Ddc and amd Sequences Resolve Phylogenetic Relationships of Drosophila

With about 3000 species, the family Drosophilidae is large and diverse even by Dipteran standards. This diversity provides biologists with distinctive opportunities to investigate evolutionary patterns, but also poses taxonomic and other challenges. Thus, the traditional classification (e.g., Wheeler, 1981) is inconsistent with phylogenetic relationships based on morphology (Throckmorton, 1975; Grimaldi, 1990) or molecular data (review in Powell, 1997). The two comprehensive phylogenetic hypotheses of Throckmorton (1975) and Grimaldi (1990) have been tested against recent molecular studies, which have resolved some important discrepancies between them (see Remsen and DeSalle, 1998; Tatarenkov et al., 1999a; Kwiatoński and Ayala, 1999; Katoh et al., 2000). Yet, many phylogenetic relationships remain unsolved, such as those among Hirudoassocida, Zapriosus, Dorsilophia, and s.g. Drosophila (e.g., Tatarenkov et al., 1999a; Kwiatoński and Ayala, 1999). Some genera, such as Lioemtersofoila and Samoaoia, have received scarce attention (Plandakis and Solignac, 1993; Tamura et al., 1995; Tatarenkov et al., 1999a) and their phylogenetic placement remains largely unknown. One problem with the molecular phylogenies is the incompleteness of taxa sampling. Although representatives of some Drosophila groups have been included together in some studies, often different studies include different groups, which prevents construction of a reliable higher-level phylogenetic framework.

We seek to define a robust framework of relationships in the Drosophilidae at the species-group and higher taxonomic levels. We have investigated 29 species (Table 1) from several drosophilid genera and subgenera and from representative species groups for two nuclear genes, dopa decarboxylase (Ddc) and α-methyl dopa (amd). These are closely linked paralogous genes, arisen by an ancient gene duplication (Eveleth and Marsh, 1986; Tatarenkov et al., 1999b). We earlier used Ddc to address some issues of Drosophila systematics (Tatarenkov et al., 1999a). We now extend our previous investigation by including additional taxa for longer sequences of Ddc and a new gene, amd, which previously had been sequenced only in D. melanogaster (Marsh et al., 1986), D. simulans, and Scaptodrosophila lebanonensis (Tatarenkov et al., 1999b).

DNA preparation and sequencing were as described by Tatarenkov et al. (1999a; method b). The 963- to 966-bp-long sequences of Ddc previously reported are now extended to 1131–1134 bp. Ddc sequences of five more species, D. pseudoobscura, D. robusta, D. phalerata, D. funebris, and D. gymnobasis, are added. Amplification and sequencing of amd was as for Ddc, except that the annealing temperature was 60°C and the extension time was 3 min. The amd amplifying primers were 5’-MAYATGCA yogGCTAYTAYCCACCAT-3’ (Amd-un2, forward primer) and 5’-ACCA TRTAGATYTTYTTCNGCTCAT-3’ (Amd-bw, reverse primer). The amplified region of amd encompasses an intron. The amplified fragment varied in length from 1269 bp in D. hydei to nearly 2600 bp in D. tripunctata, depending on the length of the intron. Only the coding regions, 1032-bp-long, were used in this study (66 bp

J. R. Anderson
Paul R. Grimstad
David W. Severson

Department of Biological Sciences
University of Notre Dame
Notre Dame, Indiana 46556
doi:10.1006/mpev.2001.0964

LETTERS TO THE EDITOR

321
from exon 1 and 966 bp from exon 2), because the large divergence of intron sequences made their alignment impossible. Internal primers for sequencing were Amd1: 5'-GNACNTGYGCNTAYGAYGA-3'; Amd1-Rev: 5'-GATCNRCTGNNACSCAS-3'; Amd2: 5'-GTGATGTYTGGYTG-3'; and Amd2-Rev: 5'-GTGCANGRCRCADAT-3'. We succeeded in amplifying only a short fragment (927 bp) from the second exon of Liodrosophila aerea using Amd2 and Amd-bw. Sequences were aligned with programs PILEUP and LINEUP of the GCG package (Wisconsin package, Version 9.1). The alignment required that a 3-bp-long gap be inserted at positions 433–435 in all Sophophora Ddc sequences. Phylogenetic analyses were performed with PAUP (version 4.0b1 for Macintosh; Swofford, 1998) and MEGA (Kumar et al., 1993). We present trees from only the neighbor-joining (NJ) analyses, but all data sets were also analyzed with maximum-parsimony (MP), minimum-evolution (ME), and maximum-likelihood (ML, HKY-gamma substitution model) methods. We use the incongruence length difference (ILD) test (Farris et al., 1994), called the partition-homogeneity test in PAUP, to evaluate incongruence between data partitions. Invariant characters were removed before the ILD test was applied.

