ARTICLE IN PRESS

Available online at www.sciencedirect.com
SCIENCE DIRECT®

Molecular Phylogenetics and Evolution xxx (2005) xxx–xxx

MOLECULAR PHYLOGENETICS AND EVOLUTION

www.elsevier.com/locate/ympev

The phylogeny of the subgroups within the *melanogaster* species group: Likelihood tests on *COI* and *COII* sequences and a Bayesian estimate of phylogeny

Rebecca L. Lewis, Andrew T. Beckenbach, Arne Ø. Mooers *

Simon Fraser University, 8888 University Drive, Burnaby, BC, Canada V5A 1S6

Received 4 September 2004; revised 12 January 2005

Abstract

The relationships among the majority of the subgroups in the *Drosophila melanogaster* species group remain unresolved. We present a 2223 basepair dataset for mitochondrial *cytochrome oxidase I* and *cytochrome oxidase II* for 43 species (including new data from 11 species), sampled to include the major subgroups. After a brief review of competing hypotheses for the *ananassae*, *montium*, *suzukii*, and *takahashii* subgroups, we combine the two genes based on a new use of the SH test and present KH and SH likelihood comparisons (Kishino and Hasegawa, 1989. J. Mol. Evol. 29, 170–179; Shimodaira and Hasegawa, 1999) to test the monophyly and placement of these subgroups within the larger species group. Although we find insignificant differences between the two suggested placements for the *ananassae* subgroup, the *ananassae* is sister to the rest of the subgroups in the *melanogaster* species group in every investigation. For the *takahashii* subgroup, although we cannot reject monophyly, the species are so closely related to the *suzukii* subgroup for these data that the two subgroups often form one clade. Finally, we present a Bayesian estimate of the phylogeny for both genes combined, utilizing a recently published method that allows for different models of evolution for different sites.

Keywords: *Drosophila melanogaster*; *Takahashii*; *Suzukii*; *Ananassae*; *Montium*; COI; COII; Phylogeny; Likelihood; Ratchet; Bayesphylogenies

1. Introduction

Drosophila melanogaster, a member of the order *Diptera* (two-winged flies), is a diversely studied species in many fields of biology and has served as a model organism in, for example, both developmental biology and evolutionary genetic analysis (see, e.g., Presgraves et al., 2003; Sempere et al., 2003). This species was also the second metazoan to have its genome completely sequenced (Adams et al., 2000), and is also a model species for many behavioral investigations (see, e.g., Dukas and Mooers, 2003; Friberg and Arnqvist, 2003). However, *D. melanogaster* is not the only member of the *Drosophila* genus to be extensively studied: *Drosophila yakuba* was the first species to have its mitochondrion completely sequenced (Clary and Wolstenholme, 1985) and now serves as the standard for comparison when sequencing and aligning mitochondrial sequences. Extensive studies into the phylogenies of the *obscura* (Beckenbach et al., 1993; Gleason et al., 1997), *saltans* (O'Grady et al., 1998), and *willistoni* (Gleason et al., 1998) species groups (which are all members of the *Sophophora* subgenus) have been reported over the years; however, as of yet, there is no complete phylogeny for *melanogaster* species group.

Here, we examine four subgroups within the *melanogaster* species group whose placement and clade status is the subject of ongoing work: the *ananassae*, *montium*,

* Corresponding author.
E-mail address: amooers@sfu.ca (A.Ø. Mooers).

1055-7903/$ - see front matter © 2005 Published by Elsevier Inc.
doi:10.1016/j.ympev.2005.02.018

suzukii, and *takahashii* subgroups. To test the monophyly of the these subgroups and their placement within the phylogenetic tree, we sequenced *cytochrome oxidase I* and *cytochrome oxidase II* for 11 species within the *melanogaster* species group and combined them with sequences for 31 other species obtained from GenBank (summarized in Appendix A). We propose a novel test that compares the fit of data to topologies generated by different models on the same genes to determine whether genes can be combined into one molecular evolution model. We test specific hypotheses for monophyly of all subgroups, and for placement of these four clades using SH and KH likelihood comparison tests (Kishino and Hasegawa, 1989; Shimodaira and Hasegawa, 1999). We also present a Bayesian estimate of the phylogeny using Bayesphylogenies (Pagel and Meade, 2004). Bayesphylogenies is a unique program that allows the combination of multiple genes and models into one dataset without a priori partitions (methods that allow likelihood models on predefined partitions have also been described: see, e.g., Nylander et al., 2004). However, our focus is on hypothesis testing, and we first present a review of competing hypotheses for the placement of the *ananassae*, *montium*, *suzukii*, and *takahashii* subgroups.

