Title
On the Evolution of Dopa decarboxylase (Ddc) and Drosophila Systematics

Journal
Journal of Molecular Evolution, 48(4)

ISSN
0022-2844

Authors
Tatarenkov, Andrey
Kwiatowski, Jan
Skarecky, Douglas
et al.

Publication Date
1999-04-01

DOI
10.1007/pl00006489

J Mol Evol (1999) 48:445–462

JOURNAL OF MOLECULAR EVOLUTION

On the Evolution of *Dopa decarboxylase (Ddc)* and *Drosophila* Systematics

Andrey Tatarenkov, Jan Kwiatowski,* Douglas Skarecky, Eladio Barrio,† Francisco J. Ayala

Department of Ecology and Evolutionary Biology, 321 Steinhaus Hall, University of California, Irvine, CA 92697-2525, USA

Received: 2 June 1998 / Accepted: 3 September 1998

Abstract. We have sequenced most of the coding region of the gene *Dopa decarboxylase (Ddc)* in 24 fruitfly species. The *Ddc* gene is quite informative about *Drosophila* phylogeny. Several outstanding issues in *Drosophila* phylogeny are resolved by analysis of the *Ddc* sequences alone or in combination with three other genes, *Sod, Adh,* and *Gpdh.* The three species groups, *melanogaster, obscura,* and *willistoni,* are each monophyletic and all three combined form a monophyletic group, which corresponds to the subgenus *Sophophora.* The *Sophophora* subgenus is the sister group to all other *Drosophila* subgenera (including some named genera, previously considered outside the *Drosophila* genus, namely, *Scaptomyza* and *Zaprionus,* which are therefore downgraded to the category of subgenus). The Hawaiian *Drosophila* and *Scaptomyza* are a monophyletic group, which is the sister clade to the *virilis* and *repleta* groups of the subgenus *Drosophila.* The subgenus *Drosophila* appears to be paraphyletic, although this is not definitely resolved. The two genera *Scaptodrosophila* and *Chymomyza* are older than the genus *Drosophila.* The data favor the hypothesis that *Chymomyza* is older than *Scaptodrosophila,* although this issue is not definitely resolved. Molecular evolution is erratic. The rates of nucleotide substitution in 3rd codon position relative to positions 1 + 2 vary from one species lineage to another and from gene to gene.

Key words: *Dopa decarboxylase* (*Ddc*) — Divergence — *Drosophila — Scaptodrosophila — Chymomyza — Zaprionus — Scaptomyza*

Introduction

The received classification of the Drosophilidae (e.g., Wheeler 1981) is inconsistent with the phylogenetic relationships among the species, whether these are based on morphology (Throckmorton 1975; Grimaldi 1990) or molecular data (Kwiatowski et al. 1997; Powell 1997).

Throckmorton (1975) advanced a comprehensive scheme of the phylogenetic relationships in the Drosophilidae and showed that paraphyly is widespread among the various groups. However, he did not make any attempt to bring the classification of the Drosophilidae in correspondence with his hypothesis of phylogenetic relationships. Grimaldi (1990) has more recently constructed a phylogeny of the family, using a number of morphological characters and relying on strict cladistic methods, concluding also that Wheeler's classification implies extensive paraphyly. Grimaldi (1990) has accordingly proposed a new classification of the Drosophilidae, which is consistent with his hypothesis of phylogenetic relationships. Grimaldi's phylogenetic hypothesis displays important disparities with Throckmorton's and has been shown also to be inconsistent with extensive molecular data (DeSalle 1992; Thomas and Hunt 1993; Kwiatowski et al. 1994, 1997; Powell 1997; Remsen and DeSalle 1998).

One particularly noteworthy discrepancy between Grimaldi's and Throckmorton's phylogenies concerns the position of the subgenus *Sophophora* (which includes

**Present address:* Institute of Botany, Warsaw University, 00-478 Warszawa, Al. Ujazdowskie 4, Poland; *e-mail:* jmkwiato@PLEARN.EDU.PL

†*Present address:* Department de Genetica, Universitat de Valencia, Dr. Moliner 50, 46100 Valencia, Spain; *e-mail:* Eladio.Barrio@uv.es

Correspondence to: F.J. Ayala; *e-mail:* fjayala@uci.edu

D. melanogaster). Grimaldi (1990) considers *Sophophora* to be a sister-taxon of the subgenus (s.g.) *Drosophila,* which together with the s.g. *Dorsilopha,* would make up the genus *Drosophila.* In contrast, Throckmorton (1975) considered the s.g. *Drosophila* to be phylogenetically closer to several genera and subgenera (such as *Zaprionus, Samoaia, Dorsilopha, Hirtodrosophila,* and *Scaptomyza*) than to *Sophophora.* Molecular data have, on the whole, favored Throckmorton's rather than Grimaldi's hypothesis in this respect (e.g., Kwiatowski et al. 1994, 1997; Russo et al. 1995; Powell 1997; Remsen and DeSalle 1998) but have left unresolved the relationships among several genera and subgenera (and brought into question whether the *willistoni* group, usually included in the subgenus *Sophophora,* may actually be a sister taxon to the *Drosophila* genus (e.g., Pélandakis and Solignac 1993). These unsettled issues are significant, particularly the phylogenetic position of *Sophophora,* because this subgenus includes *D. melanogaster,* which is so extensively used as a model species for many evolutionary, developmental, and molecular biology investigations.

In this paper we study the phylogenetic relationships among 24 species of the family Drosophilidae, using the nucleotide sequences of *Dopa decarboxylase (Ddc),* a nuclear gene involved in morphological differentiation and in the production of the neurotransmitters, dopamine and serotonin. The product of this gene, DDC, catalyzes the decarboxylation of dopa to dopamine and is essential for the sclerotization and melanization of the cuticle (Wright 1996, and references therein). This gene is conserved between *Drosophila* and humans and is expressed in the central nervous system (CNS) as well as in the peripheral nervous system of insects and mammals (Wright et al. 1982; Konrad et al. 1993; Wang and Marsh 1995; Wang et al. 1996; Wright 1996). The only Drosophilid nucleotide sequence of *Ddc* already published is that of *D. melanogaster* (Eveleth et al. 1986). We have sequenced this gene in another 22 Drosophilid species and in the medfly *Ceratitis capitata.* The *Ddc* gene has been found to be a highly appropriate marker for phylogenetic analysis in a subfamily of Lepidoptera that arose within the last 20 million years (Fang et al. 1997). Comparison of *D. melanogaster Ddc* with that of other animals (such as mosquito, moth, and some mammals) indicates that it can be informative at deeper taxonomic levels as well.

Materials and Methods

Species. The 24 species studied are listed in Table 1. The Drosophilidae species originate from the National Drosophila Species Resource Center (Yoon 1996); for the source of *Ceratitis capitata* see Kwiatowski et al. (1992). We list as *Drosophila* subgenera some taxa classified as genera by Wheeler (1981), but *Scaptodrosophila* as a genus, following Grimaldi (1990) and Kwiatowski et al. (1994, 1997).

DNA Preparation and Sequencing. Genomic DNA was extracted following the procedure described by Palumbi et al. (1991). The published sequences from a moth (*Manduca sexta;* GenBank U03909), a fly (*Drosophila melanogaster;* X04661), and a mosquito (*Aedes aegypti;* U27581) were used to design PCR primers. Two slightly different methods (*a* and *b*) were followed for amplification and sequencing. Method *a* was used for species 1–12, 14, 22, and 24; method *b,* for species 13 and 15–23 (see Table 1 for ID numbers; *Scaptodrosophila* was analyzed with both methods).

Method a. The amplifying primers were 5′-CACTGGTACCGNCCCAASTTYCAYGCCTACTTCCCCAC-3′ (APF; forward primer), and 5′-CCGCTCGTTGGTACCCTTNAGCCGGAAGCAGACCA-3′ (APR; reverse primer). The amplified fragment is 963–966 bp long and encompasses most of the *Ddc* exon 4 gene, accounting for 68% of the total *Ddc* coding sequence in *D. melanogaster.* The shorter fragments (963 rather than 966 bp) are due to a 3-bp deletion shared by all species of the *Sophophora* subgenus. All PCR reactions were conducted as described by Kwiatowski et al. (1991). Amplification parameters were as follows: initial denaturation at 94°C for 5 min, followed by 31 cycles of denaturation, annealing, and extension at 94°C for 30 s, 60°C for 30 s, and 72°C for 2 min, respectively; after 30 cycles, the reaction was additionally kept at 72°C for 7 min to complete extension. PCR products were purified with Wizard PCR preps DNA purification system (Promega corporation) and cloned using the TA cloning kit (Invitrogen, San Diego, CA). DNA sequencing was done by the dideoxy chain-termination technique with Sequenase Version 2.0 T7 DNA polymerase (Amersham Life Sciences Inc., USA) using ^{32}P-labeled dATP. The internal sequencing primers were as follows: A1F, 5′-ATAGGCAAGCTGGTGGGCTA-3′; A2F, 5′-ATCCAGATTGGGAYGARCACAC-3′; A3F, 5′-TGGTGAATTTCGACTGCTCGGCCATGTGG-3′; A1R, 5′-AGAGCCACCAAGGTGGATACACTGGC-3′; A2R, 5′-TCRAAGTTSACCAGCATCCA-3′; and A3R, 5′-ASCCACATGGCMGAGCAGTC-3′.

Method b. The amplifying primers were 5′-GAYATYGARCGNGTSATCATGCCKGG-3′ (BPF; forward primer) and 5′-TSRGTGAATCGNGARCADAYKGCCAT-3′ (BPR; reverse primer). The amplified fragments were longer than with method *a,* but we analyze here only the nucleotide sequence corresponding to the PCR fragments of method *a.* PCR amplifications were performed in a 100-μl volume of the ExTAKARA buffer, containing 2.5 U of ExTAKARA Taq polymerase, 0.2 m*M* dNTP (all from TAKARA), a 0.5 μ*M* concentration of primers, and 3 μl of template DNA. The cycling parameters for the amplification were as follows: initial denaturation at 95°C for 5 min, followed by 31 cycles with denaturation for 30 s at 95°C, annealing for 30 s at 59°C, and extension for 2 min at 72°C; after 30 cycles the reaction was additionally kept at 72°C for 7 min to complete extension. PCR products were purified with Wizard PCR preps DNA purification system (Promega corporation), and both strands of the PCR fragments were sequenced directly with an ABI Model 377 autosequencer using the Dye Terminator Ready Reaction Kit according to the manufacturer's protocol (Perkin Elmer). The two amplification primers were also used for sequencing. Internal primers used for sequencing were as follows: B1F, 5′-CNCAYTCNTCNGTGGARCG-3′; B2F, 5′-YGAYTGYTCNGCYATGTGG-3′; B1R, 5′-CGYAGNCKATTRTKCTCATC-3′; and B2R, 5′-TTRAANGCRTTNACCACCCA-3′.

The sequence of *Ceratitis capitata* was obtained from three separately amplified and cloned overlapping segments. Sequencing was done with the forward and reverse primers of the vector, otherwise using the procedures of method *a.* (A sequence of *C. capitata* is available from GenBank, Y08388, but it may have come from a different species. See Appendix 2.)

Sequence Analysis and Phylogeny Reconstruction. Sequences were entered, edited, and assembled using programs of the Fragment Assembly module and aligned using PILEUP and LINEUP of the GCG package (Version 9.1). Alignment required a 3-bp-long gap to be in-

Table 1. The 24 species studied

Family	Genus	Subgenus	Group	Species	ID No.[a]
Drosophilidae	*Drosophila*	*Sophophora*	*melanogaster*	*melanogaster*	1*
				simulans	2
				teissieri	3
				erecta	4
			obscura	*bifasciata*	5
				bogotana	6*
				persimilis	7
			willistoni	*paulistorum*	8
				willistoni	9*
				nebulosa	10
		Drosophila	*virilis*	*virilis*	11*
			repleta	*hydei*	12*
			immigrans	*immigrans*	13*
			MMP[b]	*mimica*	14*
		Scaptomyza[c]		*palmae*	15
				adusta	16*
		Hirtodrosophila		*pictiventris*	17*
		Dorsilopha		*busckii*	18*
		Zaprionus[c]		*tuberculatus*	19*
		Liodrosophila[c]		*aerea*	20
		Samoaia[c]		*leonensis*	21
	Scaptodrosophila[c]			*lebanonensis*	22*
	Chymomyza			*amoena*	23*
Tephritidae	*Ceratitis*			*capitata*	24*

[a] Asterisks indicate species analyzed for three additional genes: *Sod*, *Adh*, (except *D. busckii*), and *Gpdh* (except *D. adusta*, *D. mimica*, and *D. immigrans*).