Because of the shorter sequence of amd (927 versus 1032 bp) for L. aerea, we conducted all analyses twice, using a set of shorter amd sequences that included L. aerea and a longer set without it. The trees based on these two sets are congruent, thus allowing for straightforward interpretations.

Similarly to other previously studied genes, the amd sequences show considerable variation in nucleotide composition at the 3rd codon position. Strong nucleotide compositional bias, coupled with high divergence levels (maximum values of raw sequence divergence are 25% for Ddc and 29% for amd) raise the issue of whether the 3rd codon positions remain informative for phylogenetic reconstruction. We have exploited the fact that amd and Ddc are ancient paralogous genes (the duplication preceded the divergence of Diptera-Lepidoptera, under the molecular clock assumption). Nucleotide composition varies in similar fashion among species at both genes. If only 3rd codon positions are used, we expect the sequences of both genes to intermingle in the phylogeny if these positions are not informative. Instead, an NJ tree based on only 3rd codon position shows two clusters corresponding toamd and Ddc (100% bootstrap). This indicates that the 3rd codon positions are informative. Additional evidence in favor of the use of 3rd codon positions comes from ILD tests showing that there is no incongruence among the three codon positions, for each gene separately or for their combination.

The results of the separate analyses of amd and Ddc are fully encompassed in their combined analysis and, therefore, are not presented. We will mention only that in the NJ analysis of amd, D. tripunctata and D. purpura, which were not studied for Ddc, cluster with D. funebris and D. phalerata with strong bootstrap support of 98%: the relationships among all four species are not resolved. D. immigrans is their closest sister taxon (bootstrap 71%). Each of the three species groups of Sophophora is strongly defined on the amd tree (bootstrap 100%). Taking into account the well-established monophyly of these groups, we obtained longer Ddc sequences for only one species from each of the melanogaster and obscura groups of the subgenus Sophophora, but three species of the willistoni group, because the 3rd codon nucleotide composition in species of the willistoni group is rather distinct from that of the other Drosophila.

The amd and Ddc data partitions are not incongruent, according to the ILD tests (Farris et al., 1994), whether based on all nucleotides or on the 1st + 2nd codon positions only. The NJ tree based on the analysis of all amd + Ddc nucleotides is presented in Fig. 1, with bootstrap values above branches for all codon positions and below branches for 1st + 2nd codon positions. The
first split is between the subgenus Sophophora and all other species. In the tree based on the 1st + 2nd codon positions both clades are strongly supported, whereas support is only moderate when all positions are used. These two major clades also emerge on the maximum-parsimony tree based on the 1st and 2nd codon positions, although with somewhat lesser support than that on the corresponding NJ tree.

The Hawaiian drosophila and Scaptomyza are monophyletic. The well-defined pair virilis/robusta, together with hydei, appear as the closest sister taxa to the Hawaiian/Scaptomyza in all analyses, but without strong support. Samoaia, D. Immigrans, D. phalerata, and D. funebris form a well-outlined group (85% based on 1st + 2nd codon positions, 69% based on all nucleotides). Relationships among the previously noted clusters and the other taxa are not well resolved.

To obtain better-resolved phylogenies, we used two additional genes, Adh and Sod, which have been sequenced in many of the species that we have studied (Table 1). For Adh we use D. sordidula rather than D. robusta, D. Chilea rather than D. phalerata, and D. hawaiiensis rather than D. gymnobasis. For Sod we use D. guttifera rather than D. phalerata. We have conducted separate analyses for each gene, combined analyses, and examined the data for heterogeneity with the ILD test. According to this test, the phylogenetic signals present in each of the genes that we have investigated are largely congruent. The general conclusion is that whereas some genes can be particu-