1.1. The ananassae and montium subgroups

Over the last 30 years, six studies have examined the phylogenetics of the *ananassae* subgroup. Their findings are briefly presented below and summarized in Table 1.

Bock and Wheeler (1972) described species within each of the subgroups and proposed relationships among these groups based on morphological features, though these proposed relationships were not based on any formal phylogenetic analysis. Their study did propose a phylogeny composed of two lineages: *ananassae* plus *montium* subgroups as one lineage and the (*eugracilis*, *ficusphila*, *suzukii*, *takahashii*, and *melanogaster*) as the second. Ashburner et al. (1984) presented data that supported these findings, although they separated the *ananassae* and *montium* subgroups into two separate lineages to form three clades: (1) *ananassae* subgroup, (2) *montium* subgroup, and the (3) *melanogaster* + *elegans* + *eugracilis* + *ficusphila* + *suzukii* + *takahashii* subgroups (all species subgroups other than the *melanogaster* species subgroup within this third lineage will be referred to as the "oriental cluster" in this paper).

Péladakis et al. (1991) performed the first major sequencing study of the *melanogaster* species group. This was extended by Péladakis and Solignac (1993) who investigated the largest molecular dataset to date for this subgenus. Using 28s rRNA sequences for 72 species, and both distance and parsimony analyses, their results suggested three different lineages: (1) the *obscura* plus *ananassae* subgroups, (2) *montium* subgroup, and (3) the *melanogaster* subgroup plus oriental cluster. Importantly, because of the weak bootstrap support and the polytomy at the base of their consensus MP tree, they inferred the relationship among the three lineages from a neighbor-joining tree, placing the root between the melanogaster/oriental cluster and the other three.

Inomata et al. (1997) subsequently sequenced the *amylase* multigene family for 19 members of the subgenus. They found the *ananassae* subgroup to be monophyletic and the sister group to all the other lineages within the *melanogaster* species group. However, the placement of this clade in their phylogenetic tree affected

Table 1
Previous phylogenetic comparisons

Subgroup	Study	Proposed phylogenetic relationship	Data studied	Analysis[a]
ananassae	Bock and Wheeler (1972)	(*obscura*,((*ananassae*, *montium*), (*melanogaster*, oriental cluster)))	Morphology	N/A
	Ashburner et al. (1984)	(*obscura*, (*ananassae*, *montium*, (*melanogaster*, oriental cluster)))	Chromosomes, *Adh*, morphology	PC
	Péladakis and Solignac (1993)	(((*obscura*, *ananassae*), *montium*), (*melanogaster*, oriental cluster))	*28s rRNA*	MP, NJ
	Inomata et al. (1997)	(*obscura*, (*ananassae*, (*eugracilis*, (*melanogaster*, (*montium*, oriental cluster)))))	*Amy* Multigene	NJ
	Goto and Kimura (2001)	(*obscura*, (*ananassae*, (*montium*, (*melanogaster*, oriental cluster))))	*Gdph, COI*	MP, ML, NJ
	Schawaroch (2002)	(*obscura*, ((*melanogaster*, oriental cluster), (*ananassae*, *montium*)))	*COII*, *Adh*, *hb*	MP
takahashii	Bock (1980)	(*takahashii*, *melanogaster*)	Morphology	N/A
	Ashburner et al. (1984)	(*takahashii*, *melanogaster*)	Chromosomes, *Adh*, morphology	N/A
	Inomata et al. (1997)	(*takahashii, elegans*)	*Amy* Multigene	NJ
	Goto and Kimura (2001)	(*takahashii*, *suzukii*)	*Gdph* and *COI*	MP, ML, NJ
	Schawaroch (2002)	(*takahashii*, *suzukii*)	*COII*, *Adh* and *hb*	MP

[a] MP, maximum parsimony; ML, maximum likelihood; NJ, neighbor-joining (distance); PC, unspecified phenetic clustering algorithm; N/A, no quantitative analysis.

the bootstrap support for the *melanogaster* subgroup and the oriental cluster: only when the *ananassae* clade was removed from their analysis did the tree show robust support for the *melanogaster* subgroup and oriental cluster. Only one species was sequenced from the *montium* subgroup, (*D. auraria*); they found that this species was the sister to a cluster of *D. takahashii* and *D. elegans*.