[b] Modified mouthparts, a group of Hawaiian drosophilids.

[c] *Scaptodrosophila* is classified by Wheeler (1981) as a subgenus of *Drosophila* but has been raised to genus by Grimaldi (1990; see also Kwiatowski et al. 1994, 1997). *Scaptomyza*, *Zaprionus*, *Liodrosophila*, and *Samoaia* are classified as genera by Wheeler (1981); in this paper we refer to them, as well as to *Hirtodrosophila* and *Dorsilopha*, as subgenera within the genus *Drosophila*.

serted in the same position (385–387 in Appendix 1) in all *Sophophora* sequences. The MEGA program (Kumar et al. 1993) was used to calculate distances and to construct evolutionary trees with the neighbor-joining (NJ) method (Saitou and Nei 1987), and for calculating several descriptive statistics. Maximum-parsimony trees were constructed using the PHYLIP 3.57 package (Felsenstein 1989). Alternative topologies were compared using Templeton (1983) and Kishino and Hasegawa (1989) tests, implemented in PAUP [Version 4.0.0d64 (Swofford 1998)]. Codon usage bias was measured with ENC (or n_c, the effective number of codons) (Wright 1990). Higher values of ENC indicate less codon usage bias.

Our phylogenetic analysis includes three additional genes: *Sod*, *Adh*, and *Gpdh*. These DNA sequences, mostly obtained in our laboratory, are available from GenBank.

Results

The *Ddc* gene structure, amplification, and sequencing strategy are shown in Fig. 1. The 24 sequences (23 Drosophilidae plus the medfly *Ceratitis capitata*) are given in Appendix 1. Across the 23 Drosophilidae taxa, 413 sites are variable (43% of the 966 in the sequence; nt1:nt2:nt3, 79:34:300). Three hundred sixty sites are parsimony-informative (nt1:nt2:nt3, 61:24:275). The 10 species of the subgenus *Sophophora* lack a codon (nucleotide positions 385–387 in Appendix 1) that codes for asp in the other species. There is not much bias in GC

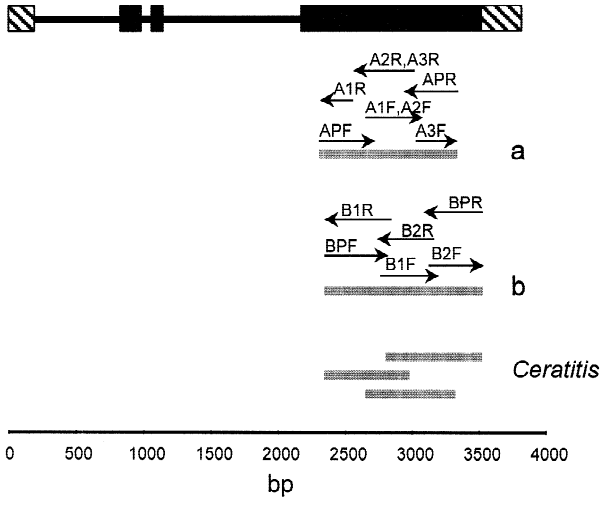


Fig. 1. Structure of the *Ddc* gene and strategy for amplification, cloning, and sequencing. The *boxes* represent exons; their coding parts are *black* and their noncoding parts are *hatched*. *Thick lines* connecting the boxes are introns. The *thick gray lines* represent the segments amplified and sequenced, with primers shown as *arrows* above them. The sequence of *Ceratitis capitata* was obtained from three separately amplified fragments, which were cloned and both strands sequenced with standard vector primers. *a* and *b* refer to two methods; for the species studied by each method, see the text.

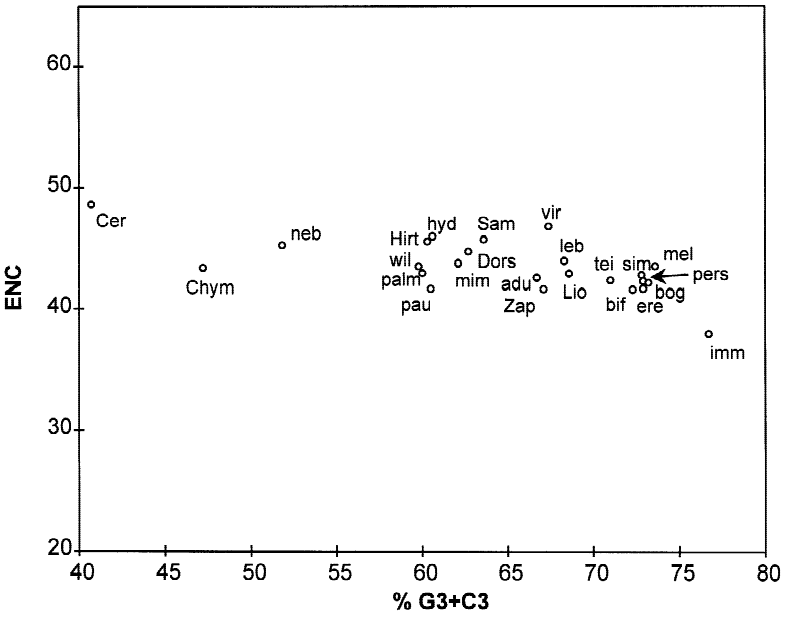


Fig. 2. Effective number of codons (ENC) versus GC content in the third codon position of *Ddc*. The abbreviations refer to the species names, as listed in Table 1.

content overall, but the variation is large at the third codon positions (Fig. 2), ranging from 0.47 in *Chymomyza* (even lower in *Ceratitis*, 0.41) to 0.77 in *D. immigrans*. The variation is notably large within the subgenus *Sophophora*, with the three *willistoni* group species having 52–61%, while the *melanogaster* and *obscura* groups have more than 70% GC. Codon usage bias, as expressed by the effective number of codons (ENC) does not differ among species (Fig. 2).

A neighbor-joining (NJ) tree based on Jukes–Cantor distances is presented in Fig. 3. *Ceratitis* (family Tephritidae) is used as the outgroup. A few species clusters are well resolved on the tree. All three species groups of the subgenus *Sophophora* form well-supported monophyletic groups, but the relationships among the three groups or between them and other Drosophilids are not well defined in this tree. The two species of *Scaptomyza*, one from Hawaii *(D. palmae)* and the other from Texas *(D. adusta)*, form a monophyletic group that clusters in turn with the Hawaiian *D. mimica* with a high statistical (bootstrap) support. The two Drosophilidae genera, *Chymomyza and Scaptodrosophila*, are outside all other species, consistent with previous results (Kwiatowski et al. 1994, 1997), but with unreliable bootstrap values in the present case. Other NJ trees based on Kimura's (1980) two-parameter distance and on the *p*-distance (proportion of different nucleotide sites) are consistent with Fig. 3 and yield similar statistically dependable relationships. A maximum-parsimony tree has a somewhat different topology, but with very low support for its nodes and yields monophyly for each of the *Sophophora* species groups, as well as for the association of *Scaptomyza* with *D. mimica*.

We have studied the same set of species (with the exceptions noted in Table 1) for three other genes (with the number of coding nucleotides analyzed, in parentheses): *Sod* (342), *Adh* (516), and *Gpdh* (729). For simplicity, only one species from each of the three *Sophophora*

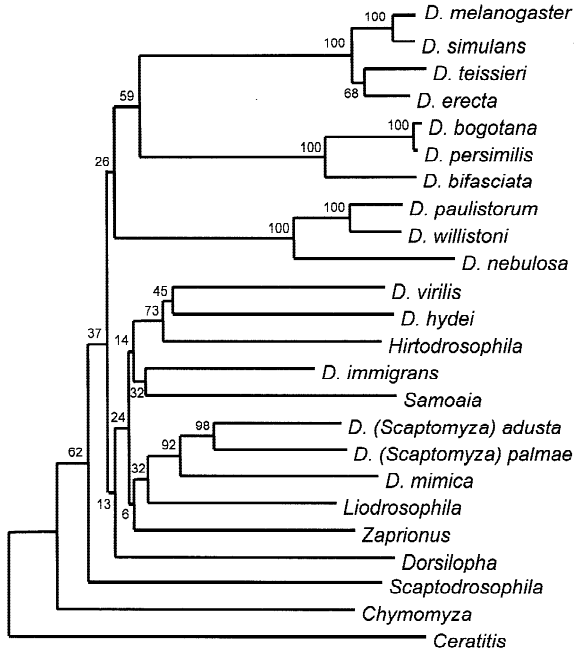


Fig. 3. Neighbor-joining tree based on Jukes–Cantor distances for *Ddc* nucleotide sequences. The bootstrap confidence level (1000 replications) is shown for each interior branch tested.

species groups is included in the analyses that follow, given that the monophyly of each group is so strongly supported in the *Ddc* tree (100% in each case) and otherwise. In the case of *Adh* we have replaced one species with another in two cases because of unavailability: the *Scaptomyza* species *albovittata* (rather than *adusta*) and *Chymomyza procnemis* (rather than *amoena*). Tamura et al. (1995) have studied *Adh* in numerous *Scaptomyza* species and concluded that they all form a monophyletic cluster.

Figure 4 displays four NJ trees based on Jukes–Cantor

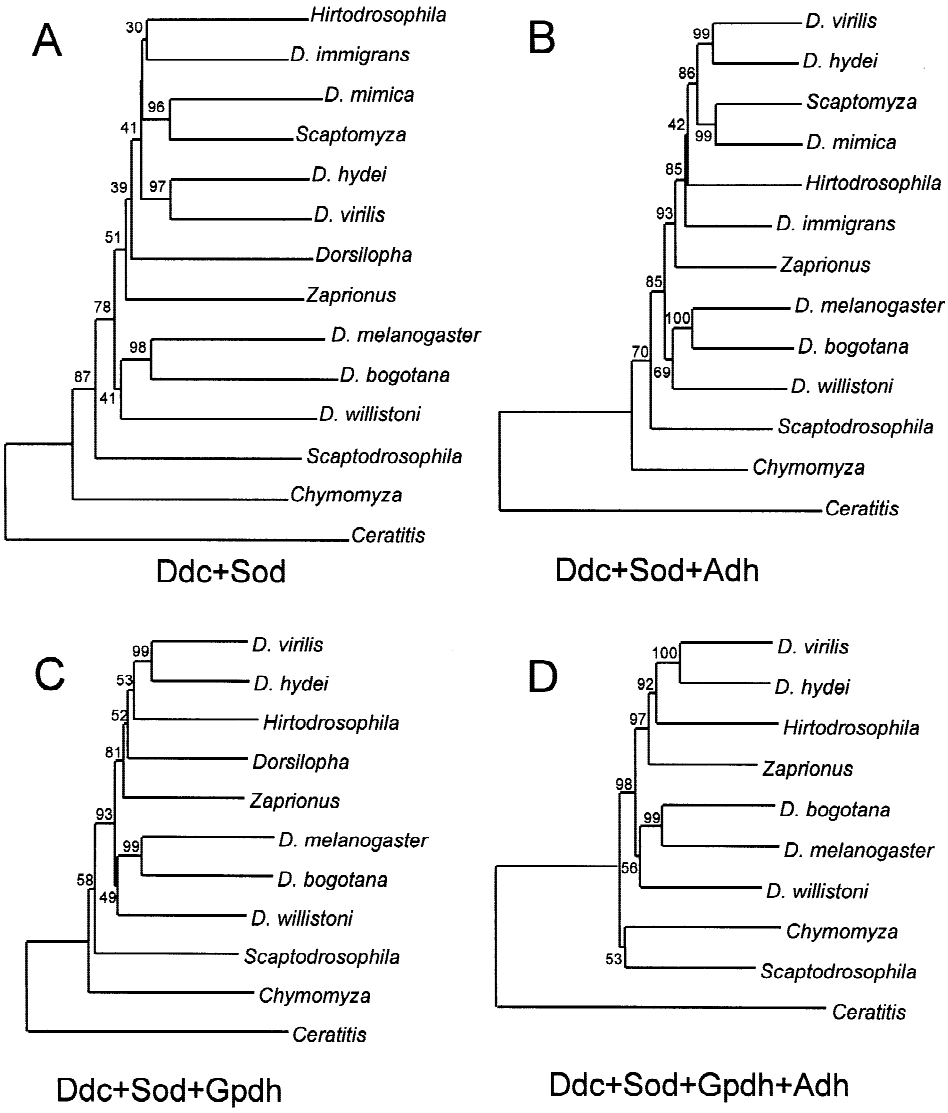


Fig. 4. Neighbor-joining trees based on Jukes–Cantor distances using combined data sets for four genes. Bootstrap confidence levels (1000 replications) are shown for all interior branches tested.

distances obtained by considering other loci in addition to *Ddc*. Trees obtained with Kimura's (1980) two-parameter or Tamura's (1992) distance have precisely the same topologies as those shown in Fig. 4, and with similar bootstrap support. Maximum-parsimony trees also yield identical, or very similar topologies, but typically with lower bootstrap values than the NJ trees.

