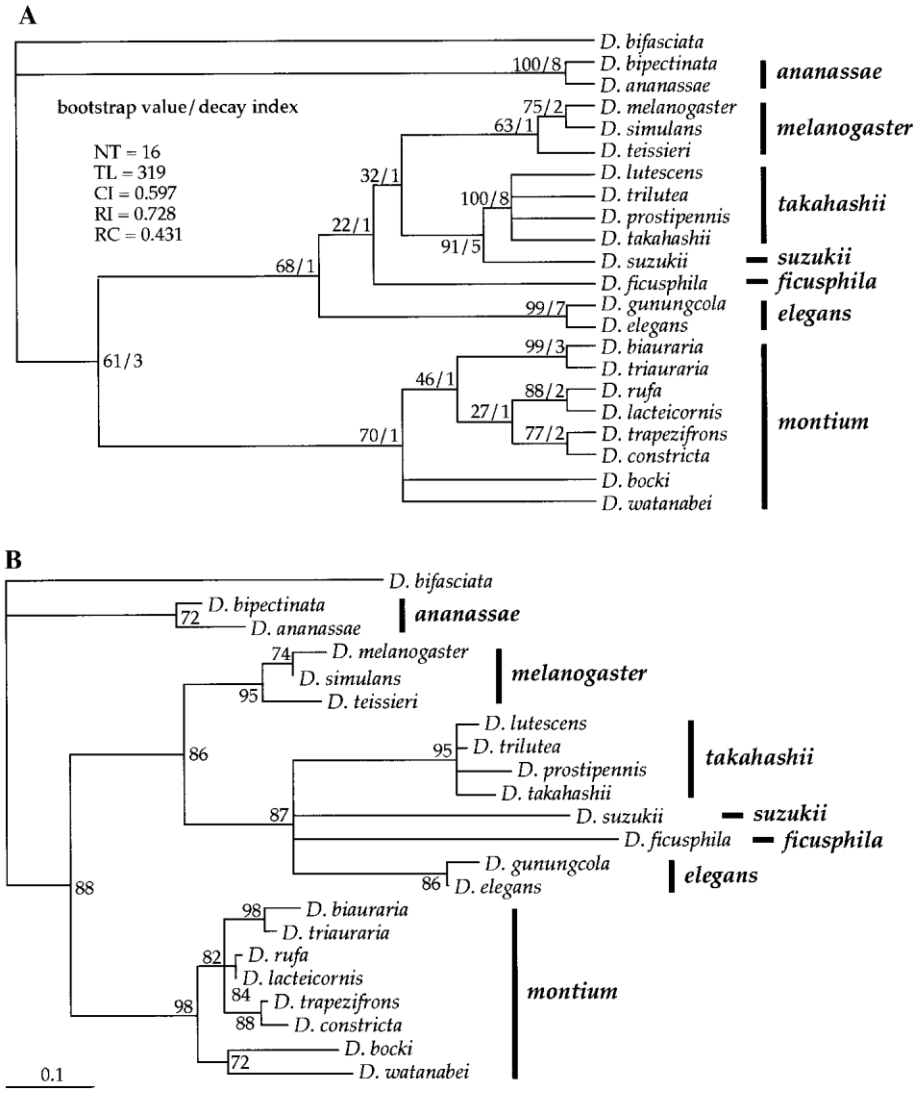

FIG. 5. MP (A) and ML (B) trees based on *Gpdh.* Bootstrap values (percentage of 1000 pseudoreplicates) and decay indexes, and quartet-puzzling scores (percentage of 1000 steps) are shown at each node of the MP and ML trees, respectively. In the MP tree, the number of most-parsimonious trees recovered (NT), total tree length (TL), consistency index (CI), retention index (RI), and rescaled consistency index (RC) are also shown. Branch lengths are proportional to the scale given in substitutions per nucleotide for the ML tree. Subgroups are given at the right.