Goto and Kimura, in 2001, presented exhaustive experiments of phylogenetic applications of algorithms on molecular data, with the *melanogaster* species group as their test case. For their *Gpdh* investigations, the *ananassae* subgroup was monophyletic and the sister group to all of the other subgroups within the *melanogaster* species group. However, for the *COI* dataset, there was no support for *ananassae* subgroup monophyly. *D. ananassae* was actually placed within the *montium* subgroup for the COI analyses. For their combined dataset, there was support for both the monophyly and placement of the *ananassae* subgroup, as sister group to the rest of the subgroups (as seen in the *Gpdh* trees). The combined dataset supported a monophyletic *montium*, sister to the *melanogaster* plus oriental subgroups. Importantly, though these trees were different from those generated from *COI*, they could not be rejected by the *COI* data.

Schawaroch published a total evidence investigation (Kluge, 1989) in 2002 for three genes (one mitochondrial and two nuclear) and 43 species within the *melanogaster* species group. She provided support for the monophyly of both the *ananassae* and the *montium* subgroups. Contrary to previous findings (Ashburner et al., 1984; Bock, 1980; Goto and Kimura, 2001; Inomata et al., 1997; Péladakis and Solignac, 1993) her results suggested an alternative rooting of the tree, one that placed the *ananassae* and *montium* subgroups as sister groups, with this clade being sister to the *melanogaster* subgroup plus the oriental cluster.

Comparing these investigations (Table 1) is difficult because different species were used in each phylogenetic analysis. Indeed, for some cases (Ashburner et al., 1984; Goto and Kimura, 2001) only one or two species were used, such that tests of monophyly are weak or impossible. Therefore, to better test both the monophyly and the placement of these clades we examined the largest relevant dataset to date: seven species of *ananassae* subgroup and seven of *montium* subgroup.

1.2. The takahashii and suzukii subgroups

The *takahashii* subgroup's position within the phylogenetic tree of the *melanogaster* species group (summarized in Table 1) is also controversial. Bock (1980) hypothesized that the *takahashii* subgroup was the sister group to the *melanogaster* subgroup; he also moved one species, *D. tanorum*, from the *montium* subgroup to the *takahashii* subgroup. This movement identified similarities not only between the *takahashii* and the *melanogaster* subgroups, but also between the *takahashii* and the *montium* subgroups. Following this, the analysis of the polytene chromosomes by Ashburner et al. (1984) "attested a closer relationship between [the *melanogaster* and the *takahashii*] subgroups than with either the *montium* or the *ananassae* subgroups." But Bock (1980) also found similarities between the *takahashii* and the *suzukii* species. He noted that several of the *takahashii* species had similar wing dimorphism with four *suzukii* species (*D. biarmipes*, *D. pulchrella*, *D. suzukii*, and *D. tristipennis*) as well as some similarities in male genitalia.

Inomata et al. (1997) sequenced only one species of the *takahashii* subgroup, and found that it grouped as the sister to members of the *elegans* subgroup (though this placement had very low bootstrap value). They did not sequence any species for the *suzukii* subgroup. Goto and Kimura (2001) found through all their investigations a very high bootstrap value for monophyly for their four *takahashii* species. Since they only sequenced *D. suzukii*, monophyly tests for this group could not be done. In all their tests, they found that the *takahashii* species and *D. suzukii* were sister groups, though the relationships within the *takahashii* clade varied.

Finally, Schawaroch (2002) supported the monophyly for the same four *takahashii* species as Goto and Kimura, but with low bootstrap values. She sequenced three species for the *suzukii*, and found that they were polyphyletic. Schawaroch's single most parsimonious tree made the *takahashii* subgroup and *D. mimetica* sister groups. She suggested *D. biarmipes* as the sister to the *melanogaster* and *eugracilis* subgroups, and *D. lucipennis* as the sister to *D. elegans*. Her tree also presented new relationships between the four species within the *takahashii* subgroup. She grouped *D. paralutea* with *D. prostipennis* and *D. takahashii* with *D. lutescens*. This contradicted the findings from offspring hybridization tests performed on these species (Bock and Wheeler, 1972; Lemeunier et al., 1986; Watanabe and Kawanisi, 1983).

As with the *ananassae* and *montium* subgroups, studies of the *takahashii* and *suzukii* subgroups have also been based on small sample sizes. Our study examines the relationship between at least five species within each of the *takahashii* and *suzukii* subgroups, and tests both their monophyly and their placement within the tree.