The combination of *Ddc* and *Sod* (Fig. 4A) brings bootstrap reliability to several nodes that were unresolved by *Ddc* alone. Incorporating also *Adh* (Fig. 4B) resolves most of the nodes of interest. *Chymomyza* and *Scaptodrosophila* are outside all other Drosophilids, with moderately strong indication that *Chymomyza* is the outgroup to the rest. The order of branching of these two genera has remained largely unresolved in the past. Throckmorton (1975) puts *Scaptodrosophila* in the ancestral position, while on Grimaldi's (1990) tree their branching order was not resolved. Of the molecular studies that include both species, DeSalle (1992) considers *Scaptodrosophila* the most ancient, a position also favored by Kwiatowski et al. (1994, 1997), who point out the absence of statistical support for this hypothesis. Beverly and Wilson (1984) favored *Chymomyza* as the ancestral lineage. This ancestry of *Chymomyza* is also favored by combining *Ddc* + *Sod* + *Gpdh* (Fig. 4C), but with a low statistical reliability. The combination of all four genes (Fig. 4D) leaves the matter unresolved. If we use only codon positions 1 + 2, the NJ as well as the maximum-parsimony trees combining any three or all four genes place *Scaptodrosophila* as the outgroup to *Chymomyza* + *Drosophila* (Fig. 5).

Figure 4 shows the *Sophophora* subgenus (*melanogaster*, *obscura*, and *willistoni* groups) as the sister group to all other *Drosophila*, namely, the cluster of the *Drosophila* subgenus plus *Scaptomyza*, *Hirtodrosophila*, and *Zaprionus* (93% bootstrap value in Fig. 4B and 81% in

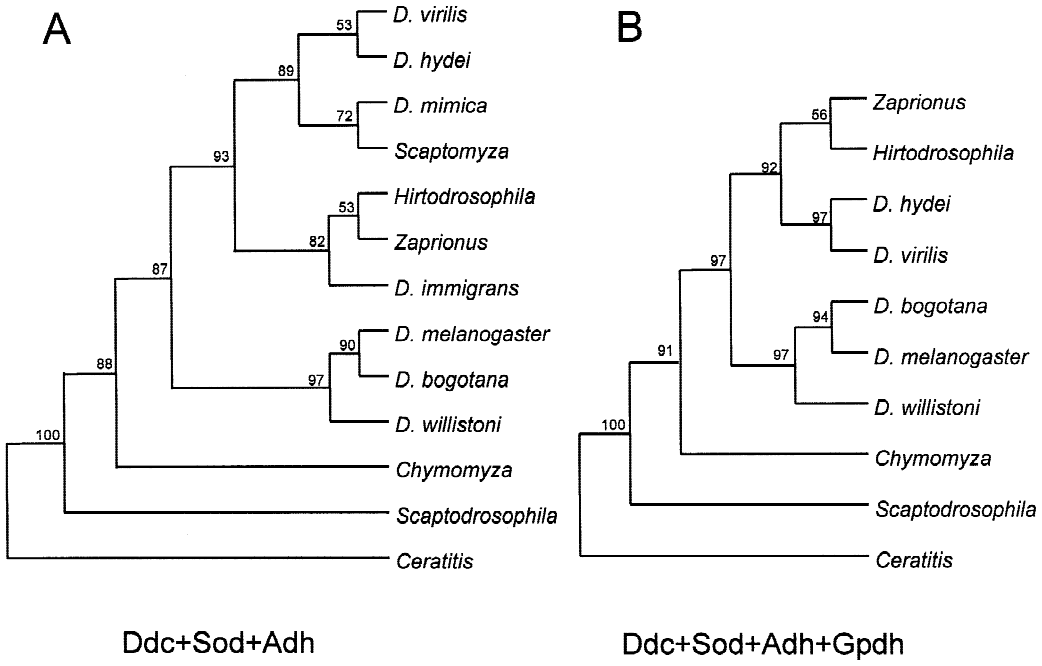


Fig. 5. Two most parsimonious trees based on the 1 + 2 positions of the combined data sets for three and four genes. Other combinations of two and three genes yield similar trees, but less well resolved. Bootstrap confidence levels (1000 replications) are shown for all interior branches tested.

Fig. 4C, which includes, in addition, the subgenus *Dorsilopha*, but not *D. immigrans* or the cluster *Scaptomyza* + *D. mimica*). The position of the *Sophophora* subgenus as the outgroup to the other *Drosophila* subgenera (*Drosophila*, *Hirtodrosohila*, and *Zaprionus*) is also firmly supported (97% bootstrap) by the combination of all four genes. The same conclusion is obtained if we use only codon positions 1 + 2 and was also reached by Tamura et al. (1995), based on the analysis of an *Adh* sequence longer than the one used in our analysis.

The conclusion that the subgenus *Sophophora* is monophyletic to all other *Drosophila* subgenera is supported by our observation of a 3-bp deletion (385–387 in Appendix 1) that appears in all *Sophophora* species (the three *willistoni* group species as well as in the *obscura* and *melanogaster* groups) but not in any of the other *Drosophila* subgenera (or in any of the outgroup genera, *Scaptodrosophila*, *Chymomyza*, and *Ceratitis*). The position of the *willistoni* group species based on genetic distances is equivocal, since the *willistoni* species often appear outside all other *Drosophila* lineages, including the other *Sophophora* (e.g., Pélandakis and Solignac 1993; Powell 1997), which may be a consequence of untypical molecular evolution in the *willistoni* group, as it is apparent in Fig. 2 with respect to third-position GC content. The monophyly of the *Sophophora* species is firmly supported when we analyze our data using only 1 + 2 codon positions.

All trees in Figs. 4 and 5 show *D. virilis* and *D. hydei* as a well-defined monophyletic cluster, as has also been determined in other molecular studies (Kwiatowski et al. 1994, 1997). The monophyly of *D. mimica* and *Scaptomyza* is also highly reliable (Figs. 4A and B, 5A), which is consistent with the Hawaiian origin of *Scaptomyza*, although this was classified as a separate genus by Wheeler (1981). Presumably, *Scaptomyza* shared with *D. mimica* a common ancestor within the *Drosophila* subgenus, in which *D. mimica* is usually included. The incorporation of *Scaptomyza* within the *Drosophila* subgenus is statistically supported in Figs. 4B and 5A by the association of the two pairs *D. mimica* + *Scaptomyza* and *D. virilis* + *D. hydei* (86 and 89% bootstrap, respectively). Nevertheless, the subgenus *Drosophila* would not seem to be monophyletic, even if we include *Scaptomyza*, because the species just mentioned appear to be equally or more closely related to the subgenus *Hirtodrosophila* than to other species of the subgenus *Drosophila* (*D. immigrans;* see below and Fig. 4A, B) when all sites are used. The subgenus *Drosophila* is not monophyletic either when the trees are based only on codon positions 1 + 2.

Figure 4 consistently shows *Zaprionus* as the outgroup to all *Drosophila* subgenera, other than *Sophophora*, with a high statistical reliability in Fig. 4B (85% bootstrap) and Fig. 4D (92% bootstrap). However, when only codon positions 1 + 2 are used, the phylogenetic relationships are somewhat changed, so that *Zaprionus*, *Hirtodrosophila*, and *D. immigrans* form a well-defined monophyletic group (82% bootstrap; Fig 5A). The reason for this discrepancy between the trees based on all positions or only 1 + 2 are not clear. One possibility could be differences in GC content in the third positions.

In order to address the problem of compositional bias, we have analyzed the data sets represented in Fig. 4 by

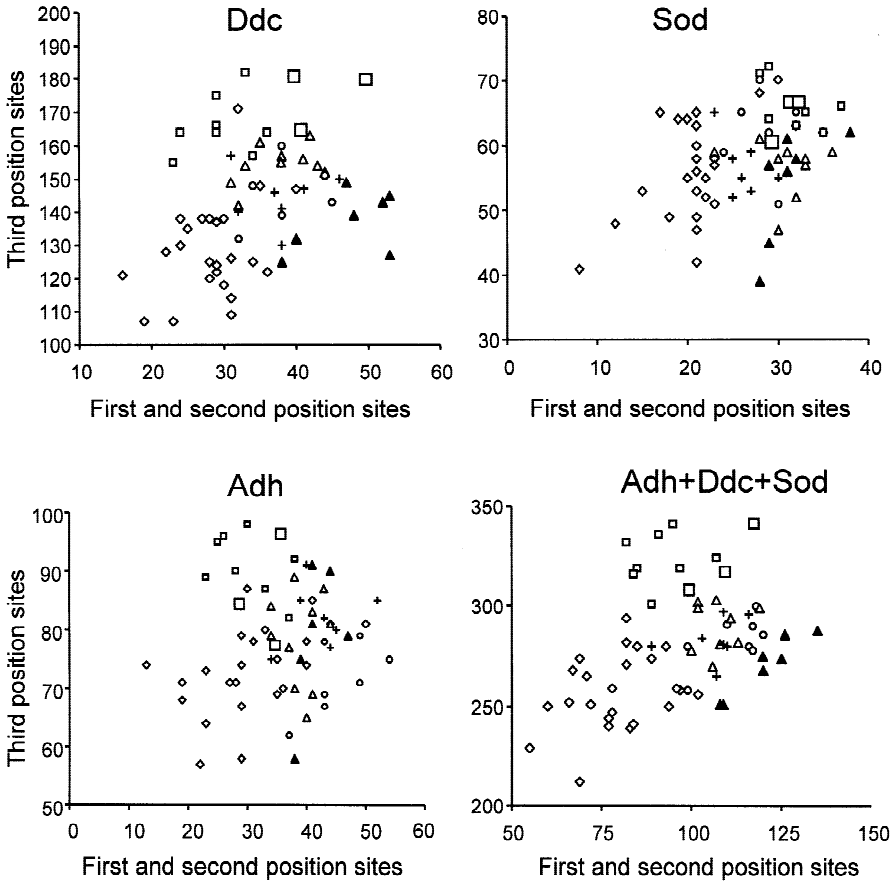


Fig. 6. Number of nucleotide differences between Drosophilids in the third versus the 1 + 2 codon positions. The data are for each of *Adh*, *Ddc*, and *Sod* and for all three genes combined. The figures include all pairwise comparisons between the species shown in Fig. 4B. *Squares*, comparisons between *Chymomyza* and all other Drosophilids (*large squares* for *Sophophora* species; *small squares* for all others). *Triangles*, comparisons between *Scaptodrosophila* and other *Drosophila* species. *Filled triangles*, *circles*, and *crosses*, comparisons between, respectively, *D. melanogaster*, *D. bogotana*, and *D. willistoni* and the seven top species in Fig. 4B. *Diamonds*, comparisons between the seven top species in Fig. 4B.

excluding species that are at the two opposite ends of the spectrum with respect to G3 + C3 content. In all sets, these species are *Ceratitis*, *Chymomyza*, *D. melanogaster*, and *D. bogotana*. (For *Ddc* + *Sod* we did the analyses with and without *D. immigrans*, which has high G3 + C3 in *Ddc*.) The differences in G3 + C3 content for the remaining species are small ($\leqslant$10%), although these species represent all groups of interest. With this procedure the branching order remains the same as in Figs. 4A–C. Thus, GC content differences in the third codon positions do not seem to be the reason for the differences in branching order of *Zaprionus*, *Hirtodrosophila*, and *D. immigrans*, when based on different codon position sites. As we show below, the possible saturation at third position sites is not a factor either, because a plot of the divergences at 1 + 2 versus third positions clearly shows the absence of saturation, especially when *Ceratitis* is not used (as it has not been used in the above analysis).