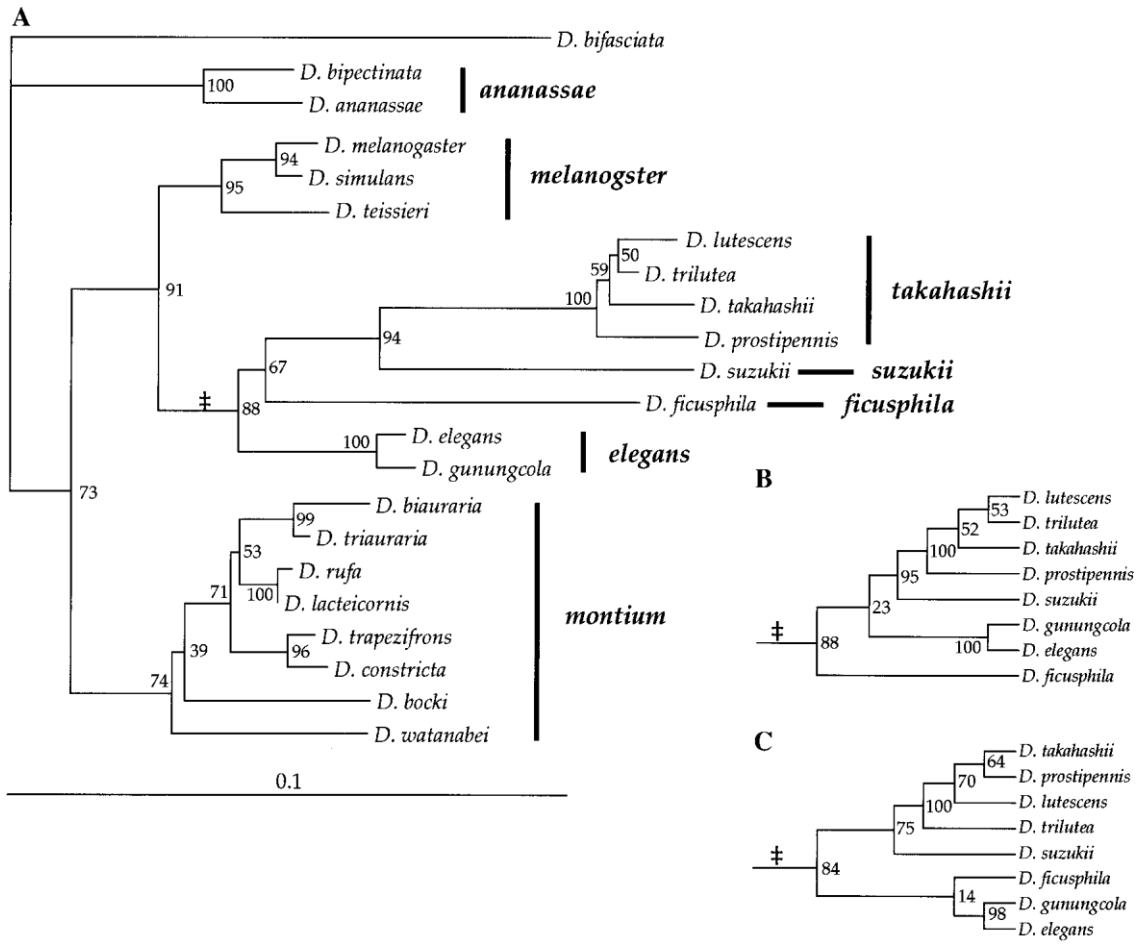

FIG. 6. NJ trees based on *Gpdh* with the K80 (A), Log Det/Paralinear (B), and ML-estimated parameter (C) model. These trees revealed nearly identical topology except the marked (‡) node (only details are shown in B and C). Bootstrap values (percentage of 1000 pseudoreplicates) are shown at each node. Branch lengths are proportional to the scale given in substitutions per nucleotide only for A. Subgroups are given at the right.