2. Materials and methods

2.1. Data collection

For this experiment, adult flies were obtained from the Tucson Fly Stock Center for the following species: *D. pseudoananassae*, *D. orena*, *D. parabipectinata*, *D. pseudotakahashii*, *D. jambulina*, and *D. lacteicornis*, *D.*

Table 2
PCR primers used

Primer 1	Sequence (5′–3′)	Primer 2	Sequence (5′–3′)
TW-J1301	GTTAAWTAAACTAATARCCTTCAAA	C1-N2353	GCTCGTGTATCAACGTCTATWCC
C1-J2231	TACCTGGATTYGGRATRATTTC	C1-N2776	TAATCTGAATAACGTCGNGG
C1-J2636	ATAGGRGCWGTATTTGCYATTAT	C2-3389	TCATAACTTCAGTATCATTG
TL-J3034	TAATATGGCAGATTAGTGCA	TK-N3796	ACTATTAGATGGTTTAAGAG

bipectinata, "*D. ercepeae*," *D. malerkotliana*, *D. paralutea*, and *D. phaeopleura*. Their taxonomic placement is presented in Appendix A. DNA extraction was carried out as described in Beckenbach et al. (1993). Whole flies were killed by freezing, and then chemically digested in 100 µl of protease buffer (0.01 M Tris, pH 7.8, 5 mM EDTA, 0.5% SDS, and 50 ng/µl proteinase K) for 3 min at 65 °C. The protease solution was extracted with 100 µl Tris buffer saturated phenol then with 70 µl chloroform/isoamyl alcohol (24:1) wash. The resultant 50 µl extraction was precipitated in 125 µl of 95% ethanol overnight at −20 °C, and DNA pellets were collected through centrifugation at 40,000 rpm. The DNA pellets were washed with 100 µl of ice cold 70% ethanol then centrifuged. The pellets were allowed to dry at room temperature. The pellets were stored frozen for further use. At that time, the pellets were dissolved in 50 µl of double distilled water, and this became the template for PCR.

For all samples, amplifications were conducted using Qiagen *Taq* DNA polymerase enzyme (Qiagen). The primer sequences used for each PCR amplification are summarized in Table 2. All thermocycling reactions were performed in either the Labline Programmable Thermal Blok II or the Eppendorf Thermocycler. The protocol for thermocycling consisted of 40 cycles, each consisting of 95 °C denaturing (30 s), 50 °C annealing (30 s), and 72 °C elongation (90 s). The products were refined and isolated with the QIAquick PCR Purification Kit for direct sequencing. All samples were run on 0.8% agarose gels with TBE (Tris borate, EDTA) buffer at 70 mV until the 600 bp marker had traveled approximately 2 cm. This allowed approximate determination of the length of the DNA segments before they were sent for sequencing.

The final DNA sequencing was conducted at the DNA Sequencing Laboratory, University of Calgary Core DNA and Protein Services. Sequencing reactions were conducted using an ABI PRISM dye-terminator sequencing kit and resolved on an ABI PRISM 373S or 377XL sequencer.

2.2. Data analysis

Sequences fragments were initially aligned and combined by visual inspection in BioEdit version 5.0.9. The species from GenBank were selected for this study if they had both *cytochrome oxidase I* and *cytochrome oxidase II* sequences and belonged to the members of the melanogaster species group. Five species from the *obscura* and *willistoni* species groups were included to root the tree (though no rooting constraints were imposed). Species were eliminated from the dataset if their sequence did not overlap with the other species' sequences. Species were also added to the analysis if they were the only species available for the subgroup (even if they only had sequence for one of the genes). The *cytochrome oxidase II* sequences were 684 bp in length; however, some of the GenBank sequences were partial sequences, the shortest being 384 bp. Due to lack of introns and contiguous sequences, the alignment was straightforward. Missing data in the partials were coded as question marks, and ambiguous positions were coded according to IUPAC ambiguity symbols. Multiple alignments were performed using CLUSTALW version 1.81 (Thompson et al., 1994) pairwise DNA alignments with default parameters. This alignment determined approximately where the partial sequences "sat" on a complete alignment. However, this alignment was not used in any of the phylogenetic studies, as the alignment was then visually realigned in SeAl v2.0A11 (Rambaut, 1996). The *cytochrome oxidase I* dataset was more involved, as there were fewer complete sequences, and more small partials. The total length of this gene is 1539 bp. Species with two or more sequences were combined: for non-overlapping sequences, the two sequences were amalgamated into one sequence. When the two sequences overlapped, and the basepairs were not identical, the position was coded as ambiguous following IUPAC. All the *cytochrome oxidase I* sequences were then visually aligned in SeAl v2.0a11 (Rambaut, 1996.) No extra gaps were needed in the alignments of either gene.

For analysis, MODELTEST version 3.06 (Posada and Crandall, 1998) was used to determine the model of evolution of each gene. As well, the two genes were joined end to end, and a model of evolution was determined for the combined dataset. After assembly and alignment, a maximum of seven of the species with the largest datasets per subgroup were retained for final analyses.