The branching sequence of *Scaptodrosophila* and *Chymomyza* (Figs. 3 and 4A–C) becomes reversed if we exclude the third codon positions (Fig. 5). Is this a consequence of substitutional saturation at third positions? The evidence favors a negative answer. For the combined data set of four genes, third-position sites remain informative throughout the Drosophilidae and even for the more distant *Ceratitis*. Plots of the divergences of the Drosophilidae species at position 3 versus positions 1 + 2 do not indicate saturation at the third position for any gene or combination thereof (Fig. 6). Similarly, the number of differences at the third position is greater for the comparison between *Ceratitis* and any Drosophilidae species than for any comparisons between Drosophilidae (data not shown). 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. Simulations show that saturation occurs only at a much higher level of sequence divergence than has previously been suggested. 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 Yang (1998), a much more serious problem than saturation is the absence of sufficient information at low levels of divergence.

Another potentially confounding effect may arise from differences in GC content in the third position (G3 + C3). Figure 6 shows the pairwise comparisons between all Drosophilidae species for third versus 1 + 2 positions. It is apparent that comparisons involving *Chymomyza* (squares in Fig. 6) generally show a relatively higher divergence at the third-position sites (Table 2). However, *Chymomyza* has the lowest G3 + C3 content of all Drosophilidae (see Fig. 2). The question is whether the higher divergence at the third position reflects an earlier split of *Chymomyza* from the other Drosophilidae or,

Table 2. Average number of nucleotide substitutions (mean ± SE) between the listed species and the *Zaprionus* clade at each of the genes *Ddc, Sod,* and *Adh* for different codon positions[a]

	Ddc		*Sod*		*Adh*	
	1 + 2	3rd	1 + 2	3rd	1 + 2	3rd
Chymomyza	29.7 ± 1.9	166.1 ± 3.6	31.9 ± 1.3	66.1 ± 1.5	31.0 ± 2.0	91.4 ± 2.1
Scaptodrosophila	36.1 ± 1.7	153.0 ± 2.3	32.9 ± 0.8	56.3 ± 1.9	38.6 ± 1.5	80.3 ± 3.0
D. melanogaster	47.3 ± 2.3	137.1 ± 3.5	31.1 ± 1.2	54.0 ± 3.3	42.0 ± 1.2	79.3 ± 4.2
D. bogotana	38.4 ± 1.8	147.0 ± 3.7	28.4 ± 1.0	63.1 ± 2.5	45.4 ± 2.1	71.6 ± 2.3
D. willistoni	37.6 ± 1.9	144.4 ± 3.2	26.1 ± 0.8	56.7 ± 1.7	42.4 ± 2.1	82.1 ± 2.0

[a] The *Zaprionus* clade includes *D. virilis, D. hydei, D. (Scaptomyza) adusta, D. mimica, D. pictiventris, D. immigrans,* and *D. (Zaprionus) tuberculatus*.

rather, the number of differences at the third position becomes inflated because of the lower incidence of G3 + C3 in *Chymomyza*. To the extent that this effect of nucleotide composition exists at all, it does not seem to be large, since we have found no correlation between the number of nucleotide differences and the differences in G3 + C3 content for all comparisons between *Chymomyza* and the *Drosophila* species (data not shown).

A more serious problem affecting phylogenetic inferences derives from the heterogeneity of substitution rates. Figure 6 shows that the number of substitutions between *Chymomyza* and the other species is relatively small with respect to positions 1 + 2, i.e., most squares are about midrange along the *x* axis, even though a majority of comparisons are between pairs of species more closely related to each other than they are to *Chymomyza;* the only partial exceptions are the comparisons with *Scaptodrosophila* (triangles in Fig. 6). This observation contrasts with the large number of substitutions in third positions, as already noted. The discrepancy is most extreme for *Adh,* but it is also clear for *Ddc* and the three genes combined. In Table 2 we show the average number of differences between species of the *Zaprionus* clade (the seven top species in Fig. 4B) and each of five species ancestral to this clade. For two genes, *Ddc* and *Adh,* the number of substitutions at 1 + 2 positions is consistently smaller between *Chymomyza* and the species of the *Zaprionus* clade than between the *Sophophora* species and the *Zaprionus* clade. With respect to the third position, the opposite is the case; at both *Ddc* and *Adh,* the number of substitutions is consistently greater for the comparisons with *Chymomyza* than with the *Sophophora* species. A similar but much reduced discrepancy occurs for the comparisons with *Scaptodrosophila.* With respect to *Sod,* however, the number of differences at positions 1 + 2 is somewhat greater in the comparisons involving *Chymomyza* and *Scaptodrosophila,* as expected; but at the third positions, the *Sophophora* species are as different from *Chymomyza* and *Scaptodrosophila* as from the *Zaprionus* clade. The conclusion of this analysis is that the rates of nucleotide substitutions, as reflected in the comparison of 1 + 2 versus third position, are variable according to patterns that are inconsistent from gene to gene and from lineage to lineage. This is likely to impact phylogenetic inferences based on numbers of nucleotide substitutions. We may add that, with respect to the number of amino acid replacements in *Gpdh,* there seems to have occurred a rapid acceleration in the *Chymomyza* lineage (Ayala et al. 1996; Kwiatowski et al. 1997), which is just the opposite of the pattern we have just noted for *Ddc* and *Adh.* In any case and for the time being, it seems safe to conclude that the branching order of *Scaptodrosophila* and *Chymomyza* relative to *Drosophila* remains unresolved, although our analysis favors somewhat the hypothesis that the *Chymomyza* lineage is older than *Scaptodrosophila.*

Figure 7 displays six trees with 12 Drosophilidae taxa (and *Ceratitis* as the outgroup). We have tested them statistically, using the combined data for *Ddc, Adh,* and *Sod,* by the methods of Templeton (1983) and Kishino and Hasegawa (1989), both of which yield qualitatively identical results. Table 3 gives the results of the Kishino–Hasegawa tests, which have been performed for the same trees, using all sites or only codon position sites 1 + 2. Tree 1 is favored by our analysis of all sites (the same topology as Fig. 4B). Tree 2 differs from tree 1 only in the position of *Chymomyza* and *Scaptodrosophila.* Tree 3 is favored by the analysis of positions 1 + 2. Trees 2 and 3 are statistically not worse than tree 1 when we use all sites. Trees 4, 5, and 6 represent, respectively, the phylogenetic hypotheses of Throckmorton (1975), Grimaldi (1990), and DeSalle (1992a,b, 1995). Every one of trees 4, 5, and 6 is statistically worse than tree 1, if based on all sites. When 1 + 2 positions are used, tree 3 is statistically preferred over all others, except tree 2.

Figure 8 displays trees that include the subgenus *Dorsilopha* and that are tested using data for only two genes, *Ddc* and *Sod.* Tree 1 has the same topology as tree 1 in Fig. 7 (and Fig. 4B), but with the inclusion of *Dorsilopha* between *Zaprionus* and *D. immigrans,* as favored by our data (Fig. 4A). Tree 2 is the phylogeny favored by analysis of 1 + 2 codon position sites. Trees 3 and 4 correspond, respectively, to the phylogenetic hypotheses of Throckmorton (1975) and Grimaldi (1990). Trees 1 and

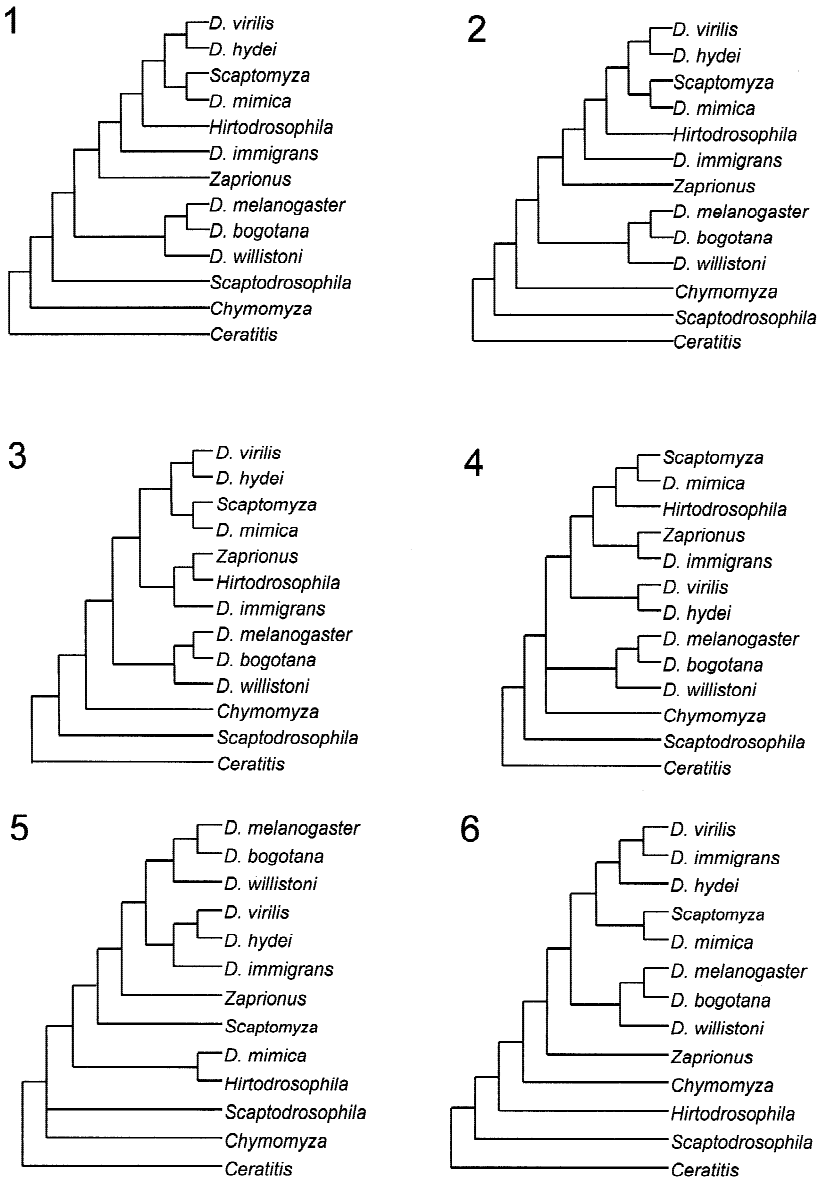


Fig. 7. Alternative topologies for 12 Drosophilid species tested by the methods of Templeton (1983) and Kishino and Hasegawa (1989), using the combined nucleotide sequences of *Adh*, *Ddc*, and *Sod*. The topologies of trees 4–6 represent, respectively, the phylogenetic hypotheses of Throckmorton (1975), Grimaldi (1990), and DeSalle (1992a, b). Results of the tests are given in Table 3.

2 do not differ statistically from each other, whether all positions or only positions 1 + 2 are used; trees 3 and 4 are statistically inferior to trees 1 and 2 (Table 4).