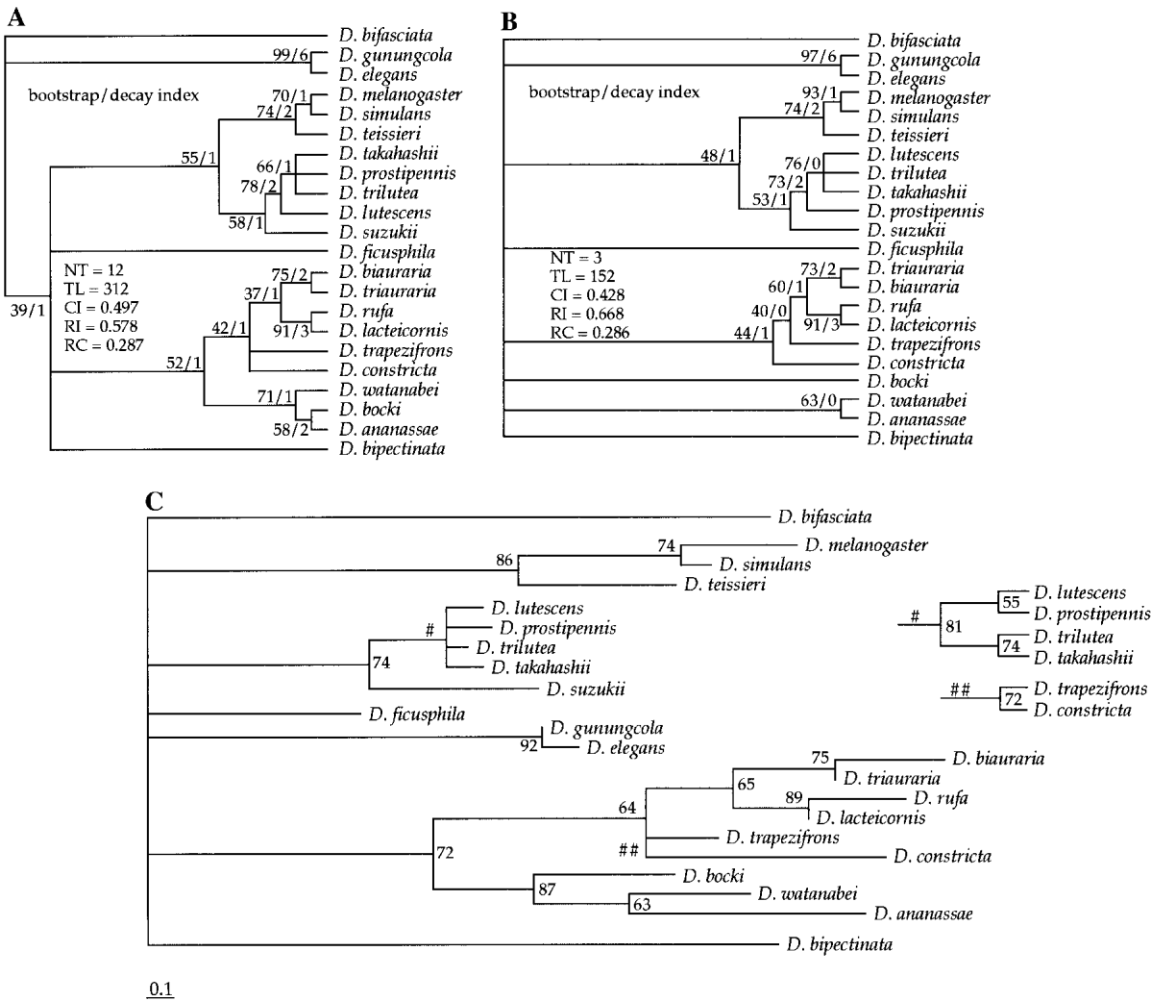

FIG. 7. MP trees based on *COI* using all substitutions (A) and only transversions (B) and ML tree (C). Bootstrap values (percentage of 1000 pseudoreplicates) and decay indexes and quartet-puzzling scores (percentage of 1000 steps) are shown at each node of MP and ML trees, respectively. In the MP trees, number of most-parsimonious trees recovered (NT), total tree length (TL), consistency index (CI), retention index (RI), and rescaled consistency index (RC) are also shown. Branch lengths are proportional to the scale given in substitutions per nucleotide for the ML tree. Details of the marked (# and ##) nodes are shown at the right.