<table>
<thead>
<tr>
<th>Genus</th>
<th>Subgenus</th>
<th>Group</th>
<th>Species</th>
<th>GenBank Accession No.</th>
<th>amd</th>
<th>Ddc</th>
<th>Adh</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drosophila</td>
<td>Sophophora</td>
<td>melanogaster</td>
<td>melanogaster*</td>
<td>X04695</td>
<td>AF091328</td>
<td>X78384</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>simulans</td>
<td></td>
<td>AF293726</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>tassieri</td>
<td></td>
<td>AF293727</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>erecta</td>
<td></td>
<td>AF293708</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>obscura</td>
<td>bifasciata</td>
<td>AF293705</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>bogotana</td>
<td>AF293706</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>persimilis</td>
<td>AF293720</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>pseudoobscura*</td>
<td>AF293722</td>
<td>AF293746</td>
<td>X62181</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>willistoni</td>
<td>paulistorum*</td>
<td>AF293719</td>
<td>AF293744</td>
<td>AB026529</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>willistoni*</td>
<td>AF293730</td>
<td>AF293750</td>
<td>L08648</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>nebulous*</td>
<td>AF293717</td>
<td>AF293742</td>
<td>DNU95275</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>virilis</td>
<td>virilis*</td>
<td>AF293729</td>
<td>AF293749</td>
<td>AB033640</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>robusta</td>
<td>robusta</td>
<td>AF293724</td>
<td>AF293747</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>sordidula</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>hydei*</td>
<td></td>
<td>AF293712</td>
<td>AF293737</td>
<td>X58694</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>immigrants</td>
<td>immigrants*</td>
<td>AF293713</td>
<td>AF293738</td>
<td>M97638</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>testacea</td>
<td>putrida</td>
<td>AF293723</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>quinaria</td>
<td>phalerata</td>
<td>AF293721</td>
<td>AF293745</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>brachinpectera</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>funebris</td>
<td>funebris</td>
<td>AF293709</td>
<td>AF293734</td>
<td>AB033643</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>tripectata</td>
<td></td>
<td>AF293728</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AF293710</td>
<td>AF293735</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>hawaiensis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>mima*</td>
<td></td>
<td>AF293716</td>
<td>AF293741</td>
<td>M60792</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>adusta*</td>
<td></td>
<td>AF293718</td>
<td>AF293743</td>
<td>AB033649–AB033651</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AF293710</td>
<td>AF293732</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AF293704</td>
<td>AF293732</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AF293706</td>
<td>AF293732</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AF293717</td>
<td>AF293732</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AF293709</td>
<td>AF293732</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AF293711</td>
<td>AF293736</td>
<td>AB026530</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AF293707</td>
<td>AF293733</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AF293731</td>
<td>AF293751</td>
<td>X63955</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AF293715</td>
<td>AF293740</td>
<td>AB033655–AB033657</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AF293725</td>
<td>AF293748</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AF293714</td>
<td>AF293739</td>
<td>M97637</td>
<td></td>
</tr>
<tr>
<td>Scaptomyza b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hirtodrosophila</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dorosilopa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zapronius a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liodrosophila b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Samoaia a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scaptodrosophila b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Newly obtained and and Ddc sequences are underlined. All Ddc sequences have been extended from 966 bp (Tatarov et al., 1999a) to 1134 bp.

a Asterisks indicate species analyzed for Sod; GenBank accession numbers are given in Kwiatowski and Ayala (1999).

b Scaptodrosophila is classified by Wheeler (1981) as a subgenus of Drosophila, but has been raised to genus by Grimaldi (1990). Scaptomyza, Zapronius, Liodrosophila, and Samoaia are classified as genera by Wheeler (1981); in this paper we shall refer to them, and to Hirtodrosophila and Dorosilopa, as subgenera within the genus Drosophila.

c Hawaiian picture-winged (HPW) and modified mouth parts (MMP) are groups of Hawaiian drosophilids.
The combined analysis of the three genes yields strong support for the early branching (95%) and monophyly (85%) of Sophophora. Addition of Sod further increases bootstrap values to 98 and 89%. Moreover, the monophyly of Sophophora is strongly supported by one codon deletion in Ddc in all Sophophora species (Tatarenkov et al., 1999a).

Similarly, the position of Zapriouus as the outgroup to all non-Sophophora species becomes strongly supported on the combined trees, whereas its position is not definite on single gene trees.