We have also compared trees that are based on all four loci (shown in Figs. 4D and 5B). These trees do not differ statistically from each other by the Kishino–Hasegawa test, whether all positions or only positions 1 + 2 are used. Trees that correspond to the hypotheses of Throckmorton (1975), Grimaldi (1990), and DeSalle (1992a, b, 1995) are statistically worse in both cases than those in Figs. 4D and 5B.

Discussion

A potential benefit of combining data from several loci when testing phylogenetic hypotheses is that the phylogenetic signal weakly present in some genes becomes amplified (Baker and DeSalle 1997). The combined analysis of the three nuclear genes, *Ddc*, *Adh*, and *Sod*, produces the tree shown in Fig. 4B (see also tree 1 in Fig. 7), which has the same topology (but with more taxa included) as the tree obtained by adding a fourth gene, *Gpdh* (Fig. 4D), if all sites are used. Separate analysis of the combined data for *Ddc* and *Sod* allows us to incorporate the subgenus *Dorsilopha (D. busckii)* in that tree (Fig. 4A and tree 1 in Fig. 8). Use of only positions 1 + 2 yields trees (Fig. 5) that are largely congruent with those obtained when all sites are used. It is not clear, however, which set of trees should be given preference. While positions 1 + 2 are less prone to the effect of saturation and nucleotide-composition bias than third positions, they are more likely to be under selective constraints, and this could impact the phylogenetic analysis.

Table 3. Kishino–Hasegawa test of six tree topologies shown in Fig. 7, using the combined data for *Ddc*, *Sod*, and *Adh* with either all codon position sites or only positions 1 + 2: Differences are in comparison to the best tree[a]

	All codon positions			Positions 1 + 2		
Tree	Length	Length difference ± SD	*p*	Length	Length difference ± SD	*p*
1	3419	Best	—	925	19 ± 7	<0.01
2	3424	5 ± 10	0.60	915	9 ± 5	0.06
3	3429	13 ± 14	0.47	906	Best	—
4	3502	83 ± 16	<0.0001	936	30 ± 6	<0.0001
5	3552	133 ± 17	<0.0001	966	60 ± 10	<0.0001
6	3505	86 ± 21	<0.0001	957	51 ± 10	<0.0001

[a] Tree 1 represents the phylogeny favored by analysis of all sites; tree 2 is the same as tree 1 except for the inverted position of *Scaptodrosophila* and *Chymomyza;* tree 3 represents the phylogeny favored by analysis of 1 + 2 position sites; trees 4, 5, and 6 represent, respectively, the phylogenetic hypotheses of Throckmorton (1975), Grimaldi (1990), and DeSalle (1992a, b, 1995). The test of Templeton (1983) yields qualitatively identical results.

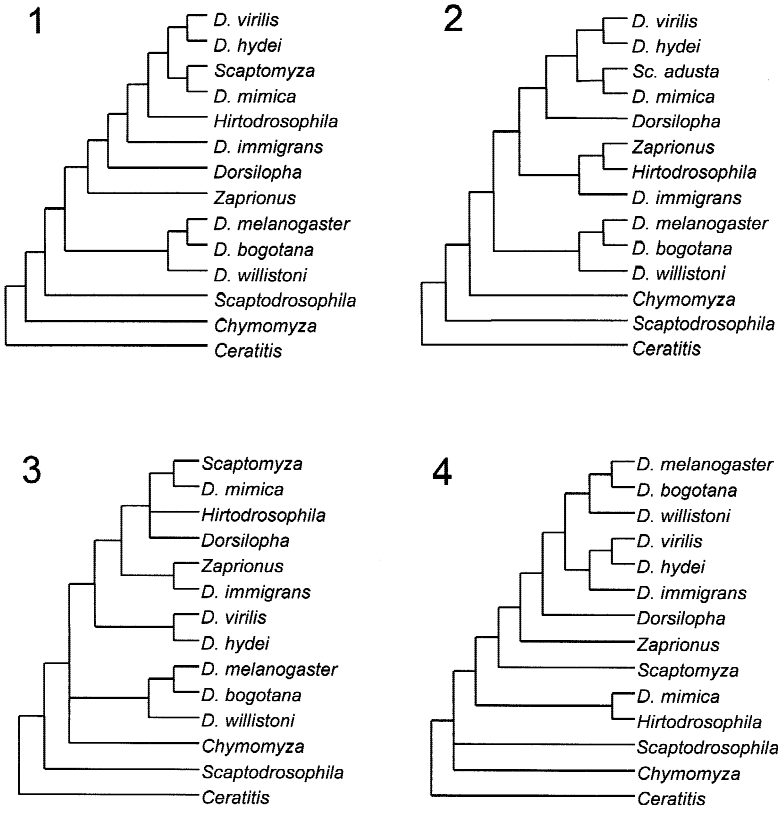


Fig. 8. Alternative topologies for 13 Drosophilid species tested by the methods of Templeton (1983) and Kishino and Hasegawa (1989), using the combined nucleotide sequences of *Ddc* and *Sod.* The topologies of trees 3 and 4 represent, respectively, the phylogenetic hypotheses of Throckmorton (1975) and Grimaldi (1990). Results of the tests are given in Table 4.

Our analysis shows some heterogeneity between all sites and positions 1 + 2 in the Drosophilid lineages, particularly with respect to *Chymomyza*. We have noted that the effects of saturation and nucleotide-composition bias do not seem to be detectable at the third positions. This suggests that trees based on all sites may be most informative. Nevertheless, it is most conservative to consider the position of *Chymomyza* relative to *Scaptodrosophila* as unresolved, especially considering that the Kishino–Hasegawa (1989) and Templeton (1989) tests show that trees based either on all positions or only positions 1 + 2 do not differ statistically. A consensus tree based on all analyses is shown in Fig. 9.

Consistent topologies are obtained and are well supported when pairs of the four genes we have studied are analyzed, although few alternatives become resolved in the separate analysis of individual genes. The combination of data from different genes has to be made with the awareness, as we have shown, that rates of evolution vary among taxa in patterns that are different from gene to gene, and even within a gene, as observed when comparing codon positions 1 + 2 versus 3 (see Results). The

Table 4. Kishino–Hasegawa test of the four tree topologies shown in Fig. 8, using the combined data for *Ddc* and *Sod* with either all codon position sites or only positions 1 + 2: Differences are in comparison to the best tree[a]

	All codon positions			Positions 1 + 2		
Tree	Length	Length difference ± SD	*p*	Length	Length difference ± SD	*p*
1	2337	Best	—	505	6 ± 6	0.32
2	2349	12 ± 13	0.34	499	Best	—
3	2404	67 ± 15	<0.0001	518	19 ± 6	<0.001
4	2408	71 ± 14	<0.0001	528	29 ± 9	<0.001

[a] Trees 1 and 2 represent phylogenetic hypotheses favored by analysis of all sites and by analysis of 1 + 2 position sites, respectively; trees 3 and 4 represent, respectively, the phylogenetic hypotheses of Throckmorton (1975) and Grimaldi (1990). The test of Templeton (1983) yields qualitatively identical results.

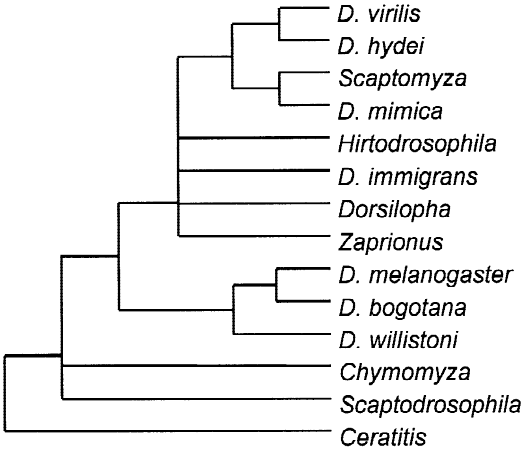


Fig. 9. Consensus tree resulting from combined analyses of four nuclear genes, based on all sites as well as on codon positions 1 + 2. All resolved nodes are strongly supported.

rational expectation is, nevertheless, that the phylogenetic signal will increase on the average, if not always monotomically, with the number of genes incorporated in the analysis. Our analysis firmly supports that *Scaptodrosophila* and *Chymomyza* are outgroups to all other Drosophilid species, in accordance with Grimaldi's (1990) proposition. Although *Chymomyza* is favored as the earliest-diverged lineage, the branching order of these two taxa may for now be considered unresolved, because the results are strongly dependent on which codon positions are included in the analysis, and because of the noted erratic rates of evolution of the various genes in the Drosophilids and, particularly, in *Chymomyza*. More data are needed to resolve the order of branching of these two taxa. DeSalle (1992, 1995) has suggested, based on mtDNA data, that the *Hirtodrosophila* lineage diverged from the other Drosophilids earlier than *Chymomyza*, a hypothesis contradicted by our results.

A controversial matter in *Drosophila* phylogeny concerns the position of *Sophophora*. Two issues are at stake: (1) whether the *Sophophora* subgenus is monophyletic and (2) whether *Sophophora* is an outgroup to the other Drosophila subgenera (and some nominal genera), namely, *Zaprionus*, *Scaptomyza*, *Hirtodrosophila*, *Dorsilopha*, and the subgenus *Drosophila*, including the Hawaiian *Drosophila*.

Traditional taxonomies consider the subgenus *Sophophora* to be a monophyletic taxon that embraces the *willistoni*, *melanogaster*, and *obscura* groups, as well as other groups not included in our study (Wheeler 1981; Patterson and Stone 1952). Several molecular analyses, however, place the *willistoni* group outside a clade that includes all other *Drosophila*, although typically with a low statistical confidence (e.g., Pélandakis and Solignac 1993; Kwiatowski et al. 1994, Figs. 3A and B; Kwiatowski et al. 1997, Fig. 3). This *willistoni* group position as the sister clade to all other *Drosophila*, including the set of the other *Sophophora* groups, such as *melanogaster* and *obscura*, may be considered correct but it may also be attributed to distinctive characteristics of the molecular evolution of the *willistoni* group, such as an accelerated rate of nucleotide substitutions and low G3 + C3 content (review by Powell 1997). The NJ *Ddc* tree shown in Fig. 2 places the *willistoni* group within the *Sophophora* clade, but with a low bootstrap value. Nevertheless, when the *Ddc* data are combined with *Sod* alone, or also with *Adh* and *Gpdh*, the monophyly of the *Sophophora* subgenus is statistically well supported (Fig. 4). This is also the case when only positions 1 + 2 are taken into account (Fig. 5). Moreover, the *Ddc* gene sequences (Appendix 1) provide unambiguous evidence that *Sophophora* is a monophyletic subgenus, because there is a deletion of three coding nucleotides (sites 385–387 in Appendix 1) shared by all *Sophophora* species but no other Drosophilid species (or by *Ceratitis*).