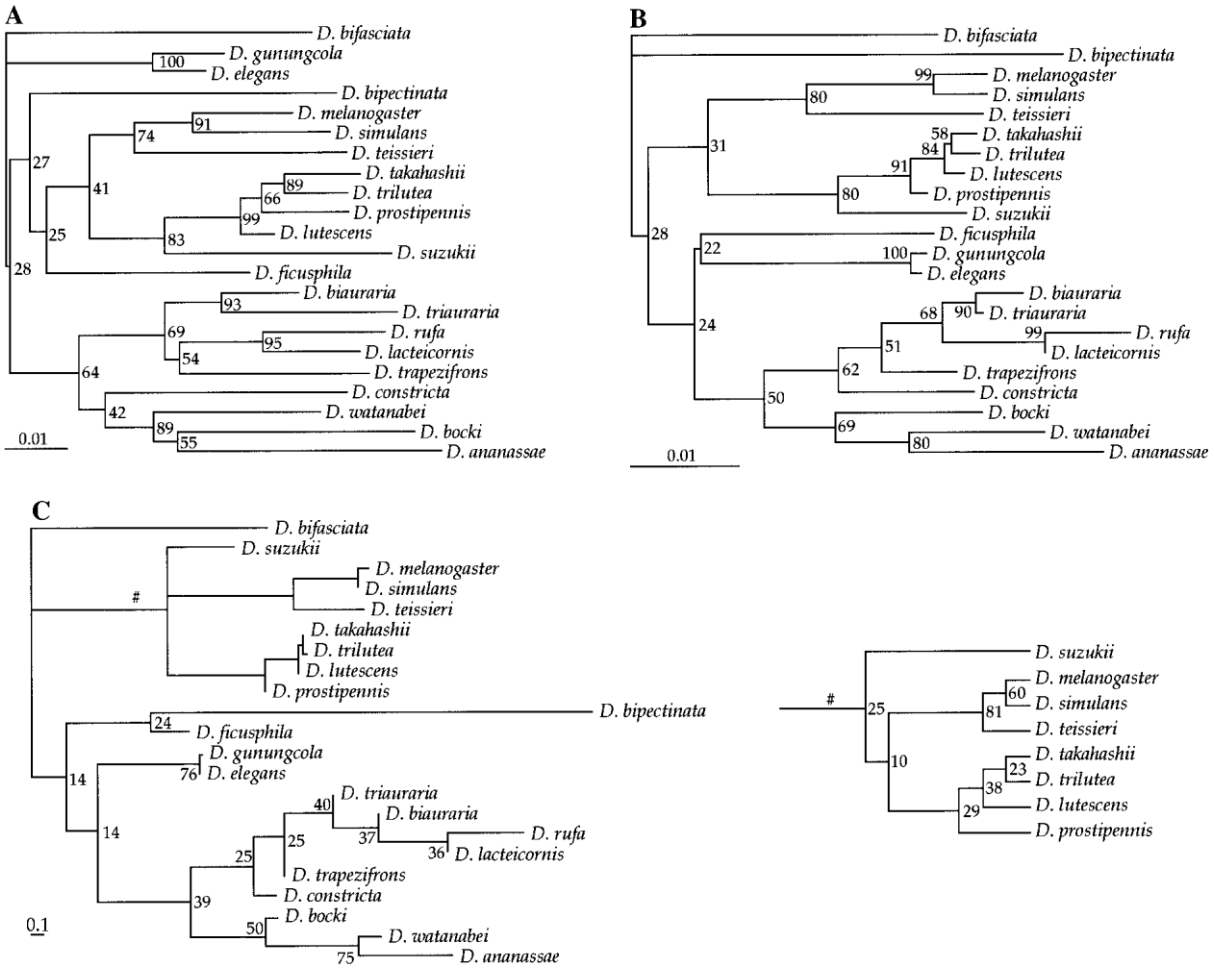

FIG. 8. NJ trees based on *COI* with the K80 model using all substitutions (A) and using only transversions (B) and with the ML-estimated parameter (C). Branch lengths are proportional to the scale given in substitutions per nucleotide. Details of the marked (#) node are shown at the right. Bootstrap values (percentage of 1000 pseudoreplicates) are shown at each node.

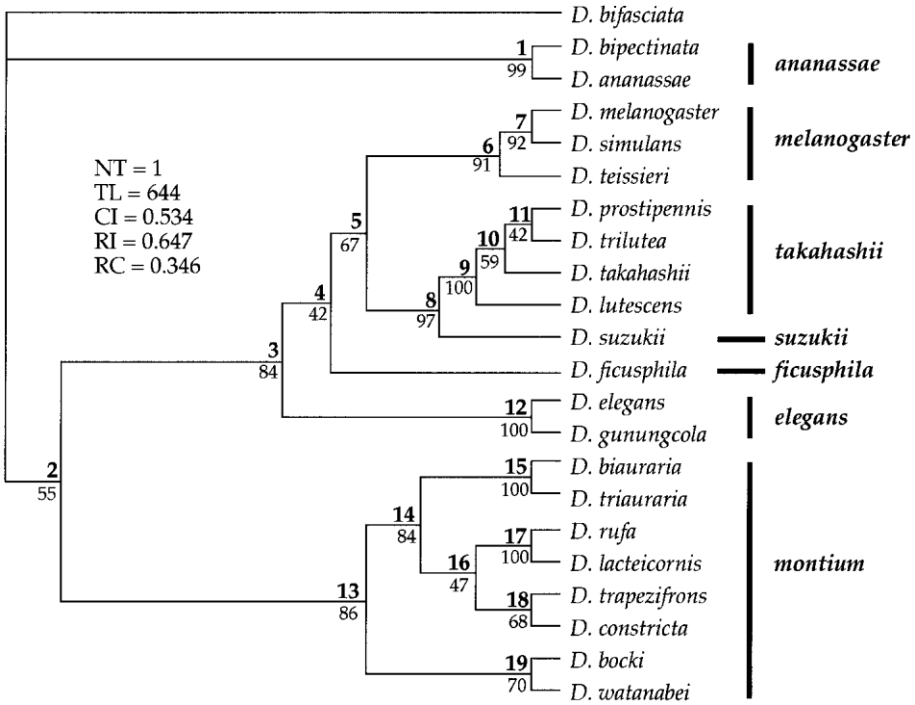

FIG. 9. MP tree based on the combined data set. Number of most-parsimonious trees recovered (NT), total tree length (TL), consistency index (CI), retention index (RI), and rescaled consistency index (RC) are shown. Boldface numbers show branches that were tested for the partitioned Bremer support (see Table 6). Bootstrap values (percentage of 1000 pseudoreplicates) are shown at each node. Subgroups are given at the right.