For testing hypotheses, all topologies were obtained using portable PAUP v 4.0b10 (Swofford, 2002) on the Beowulf Cluster at Simon Fraser University. Accelerated likelihood explorations were performed using likelihood ratchets (Vos, 2003) in conjunction with PAUPR at (Sikes and Lewis, 2001) for more efficient searching of the likelihood surface. In brief, a likelihood ratchet

employs multiple sequential truncated searches on different starting trees created by fast algorithmic searches on reweighed data, in the hopes of exploring a larger proportion of tree space, analogous to the parsimony ratchet (Nixon, 1999). Like the latter, it has been shown to increase search efficiency substantially (Vos, 2003). We ran 200 iterations with a GTR + I + G likelihood model of evolution, and uniformly reweighing 15% of the dataset per iteration. A strict consensus of trees with the same likelihood was considered the best estimate of the gene phylogeny.

One straightforward approach to ascertaining whether one model fits a data partition better than another is to ask whether trees inferred under different models actually differ substantially. If they do not, this suggests the contrasting models have limited effect on inference of relatedness. We applied this logic to determine whether our two genes could be combined, using the following novel test. We first used ModelTest (Posada and Crandall, 1998) to determine the best model for each gene, and also for the two genes taken as a single partition. All three models were GTR + I + G with varying parameters for each dataset. We then did topology searches with the likelihood ratchet (Vos, 2003) and PAUP with default settings for each gene separately with two models: their respective best model from Modeltest as well as the best model returned when both genes were combined into a single partition. The best unconstrained tree for a gene (with its own preferred model) was then compared to the best unconstrained tree for the gene produced with the combined model using a test that asks if one tree is significantly better-fit by the data than another. Here, because one tree is a posteriori and one is a hypothesis, we would use a one-tailed SH topology test using RELL bootstraps with 10,000 replicates (Shimodaira and Hasegawa, 1999; see also Felsenstein, 2004, Chapter 21). This was repeated for each gene separately. For both genes, the combined model and the gene-specific models produced the same topology, strongly supporting the hypothesis that the two genes can be modeled as evolving with a common set of parameters. We note that the two models could have returned different topologies for the same data, and the SH topology test would have been used to ask if they were different enough to reject the poorer-fit model. The combined dataset (2223 positions) and the combined model were therefore used for subsequent analysis.

A second, novel approach to combining partitions has recently been presented by Pagel and Meade (2004). They introduce "Bayesphylogenies," which allows for a posteriori partitions in a dataset. The user specifies the number of partitions, and the search determines which sites are best fit by different model. We performed three different runs, all GTR + I + G, with 1,000,000 generations: (i) a MrBayes (Huelsenbeck and Ronquist, 2001) run with twelve parallel hot swapping chains and default priors (Altekar et al., 2004); (ii) a single pattern Bayesphylogenies run with six chains, corresponding to a single model for the entire dataset; (iii) a two pattern Bayesphylogenies run with six chains, corresponding to two separate sets of model parameters. We then compared these topologies to each other and the topology returned from the unconstrained likelihood ratchet search.

2.3. Hypothesis testing

To test our hypotheses, we employed the same likelihood-based paired sites tests we used to determine model choice above (see Felsenstein, 2004, ch. 21 for an extensive exposition). To test whether the mtDNA data were consistent with monophyly of each of the seven subgroups with more than one species present (Appendix A), we employed the one-tailed SH test (Shimodaira and Hasegawa, 1999), which compares the fit of the data for a priori hypotheses to the fit of the data for the ML tree (an a posteriori hypothesis). In our case, we tested whether trees that forced subgroups to be monophyletic were significantly worse than the best tree. Monophyly constraint searches in PAUP produced the best tree under each constraint and these trees were compared to the best unconstrained tree using RELL bootstraps with 10,000 replicates (these tests are outlined in Table 3). The data cannot reject monophyly for species subgroups that were monophyletic in the unconstrained ML tree, and so these were not tested. To test whether the mtDNA data could distinguish the placement of the *ananassae* subgroup relative to the others studies (Table 1), we employed the two-tailed KH topology tests using RELL bootstraps with 10,000 replicates (Kishino and Hasegawa, 1989). This test is directly analogous to the SH test, but is used to distinguish among *a priori* hypotheses rather than to compare a priori with a posteriori ones.