Our results also provide strong support to the traditional interpretation that places *Sophophora* within the genus *Drosophila* (in the *sensu latto* we use), but as the first *Drosophila* clade to branch off, and thus as the sister group to all other *Drosophila* subgenera, as proposed by Throckmorton (1975). A majority of molecular studies supports this positioning of *Sophophora* (Thomas and Hunt 1993; Kwiatowski et al. 1994, 1997; Tamura 1995) (see Table 5). Our analysis of *Ddc* indicates, again in accordance with Throckmorton (1975), that other groups

Table 5. Position of the genera and subgenera listed in the left column with regard to the subgenera *Sophophora* and *Drosophila* according to various molecular studies

Taxon	Ancestral to *Sophophora* + *Drosophila* (Grimaldi's hypothesis)	Placed with s.g. Drosophila, while *Sophophora* is outgroup (Throckmorton's hypothesis)
Liodrosophila		*mtDNA* (DeSalle 1992a) *Adh* (Tamura et al. 1995)
Zaprionus	*mtDNA* (DeSalle 1992a)	*Adh* (Thomas and Hunt 1993; Russo et al. 1995) *Gpdh* (Kwiatowski et al. 1997) *Sod* (Kwiatowski et al. 1994) *18SRNA* (Pélandakis and Solignac 1993)
Samoaia		*18SRNA* (Pélandakis and Solignac 1993)
Dorsilopha		*Gpdh* (Wells 1996; Kwiatowski et al. 1997) *Sod* (Kwiatowski et al. 1994) *18SRNA* (Pélandakis and Solignac 1993)
Engioscaptomyza		*Adh* (Thomas and Hunt 1993; Russo et al. 1995)
Scaptomyza		*mtDNA* (DeSalle 1992b) *Adh* (Thomas and Hunt 1993; Russo et al. 1995) *18SRNA* (Pélandakis and Solignac 1993)
Hirtodrosophila	*mtDNA* (DeSalle 1992a)	*LHP* (Beverly and Wilson 1984) *Gpdh* (Kwiatowski et al. 1997) *Sod* (Kwiatowski et al. 1994) *Adh* (Tamura et al. 1995)
Hawaiian *Drosophila (Idiomya)*		*LHP* (Beverly and Wilson 1984) *mtDNA* (DeSalle 1992b) *Adh* (Thomas and Hunt 1993; Russo et al. 1995)

also are in a derived position relative to *Sophophora* (e.g., the genera *Samoaia* and *Liodrosophila*). Although data have been available for years indicating that *Sophophora* is an early-diverged lineage [e.g., *Sod* (Kwiatowski et al. 1994) and *Adh* (Tamura et al. 1995)], other authors have recently favored the hypothesis placing the *Sophophora* lineage closer to other *Drosophila* subgenera than *Zaprionus* and *Hirtodrosophila* (DeSalle 1995; Powell 1997; Powell and DeSalle 1995). Our analysis of the combined data for four genes (Figs. 4 and 5), as well as the recent analysis of Remsen and DeSalle (1998), clearly contradicts this hypothesis. The suggestion that *Zaprionus* is ''a good choice'' as an outgroup to the genus *Drosophila* (Powell 1997, pp. 275–276) can hardly be maintained.

Our analysis does not agree, however, with Throckmorton's claims concerning the branching order among the rest of the *Drosophila* species, which make up the whole sister clade to *Sophophora.* Throckmorton divides the rest of the species considered here into two clades: the ''*virilis–repleta* lineage,'' which includes *D. hydei* and *D. virilis,* and the ''*immigrans–Hirtodrosophila* lineage,'' which includes *D. immigrans, Zaprionus, Scaptomyza, Hirtodrosophila, Dorsilopha,* and the Hawaiian *Drosophila.* Tamura et al. (1995), based on analysis of *Adh,* have suggested that the Hawaiian groups of *Drosophila* and *Scaptomyza* form a monophyletic group, which is closest to the species in the *virilis–repleta* lineage, but are not included in the *immigrans–Hirtodrosophila* lineage. Our analysis of the combined data for four genes (which include *Adh*) supports Tamura and co-workers' (1995) proposal. The monophyly of *Scaptomyza* and the Hawaiian *Drosophila* is favored by virtually all molecular studies. Placing these two groups as the sister clade to the *virilis–repleta* set contradicts DeSalle's (1992, 1995) conclusion, based on mtDNA, that the Hawaiian flies are an early offshoot of the subgenus *Drosophila.* But it agrees with the recent conclusion of Remsen and DeSalle (1998), based on the combined analysis of several genes.

Our results show that the subgenus *Drosophila* (represented in our study by *D. virilis, D. repleta, D. mimica,* and *D. immigrans*) is likely to be paraphyletic (see Fig. 4, trees A–C, and Fig. 5A), although this is not definite in the consensus tree (Fig. 9), with respect to the genus *Drosophila.* Kwiatowski et al. (1997) suggested removing some paraphyly by downgrading the status of the genus *Zaprionus* to the subgeneric level. But if one is to retain *Sophophora* as a *Drosophila* subgenus, it becomes necessary by cladistic rules also to downgrade *Scaptomyza* and, possibly, the genera *Liodrosophila* and *Samoaia.* When this is done, *Drosophila* is not only a genus ''with too many species,'' but also a genus ''with too many subgenera.''

An alternative possibility would be to raise *Sophophora* to the rank of genus. This would seem justified by the old age of *Sophophora,* which diverged from the other *Drosophila* no less than 50 million years ago (and by the old age of the divergence between the *willistoni* and the *melanogaster* groups, which is no less than 40 million years old) and also by the existence of several hundred *Sophophora* species. However, it is unrealistic to expect that thousands of *Drosophila* geneticists would accept this proposal and refer henceforward to *D. mela-*

nogaster as *Sophophora melanogaster* in the thousands of papers published each year that deal with *D. melanogaster*. Rather more sensible, as a matter of practice, is to enlarge the genus *Drosophila*, as done in Table 1, so that it embraces several taxa formerly ranked as genera.

References

Ayala FJ, Barrio E, Kwiatowski J (1996) Molecular clock or erratic evolution? A tale of two genes. Proc Natl Acad Sci USA 93:11729–11734

Baker RH, DeSalle R (1997) Multiple sources of character information and the phylogeny of Hawaiian drosophilids. Syst Biol 46:645–673

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

DeSalle R (1992a) The phylogenetic relationships of flies in the family Drosophilidae deduced from mtDNA sequences. Mol Phyl Evol 1:31–40

DeSalle R (1992b) The origin and possible time of divergence of the Hawaiian Drosophilidae—Evidence from DNA sequences. Mol Biol Evol 9:905–916

DeSalle R (1995) Molecular approaches to biogeographic analysis of Hawaiian Drosophilidae. In: Wagner WL, Funk VA (eds) Hawaiian biogeography: evolution on a hot spot archipelago. Smithsonian Institution Press, Washington, DC, pp. 72–89

Eveleth DD, Gietz RD, Spencer CA, Nargang FE, Hodgetts RB, Marsh JL (1986) Sequence and structure of the dopa decarboxylase gene of *Drosophila:* evidence for novel RNA splicing variants. EMBO J 5:2663–2672

Fang QQ, Cho S, Regier JC, Mitter C, Matthews M, Poole RW, Friedlander TP, 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

Felsenstein J (1989) PHYLIP—phylogeny inference package (version 3.2). Cladistics 5:164–166

Grimaldi DA (1990) A phylogenetic revised classification of genera in the Drosophilidae (Diptera). Bull Am Mus Nat Hist 197:1–139

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

Kishino H, 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

Konrad KD, Wang D, Marsh JL (1993) Vitelline membrane biogenesis in Drosophila requires the activity of the alpha-methyl dopa hypersensitive gene (I(2)amd) in both the germline and follicle cells. Insect Mol Biol 1:179–187

Kumar S, Tamura K, Nei M (1993) MEGA: molecular evolutionary genetics analysis, version 1.0. Pennsylvania State University, University Park

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

Kwiatowski J, Skarecky D, Ayala FJ (1992) Structure and sequence of the Cu,Zn *Sod* gene in the Mediterranean fruit-fly *Ceratitis capitata:* intron insertion and deletion in the evolution of the SOD gene. Mol Phyl Evol 1:72–82

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

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

Mantzouridis TD, Sideris DC, Fragoulis EG (1997) cDNA cloning of L-dopa decarboxylase from the eclosion stage of the insect *Ceratitis capitata.* Evolutionary relationship to other species decarboxylases. Gene 204:85–89

Marsh JL, Erfle MP, Leeds CA (1986) Molecular localization, developmental expression and nucleotide sequence of the alpha-methyldopa hypersensitive gene of Drosophila. Genetics 114:453–467

Palumbi S, Martin A, Romano S, Macmillan WO, Stice L, Grabovsky G (1991) The simple fool's guide to PCR, Version 2.0. Department of Zoology and Kewalo Marine Laboratory, University of Hawaii, Honolulu

Patterson JT, Stone WS (1952) Evolution in the genus *Drosophila.* Macmillan, New York

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

Powell JR (1997) Progress and prospects in evolutionary biology: the *Drosophila* model. Oxford University Press, New York

Powell JR, DeSalle R (1995) *Drosophila* molecular phylogenies and their uses. Evol Biol 28:87–139

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

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

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

Swofford D (1998) PAUP*: phylogenetic analysis using parsimony (* and other methods). Sinauer, Sunderland, MA

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

Tamura K, Toba G, Park J, Aotsuka T (1995) Origin of Hawaiian drosophilids inferred from alcohol dehydrogenase gene sequences. In: Nei M, Takahata N (eds) Current topics on molecular evolution. Proceedings of the US–Japan Workshop, Hayama, Japan, pp. 25–27 August 1995. Pennsylvania State University, University Park: Graduate School for Advanced Studies, Hayama, Japan, pp. 9–18

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

Thomas RH, Hunt JA (1993) Phylogenetic relationships in *Drosophila:* a conflict between molecular and morphological data. Mol Biol Evol 10:362–374

Throckmorton LH (1975) The phylogeny, ecology, and geography of *Drosophila.* In: King RC (ed) Handbook of Genetics. Plenum Press, New York, pp. 421–459

Wang D, Marsh JL (1995) Developmental regulation of the alpha-methyldopa hypersensitive gene of *Drosophila melanogaster.* Dev Biol 168:598–612

Wang D, Marsh JL, Ayala FJ (1996) Evolutionary changes in the expression pattern of a developmentally essential gene in three *Drosophila* species. Proc Natl Acad Sci USA 93:7103–7101

Wheeler MR (1981) The Drosophilidae: a taxonomic overview. In: Ashburner M, Carson HL, Thompson JNJ (eds) The genetics and biology of *Drosophila.* Academic Press, New York, pp. 1–97

Wright F (1990) The effective number of codons used in a gene. Gene 87:23–29

Wright TRF (1996) Phenotypic analysis of the dopa decarboxylase gene cluster mutants in *Drosophila melanogaster.* J Hered 87:175–190

Wright TRF, Black BC, Bishop CP, Marsh JL, Pentz ES, Steward R, Wright EY (1982) The genetics of dopa decarboxylase in *Drosophila melanogaster.* V. Ddc and I(2)amd alleles: isolation, characterization and intragenic complementation. Mol Gen Genet 188: 18–26

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

Yoon J (1996) Stock list: national *Drosophila* Species Resource Center. Dros Inform Serv 77:1–44