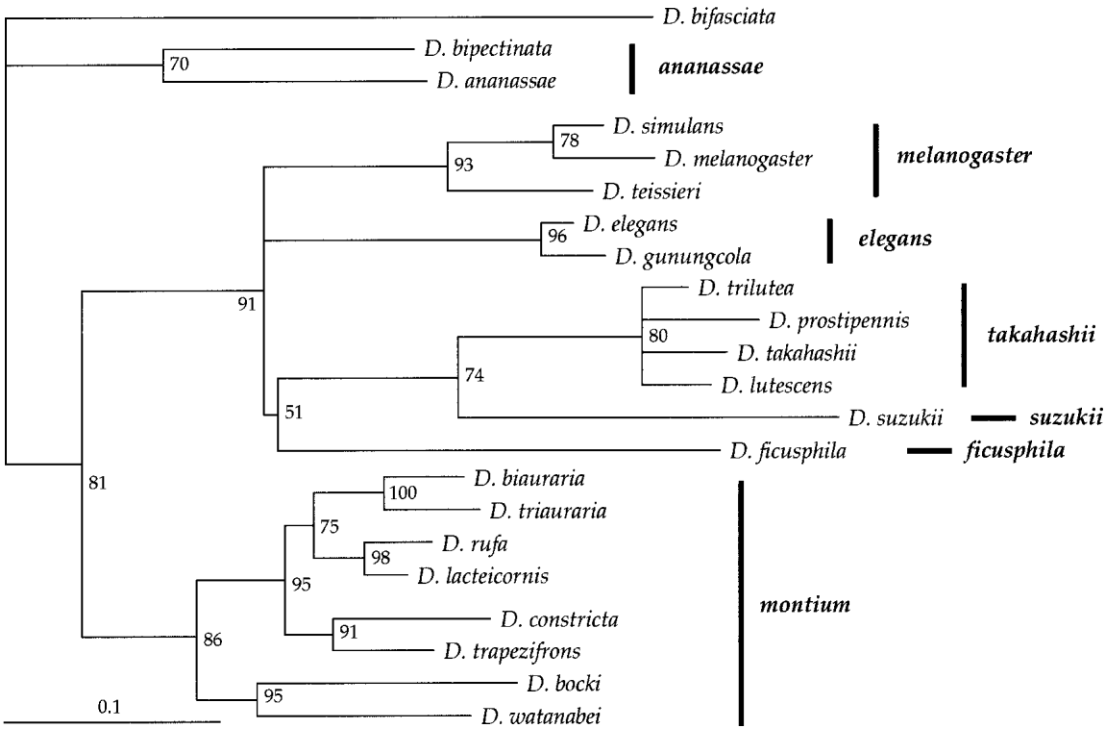

FIG. 10. ML tree based on the combined data set. Quartet-puzzling scores (percentage of 1000 steps) are shown at each node. Branch lengths are proportional to the scale given in substitutions per nucleotide. Subgroups are given at the right.