Constraints for phylogenetic tests were built using MacClade v4.03 PPC (Maddison and Madison, 2000) and translated into Newick format in Treeview by Rod Page (v. 1.6.6; http://taxonomy.zoology.gla.ac.uk/rod/treeview.html). For the *ananassae* subgroup hypothesis test, there were two constraint searches. For hypothesis

Table 3
A priori comparisons tested

Comparison	*P* value
((suzukii), remaining species) vs. unconstrained tree[a]	0.0014[b]
((takahashii), remaining species) vs. unconstrained tree[a]	0.28[b]
(obscura, (ananassae, (montium, melanogaster, oriental))) vs. (obscura, ((ananassae, montium), (melanogaster, oriental)))	0.52[c]

[a] Unconstrained tree is the 'single model' phylogeny depicted in Fig. 1.
[b] One-tailed SH topology tests using RELL bootstraps with 10,000 replicates (Shimodaira and Hasegawa, 1999).
[c] Two-tailed KH topology tests using RELL bootstraps with 10,000 replicates (Kishino and Hasegawa, 1989).

one (HYP1), that the *ananassae* subgroup is the sister group to the rest of the subgroups within the *melanogaster* species group, the constraint was: (obscura, (ananassae, (montium, melanogaster, oriental cluster))). For hypothesis two (HYP2), with an alternative rooting topology that places the *montium* and the *ananassae* subgroups as sister groups to each other with the root lying between their clade and the rest of the subgroups of the *melanogaster* species group, the constraint was: (obscura, ((ananassae, montium), (melanogaster, oriental cluster))). The best tree from each hypothesis was compared with a two-tailed KH topology tests using RELL bootstraps with 10,000 replicates (Kishino and Hasegawa, 1989).

3. Results

For all phylogenetic analyses, the model of evolution returned from MODELTEST version 3.06 (Posada and Crandall, 1998), was GTR + I + G with the following parameter values (parameter names in PAUP* notation): Base = (A = 0.3079 C = 0.1083 T = 0.1447 G = 0.4391); Nst = 6; Rmat = (AC = 2.8506 AG = 18.6492 AT = 14.8598 CG = 2.0260 CT = 74.4765 TG = 1); Rates = gamma with shape = 0.8440; and Pinvar = 0.5823. For the unconstrained tree, the average pairwise distance from the outgroup was 0.11 substitutions/site. The two mitochondrial genes showed strong signal, as expected (skew based on 10,000 random trees: $g_1 = -0.627$***); there were 1568 invariant sites and 422 parsimony-informative sites across the two genes.

A summary of our tested hypotheses and results are summarized in Table 3. For the *ananassae* subgroup, monophyly was rejected through a one-tailed SH test comparing the constrained monophyly tree to the unconstrained tree ($p < 0.0001$). However, all species save one in this clade were monophyletic in the unconstrained tree—the exception being "*D. ercepeae*," which grouped with *D. jambulina*. Because monophyly is required to constrain the topological placement of the subgroups when testing *ananassae* subgroup's position, "*D. ercepeae*" was removed from the dataset (making the diminished *ananassae* subgroup monophyletic) for hypothesis testing. Indeed, *D. ercepae* has been mislabeled at the Tucson Stock Centre, and is actually *D. greeni* (Schawaroch, 2000), making its position as sister to *D. jambulina* within the *montium* subgroup unproblematic (Fig. 1). No tests for monophyly were required for the *elegans*, *eugracilis*, *ficusphila*, *melanogaster*, *montium* and *rhopaloa* subgroups because these subgroups were monophyletic in the unconstrained tree (or these subgroups consisted of only one species—the *eugracilis* and *ficusphila*). Monophyly tests were also performed on the *suzukii*, and *takahashii*, and subgroups (which were not monophyletic in the unconstrained tree). The data rejected monophyly for the *suzukii* clade as well (one-tailed SH test, $p = 0.0014$), but this was not due to a single species. For the *takahashii*, monophyly could not be rejected ($p = 0.28$).

We then tested the placement of the *ananassae* subgroup. There was no significant difference between the two different rooting hypotheses (HYP1 and HYP2) at the base of the melanogaster species group ($p = 0.52$), although hypothesis one produced the better fit to the data. As well, there was no significant difference between either hypothesis and the unconstrained best tree. Interestingly, the unconstrained best tree was most similar to the hypothesis one tree (such that the *ananassae* clade was the sister group to the rest of the *melanogaster* species group) with rearrangement occurring in the smaller represented subgroups and the dissolution of the monophyly some of the subgroups. In the unconstrained tree, the *suzukii* species *D. mimetica* broke up the monophyly of the *takahashii* subgroup. (This is a consistent picture: in the *takahashii* subgroup monophyly constraint tree, *D. mimetica* became the sister group to the *takahashii* subgroup.)