Appendix 1

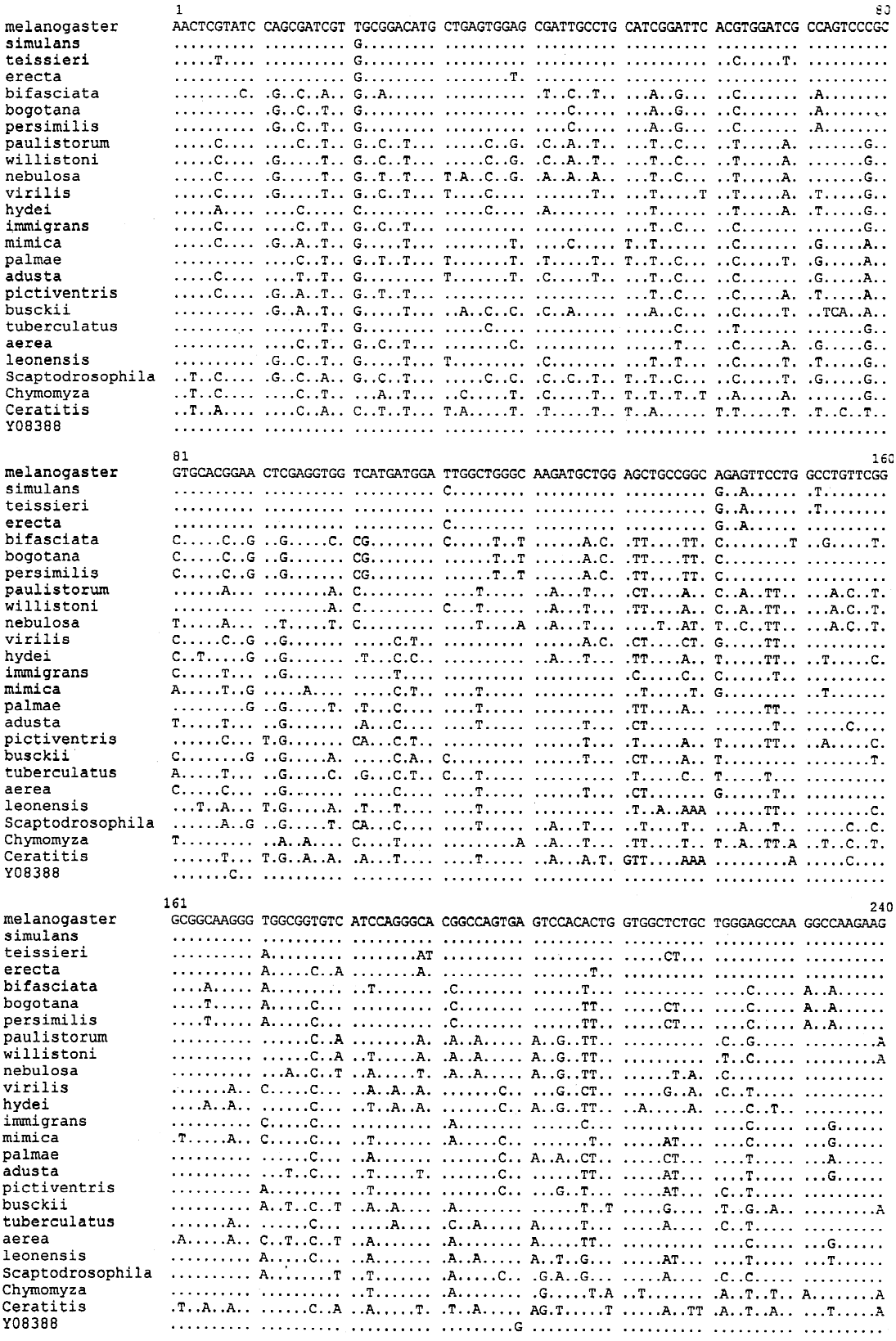


Fig. A1. Twenty-five sequences of *Ddc*. The sequence Y08388 is from Mantzouridis et al. (1997).

```
                    241                                                                                                 320
melanogaster        TTGAAGGAGG TGAAGGAGCT CCATCCGGAG TGGGATGAGC ACACCATCTT GGGCAAGTTG GTGGGCTACT GCTCGGACCA
simulans            .......... .......... G......... .......... .......TC. .......C.. .......... ..........
teissieri           G......... .......... G......... .......... ........C. .......C.. .......... ....T.....
erecta              G......... .......... G......... .......... ........C. .......C.. .......... ....C.....
bifasciata          G....A.... .CC...CA.A G........T .....C.... .......TA. T.....AC.C ..C....... ....T..T..
bogotana            G.T....... .CC...CC.A G.....A..T .......... ........A. T.....AC.C ..C....... .T..T.....
persimilis          G.T....... .CC...CC.A G.....A..T .....C.... ........A. T.....AC.C ..C....... .T..T.....
paulistorum         ...C....A. ......CTGA G........A .......... .T.....AA. T......C.. ........T. .......T..
willistoni          ...C....A. ......CTGA G........A .......... .......AA. T......C.. ........T. .......T..
nebulosa            ...C....A. ......CTGA G.....C..A .......... .T..TG.AA. T..A...C.. .....T.... .T..C..T..
virilis             C......... ......CTA. G.....T... ........T. ....G..AA. A......C.. ........TG C...T.C...
hydei               ..AC.AC... .....A.AGA G.....C... .......... ....A...A. T......C.. ..A.....T. C.....C...
immigrans           C..C....A. ......CTGC G......... .......... ....G...A. T......C.. ........T. C...C.C...
mimica              C.AC...... .A..A.CAGA A.....A..T .......... .A..G...A. T.....AC.. ..A.....TG C...CTC...
palmae              A..C...... ......C.GA G.....T..A .......... ....G...A. T.....AC.. ........TG C...T.C...
adusta              A..C...... ......C.GA G.....C..T .......... ....A...A. T.....AC.. ........TG C...T.C...
pictiventris        C..C.A..A. ....A.CAGA G.....C... ........T. .T..G..TA. T......C.T ..C.....T. ....A.CT..
busckii             ..AC....A. ......CA.A G......... .......... .T.....TA. A....... ..A.....TG C.....CT..
tuberculatus        ...C...... .T....C.GA G.....C..A .......... .T..G..TA. T......C.A .....A.... C.....CA..
aerea               C.TC...... .C....C.GA G.....C..A .......... .T......A. T.....AC.A .........G C.....CT..
leonensis           C..C..C... ....A.CCGA G......... .....C.... .T..G...A. A......C.. ..T.....T. CG..C.C...
Scaptodrosophila    A......... ...G....AC A........C .....C..T. ....G..TA. CTCG...C.T .......... C.....CG..
Chymomyza           A..C...... .T..A.C.GA A.....A..T .....A.... .T..T..TA. TTC...AC.T ........TG CT....CA..
Ceratitis           A.TG.A...A .T..A.T... A..C..T..A ...AG....A .T.....TA. CTC....... ..T..A..TA .....TCG..
Y08388              .......... .......... .......... .......... ........C. C......C.T .C........ ..AGT..T..

                    321                                                                                                 400
melanogaster        GGCTCACTCA TCCGTGGAGC GGGCTGGTCT TCTGGGCGGA GTAAAGCTCC GTTCCGTGCA GTCC---GAG AATCACAGAA
simulans            .......... .....A.... .T........ .......... .......... .......... ....---... ..........
teissieri           .........C .......... .T........ CT....A... ..G....... ....T..... ....---... ..........
erecta              ...C.....C .......... .T.....C.. CT....A... ..G....... ....T..... ....---... ..........
bifasciata          ...G..T..C ..T....... .A..G..A.. ......T..G ..C.....G. ....G....C ...G---... ..GAGTC.CT
bogotana            ...G..T..T .........A .A..G..A.. ......T..G ..C.....G. ....G....C ...G---... ..GAGTC.CT
persimilis          ...G..T..C .........A .A..G..A.. ......T..G ..C.....G. ....G....C ...G---... ..GAGTC.CT
paulistorum         ...G.....T ..G......A .......... CT....T..C A.T..A..G. .C.....CA. ...G---... ..CA.TC.CT
willistoni          ...G...... ..G......A .......... CT....T..C A.T..A..G. .C..T..CA. ...G---... ..CA.TC.CT
nebulosa            .........T ..G......A .A......C.. CT....T..G A.T....T.G. .C......A. ...T---... ...A.TC.CT
virilis             .T.A..T..G .......... .......... CT.......C A.C...T.G. .C..A..T.C .G.GGAT... ..CA.TC.TC
hydei               ...C.....T ..T....... .C........ CT.......C ..C..A.... ....A..T.C TG.TGAT..A C.CA.TC.TC
immigrans           A..G..T..G ..G....... .C..A..... G........C ..C.....G. .C.....C.C AG.TGAT... C.CA.TC.CC
mimica              ...G..T..C ..T.....A. .C.....A.. .T....T..C A.C.....G. .C..G....C CG.TGAT... C.CA.T..TC
palmae              ...G..T..T ..G....... .T........ .T....T..C ..C.....G. .C..G....C .G.GGAT... C.CA.TC.TC
adusta              .........C ..G....... .T..G..... C........C ..C.....G. .C.......C AG.GGAT... C..A.TC.TC
pictiventris        ...C..T..C ..T..A..A. .C.....G.. CT....T..C A.C..A.... .C..A....C AG.GGAT..A ..CA.....C
busckii             ...G..T..G .....T.... .C.....C.. G.......... ..C...T.G. .C.......C CG..GAT... C.CA.TC.TC
tuberculatus        ...G..T..C ..G....... .......... CT....T..C ..C.....G. ....G....C TG.GGAT... C..A..C.TC
aerea               ...G..T..C ..G....... .T..C..A.. CT....T..C ..C.....G. .C..T....C AG.AGAT... ...A.TC..C
leonensis           A..C..T..C ..A....... .C..A..... A..C..A..C ..C.....A. .C..A....C .G.GGAT... ..CA.TC..C
Scaptodrosophila    ...A..T... ..T..T.... .T..G..... G.....T... ..G.....G. .C..A....C CG.GGAT... C.GA.TC.TT
Chymomyza           .........T ..T....... .T........ G..A..A..T ..T.....G. ....T....C AG.TGAT..A ..CA.TC.TC
Ceratitis           A..A..T..C ..A..A.... .C........ AT....T..T .....AT.A. ..GGT..TGC TG.TGAT..A ..CA.T...T
Y08388              .......... ........AA .A......T. G......... ..G.....TG .A........ .AGT---... ..C...C.T.

                    401                                                                                                 480
melanogaster        TGCGTGGTGC TGCCCTGGAA AAGGCCATCG AACAGGATGT GGCCGAGGGT TTGATTCCCT TCTACGCGGT GGTCACCCTG
simulans            .......... .......... .......... ........C. .......... .......... .......... ..........
teissieri           .......... ...A...... .......... .......T. .......... .......... ....T..C.. ..........
erecta              .......... .......... .......... ........C. .......... .......... ....T..... ..........
bifasciata          ....C..C.A G........G .......... .G......C. AAAG..T..C C....A..A. ....T..... ..........
bogotana            ....C..C.A G..T...... .......... .G......C. ..AG..T..C C....A..T. ....T..T.. T........A
persimilis          ....C..C.A G..T.....G .......... .G......C. ..AG..T..C C....A..T. ....T..T.. T........A
paulistorum         .......C.A ......A..G ........T. ........T. ......T..C .....C..G. ....T..T.. C.....T...
willistoni          ....G..C.A ......C..G ........T. ........T. ......T..C .....C..G. ....T..T.. C.....T..A
nebulosa            .......A.A ......T..G ........T. ..A.....T. ......T... C....A..A. ....T..T.. C.....T...
virilis             .A..C..C.A ...TT....G .....T..T. ..A.A...T. ...T..T..C C....A.... .......... C.....AT..
hydei               .......C.A ...TT....G .....A..T. ..G.....T. ...T.CT..C C....A.... ....T..T.. T.....AT..
immigrans           ....C..C.A C..T.....G .........A .......CC. ..A...T..C .......... ....T..... ......G...
mimica              .A......C.A ...TT....G GCC..A..TA .G.....CT. ...T..T... C....A.... ....T..C.. .......T..
palmae              .........A ....T....T GC...T...A ....A...T. ......T... C.C..A.... ....T..T.. .......T..
adusta              .........A C..TT....C GC...A...A .G......T. ......T... C.C..A.... .T........ .......T..
pictiventris        .A.....C.A ...A.....G C....T..AA .G......C. ...TA.T... C.C..... ....T..C.. ......AT..
busckii             .......C.A ...TT....G C........A. ..A.....C. ...T..T... C.C..A..A. ....T..... ......T..
tuberculatus        ....C..C.A ...T.....G .......... ........T. ......T... C.T..A.... ....T..A.. C..T..A..C
aerea               .A.....C.A G..A.....T GC.......A ........CT. ......T... C.A..A.... ....T..... ......A...
leonensis           .A..A....A ....T....G ........AC ..G....CT. ..ATA.T..A ........G. ....T..... ..........
Scaptodrosophila    .......C.A G...T....G .....T..A. .G......T. ......C..C ..A..A..A. ....T..... ......T..C
Chymomyza           .......C.A G..AT....T .....T..TA ........T. ...T..T..C C....A..A. .T..T..A.. ...T..A...
Ceratitis           ....A....A G...T..... GCA..A..AA ..A....CT. ..AA.....G C.T..A..T. .T..T..C.. T..G..GT..
Y08388              ..A.A..... ....T..... .......... .G........ .........A CGT.....T. .......C.. ...G...T..
```

Fig. A1. Continued.