Some clusters of species receive high bootstrap support in each of the separate analyses. In the NJ analysis of amd, there is strong support (bootstrap 98%) for the clade comprising D. putrida (testacea group), D. funebris (funebris), D. phalerata (quinaria), and D. tripunctata (tripunctata). Ddc and Adh were not studied in species of the testacea and tripunctata groups, but funebris and representatives of the quinaria group cluster together with very high support for each gene (100 and 99%). As expected, the cluster of D. funebris and D. phalerata/brachinephros is strongly supported in the combined analysis of the three genes. This finding is significant because it resolves the previously uncertain position of D. funebris. Throckmorton (1975) places it at the base of the Drosophila radiation. According to his scheme, after the split of Sophophora, the funebris group was among the first lineages (the other is the genus Liodrosophila) to separate from the rest of Drosophila. But according to Grimaldi (1990), the funebris group occupies a more derived position on the tree within the subgenus Drosophila. D. immigants appears as the sister group to the cluster of species just discussed, in the analysis of Adh (56%) and amd (71%). When sequences from the three genes are combined, bootstrap support for the cluster comprising D. immigants, D. phalerata/brachinephros, and D. funebris becomes 98%. Grouping of D. immigants with the quinaria, tripunctata, and funebris groups is also suggested by the 28S rRNA gene (Pélandakis and Solignac, 1993), although without statistical support (~50%).

The separate analysis of the three genes yields in all cases a clade comprising Scaptomyza and the Hawaiian Drosophila (represented here by the picture-winged and modified-mouth-parts groups), with support which is strong in Ddc (92%), moderate in amd (78%), and weak in Adh (65%). When the three genes are combined, this cluster receives support of 99%. Throckmorton (1975) considered the Hawaiian groups of Drosophila and Scaptomyza sister taxa and monophyletic with respect to other Drosophila groups, whereas Grimaldi (1990) proposed that the Hawaiian Drosophila together with Hirtodrosophila represent an early offshoot in the subfamily Drosophilinae and...
raised the Hawaiian Drosophila to generic status (Id-iomyia). All molecular data indicate that Scaptomyza is the closest sister taxon to the Hawaiian Drosophila.

An unexpected early result from molecular studies was that the closest sister group to the cluster Scaptomyza/Hawaiian Drosophila was the virilis-repleta lineage [shown in the analysis of Adh by Russo et al. (1995) and Tamura et al. (1995)]. How do other genes support this hypothesis? Separate analyses of amd and Ddc are not informative on this matter, but in the combined analysis of these genes, the Hawaiian Drosophilids do cluster together with the virilis-repleta groups with weak support, and when Sod is added the support becomes 77% (Sod separately supports this cluster only weakly, at 22%). It thus appears that the clade Hawaiian Drosophila/Scaptomyza is indeed the sister taxon to the virilis-repleta lineage. Although the strongest evidence for this comes from Adh, the total evidence from three other genes also supports it. When amd, Ddc, and Adh are pooled, the cluster of the Hawaiian drosophilids and the virilis-repleta lineage achieves bootstrap support of 96%.

The positions of Liodrosophila (not studied for Sod), Hirtodrosophila, and Dorsilopha (not studied for Adh) are not definite and require additional study. Analysis of amd, Ddc, and Adh clearly indicates that Liodrosophila and Hirtodrosophila are derived with regard to Zapronius. On the other hand, in the combined analysis of amd, Ddc, and Sod, Hirtodrosophila, together with other non-Sophophora species, appears in a derived position with regard to both Dorsilopha and Zapronius. Tentative relationships among these taxa are shown in Fig. 2.

No single gene has yet produced an unequivocal phylogeny of the Drosophilidae. Instead, the pooling together of data sets from several genes seems promising. The analysis of new amd and Ddc sequences from all major radiations (sensu Throckmorton, 1975), together with the previously studied Adh and Sod, provides a reasonably detailed resolution of phylogenetic relationships of Drosophilidae (Fig. 2) that may serve as a working hypothesis in future studies.

**ACKNOWLEDGEMENTS**

We thank Mariela Macas for technical help. The National Drosophila Species Stock Center in Bowling Green, Ohio kindly provided drosophilid species. Research was supported by NIH Grant GM42397 to F.J.A.

**REFERENCES**


Andrey Tatarenkov*  
Martina Žurovčová†  
Francisco J. Ayala*

*Department of Ecology and Evolutionary Biology, 321 Steinhaus Hall, University of California, Irvine, California 92697-2525  
†Institute of Entomology, Czech Academy of Sciences, Branišovská 31, 370 05 České Budějovice, Czech Republic  
doi:10.1006/mpev.2001.0967