For the three way comparison between the likelihood ratchet, MrBayes and Bayesphylogenies, we found that a two patterned Bayesphylogenies model performed substantially better than a single pattern model (Log Bayes Factor = 244; see Gelman et al., 2004). The ratchet likelihood search, the single pattern Bayesphylogenies, and the MrBayes run returned an identical consensus topology, whereas the two patterned Bayesphylogenies consensus topology moved two species, *D. kikkawai* and *D. pseudotakahashii*. These changes are outlined in Fig. 1.

4. Discussion

In this experiment, we examined a maximum of 2223 characters from two mitochondrial genes for 41 species. These data showed strong signal for relationship resolution at the level of the (sub)group. *Cytochrome oxidase I* and *cytochrome oxidase II* are widely used genes for phylogeny, and often provide strong signal both when alone or in combination with other genes.

We first employed a new method for testing the combination of two genes under one model of evolution. We used topology tests to compare the best tree from the individual models for each gene (the best model for that partition) to the best tree from the model for both genes combined. We employed one-tailed SH likelihood comparison tests (Shimodaira and Hasegawa, 1999) because it makes allowance for an a priori hypothesis to be tested against an a posteriori topology. Since the two models agreed on the same unconstrained topologies, this statistical test does not allow us to reject using the combined dataset and the global model. This simple approach may be of use to investigators as they await more sophisti-

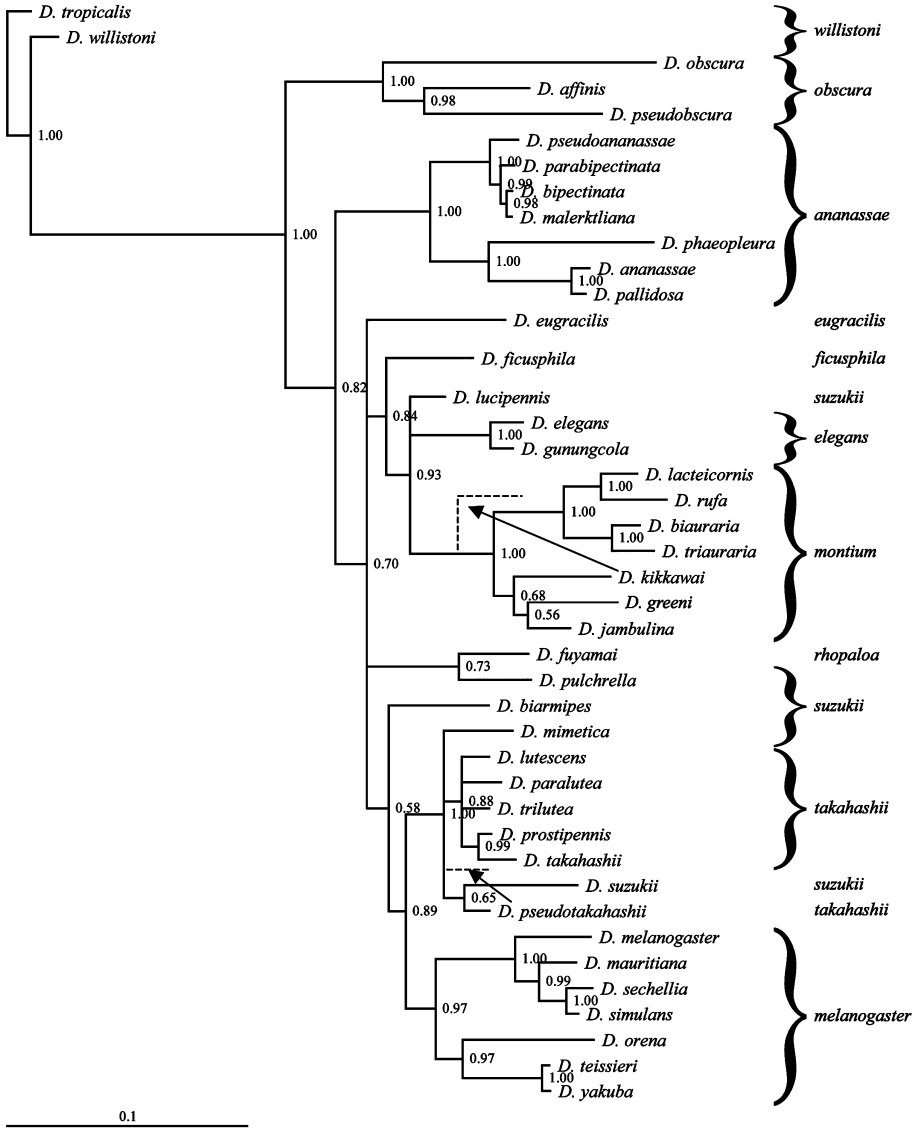

Fig. 1. Bayesphylogenies two matrix analysis of *cytochrome oxidase one* and *cytochrome oxidase two* combined dataset. The dashed lines denote the position of *D. kikkawai* and *D. pseudotakahashii* in the single model (MrBayes and Ratchet) analyses. Numbers at nodes indicate support values (posterior probabilities) from Bayesphylogenies analysis. The prior distributions used in the Bayesphylogenies search are as follows: all continuous parameters are drawn from a uniform on 1…100, while pattern heterogeneity weights and base frequencies are drawn from a dirichlet distribution; branch lengths are drawn from the exponential, while all topologies are considered equally likely a priori.