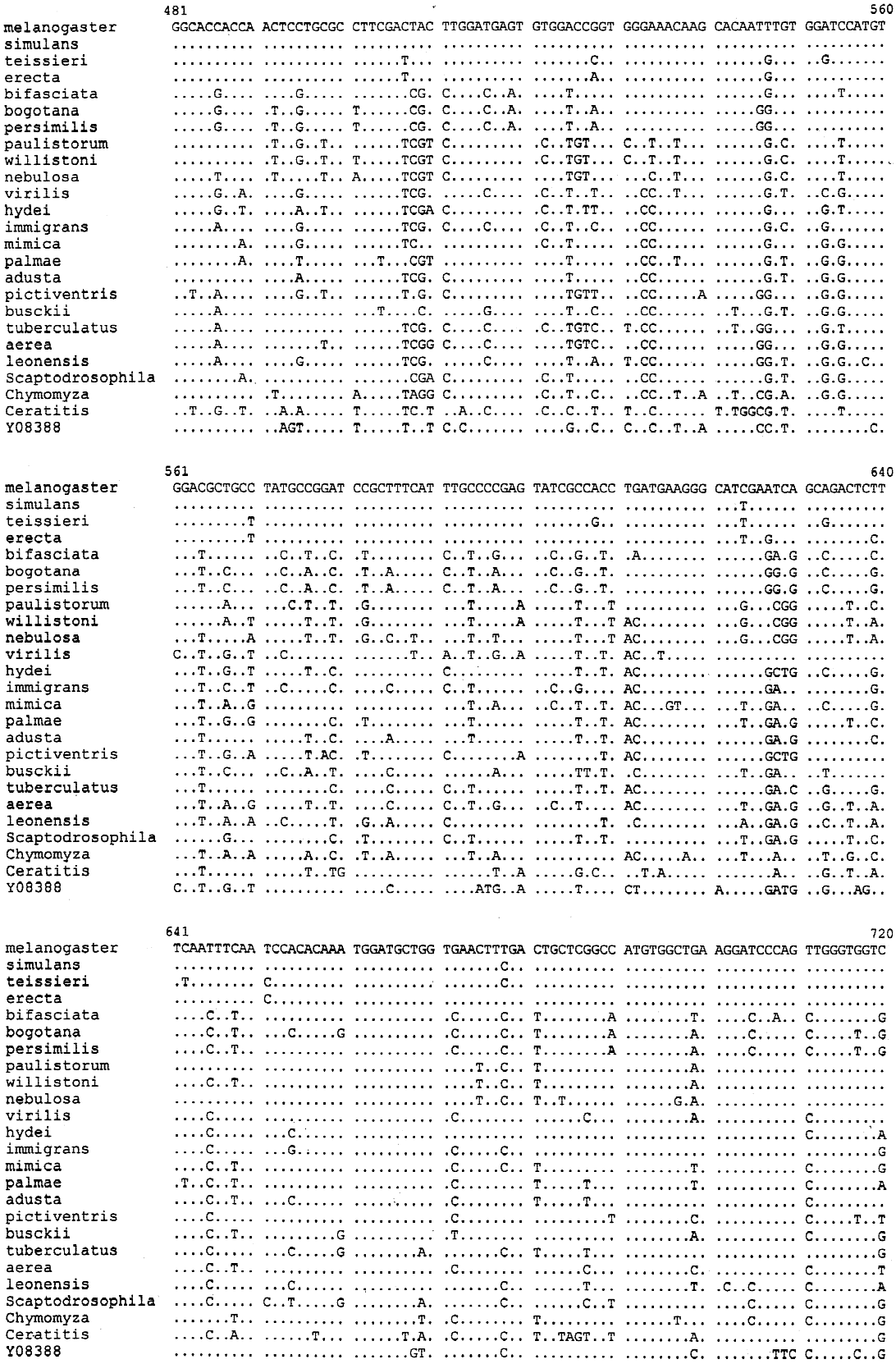
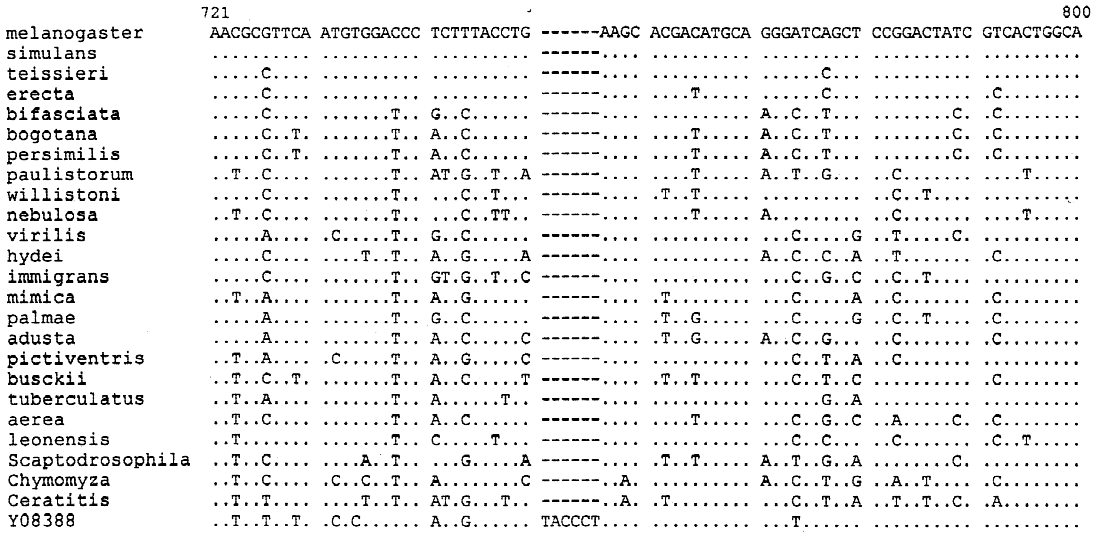
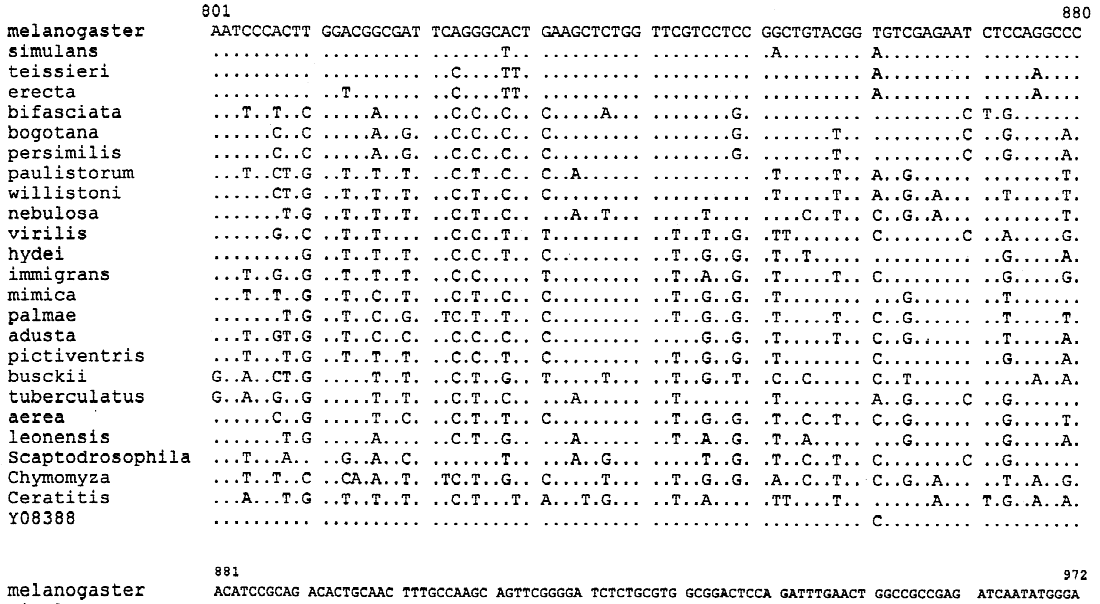


Fig. A1. Continued.





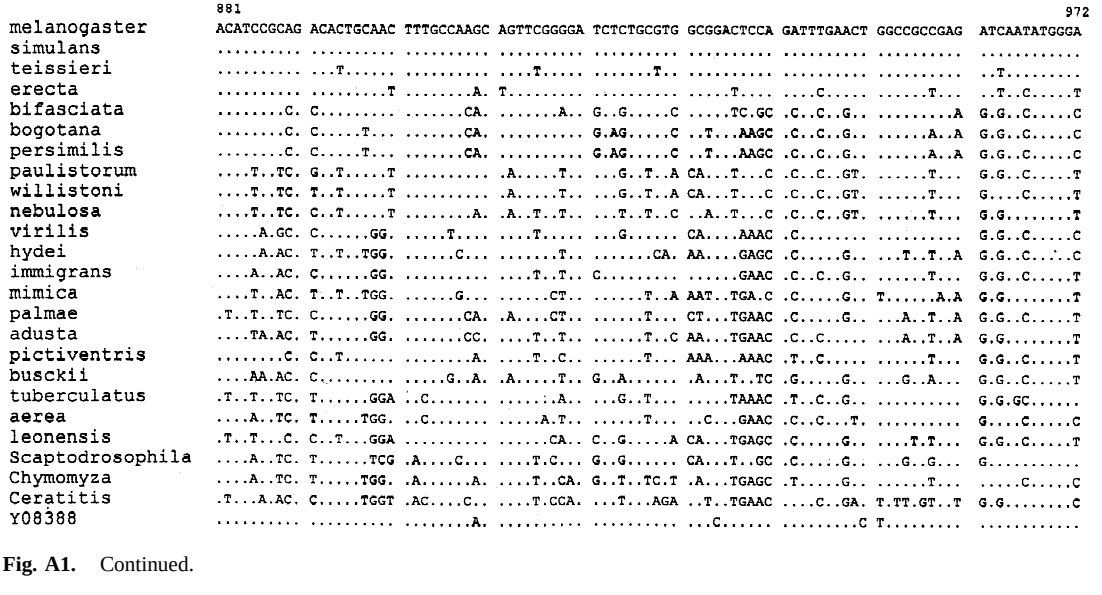


Fig. A1. Continued.

Appendix 2

T.D. Mantzouridis, D.C. Sideris, and E.G. Fragulis (Gene 204:85–89, 1997) have published a cDNA *Ddc* sequence attributed to the medfly *Ceratitis capitata.* Figure A1, bottom row, gives the alignment of this sequence, Y08388, with the others reported in this paper. Figure A2 gives the position of Y08388 in a simplified NJ tree. It is apparent from this tree (and Appendix 1) that Y08388 represents a gene sequence that has only recently (within the last 2–5 million years) diverged from *D. melanogaster*. We have also compared Y08388 with sequences from the related *amd* (α-methyl dopa-hypersensitive) gene, which is assumed to have arisen with *Ddc* from an ancient duplication event. The *amd* genes from *D. melanogaster* (Marsh et al. 1986), as well as from several Drosophilids sequenced in our laboratory, are all extremely distant from any of the *Ddc* genes. Indeed, the fruitfly *amd* gene is more remotely related to any fruitfly *Ddc* gene than any of these is to human *Ddc.* It seems likely that Y08388 comes from a species closely related to *D. melanogaster* rather than from *Ceratitis capitata*. A possible alternative explanation is that Y08388 represents a second *Ceratitis Ddc* gene, acquired by lateral transfer from one of the *melanogaster*-group species within the last 2–5 million years. The transfer of a functional gene between two animals has no known precedent, and it must be therefore considered very unlikely.

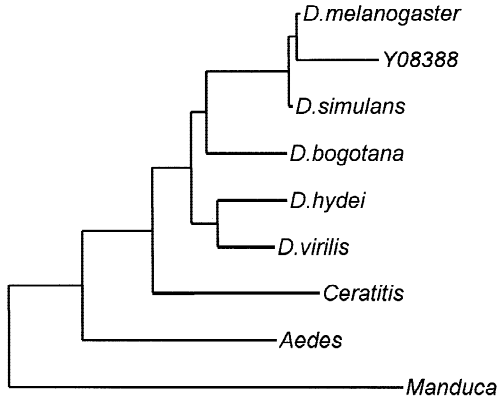


Fig. A2. Neighbor-joining tree of the DDC amino acid sequences from fruitflies, a mosquito, and a moth. Y08388 has been reported to be from the medfly *Ceratitis capitata* (Mantzouridis et al. 1997), but its great difference from the *Ceratitis* sequence we have obtained and great similarity to species of the *D. melanogaster* group make this origin uncertain.