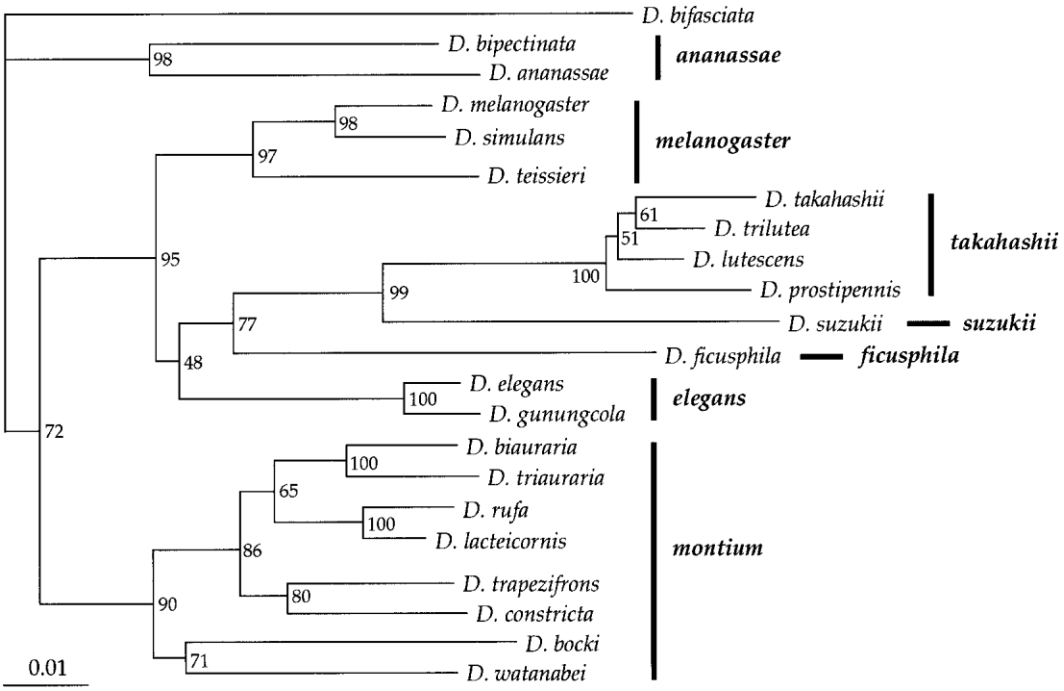

FIG. 11. NJ tree based on the combined data set with the K80 model. Bootstrap values (percentage of 1000 pseudoreplicates) are shown at each node. Branch lengths are proportional to the scale given in substitutions per nucleotide. Subgroups are given at the right.

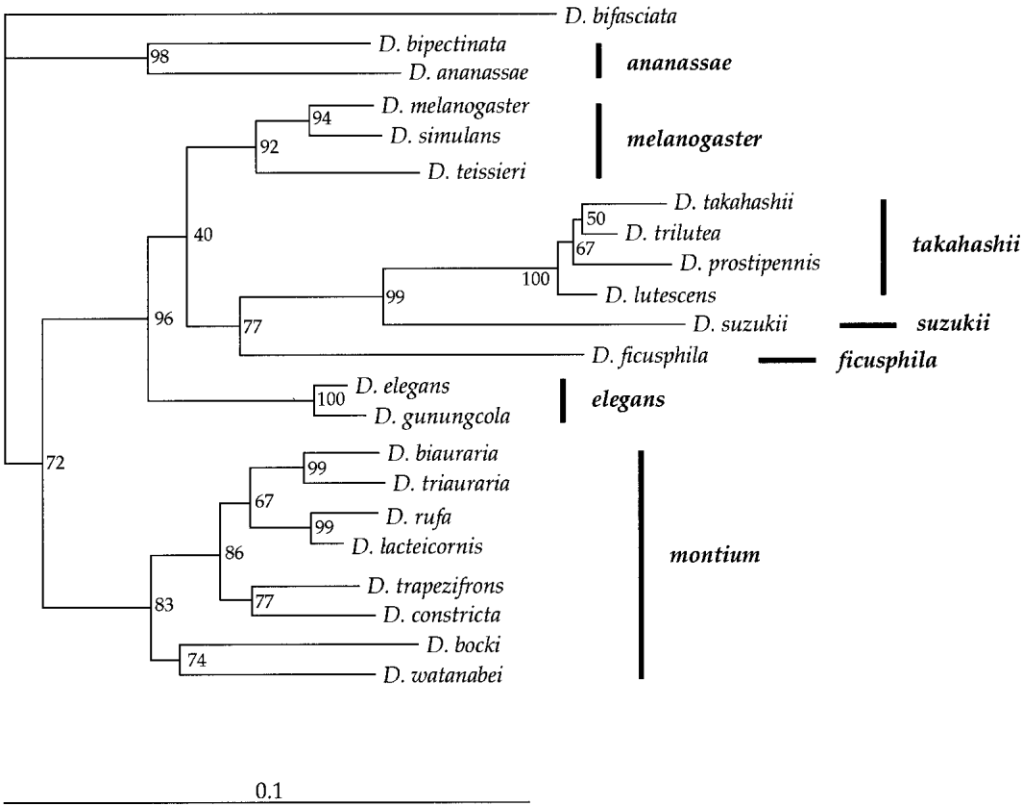

FIG. 12. NJ tree based on the combined data set with the ML-estimated parameter (Table 5). Bootstrap values (percentage of 1000 pseudoreplicates) are shown at each node. Branch lengths are proportional to the scale given in substitutions per nucleotide. Subgroups are given at the right.