cated methods of combined data analysis (for a discussion, see Nylander et al., 2004).

4.1. The ananassae and montium subgroups

The dataset we collected could not statistically distinguish between the hypothesis one and hypothesis two for the *ananassae* subgroup. Schawaroch's (2002) single most parsimonious tree suggested the *ananassae* and the *montium* subgroups were sister groups (hypothesis two for this study), though the data analysis offered negligible support for this hypothesis (Bremer support = 1 (Bremer, 1988) and a bootstrap value ⩽ 50 (Felsenstein, 1985)). Though Schawaroch (2002) used similarities between these subgroup's male genitalia—both having surstylar and cercal claspers—as further support for their sister-clade status, male sex comb morphology is not consistent with this grouping. The *ananassae* subgroup species have highly varying sex-combs (from absent, to transverse, oblique or longitudinal row/rows) whereas almost all the *montium* species have a sex-comb in longitudinal rows of teeth along metatarsus and second tarsal segment (Bock, 1980; Lemeunier et al., 1986). Of course, these observations would require full morphological analysis to be phylogenetically meaningful. Although our data could not conclusively reject either

hypothesis, in our most likely unconstrained tree, the *ananassae* subgroup was sister group to the rest of the subgroups within the melanogaster species group (with a Bayesian posterior probability = 0.82, see below).

For the *montium* subgroup, the most likely unconstrained tree suggested no sister group for the *montium*, but a trichotomy between *D. lucipennis*—a *suzukii* subgroup species—and the *elegans* subgroup (Bayesian posterior probability = 0.93, see below). This grouping was unexpected, as there are no similarities in, e.g., sex-comb morphology (*D. lucipennis* lacks a sex-comb) and there are notable differences in male genitalia—*suzukii* species' male genitalia only have large primary claspers with several teeth whereas *montium* species have primary and secondary claspers (Bock, 1980).

4.2. The takahashii and suzukii subgroups

The data in this examination support the hypothesis that our members of the *takahashii* subgroup are monophyletic. The fact that many species of the *takahashii* subgroup have very similar external morphology strengthens this hypothesis (Lemeunier et al., 1986). Bock (1980) also reinforces this hypothesis with his description of this subgroup's consistent male genitalia and sex combs (though these species do show some interspecific variation in the number and arrangement of teeth). However, there are interesting results in the best-fit unconstrained tree. *Drosophila mimetica* from the *suzukii* subgroup consistently breaks up the monophyly of the *takahashii* subgroup and *D. suzukii* becomes this combined clade's sister group. One of the defining characteristic of the *suzukii* subgroup is the homogeneity of the male genitalia across species; however, they are more heterogeneous in external morphology (e.g., wing spots and sex-combs) than most other species subgroups (Bock, 1980). The two subgroups also have similarities in male genitalia: they both possess primary claspers, and rows of large teeth (Bock, 1980). Along with similarities in genitalia, four species of the *suzukii* subgroup (*D. pulchrella*, *D. biarmipes*, *D. suzukii*, and *D. tristipennis*) have similar wing dimorphism to the *takahashii* subgroup. One of these, *D. suzukii*, as previously stated, becomes the sister group to the *takahashii* subgroup plus *D. mimetica* clade. These similarities attest to the close relationship between the two subgroups, consistent with our mtDNA results and the results from previous studies (Goto and Kimura, 2001; Péladakis and Solignac, 1993; Schawaroch, 2002). Again, without a full morphological analysis, these observations are of limited systematic value. Also, more species need to be sequenced for more genes to further elucidate the relationships between the *suzukii* and the *takahashii* subgroups.

To organize our present dataset in anticipation of these further analyses, we present the Bayesphylogenies (Pagel and Meade, 2004) two-pattern consensus topology. This two-pattern Bayesphylogenies model performed substantially better than single pattern model (Log Bayes Factor = 244; according to Pagel and Meade (2004), a second matrix must contribute a minimum of 70–80 log units lower likelihood to be added for analysis.) We present this two pattern topology in Fig. 1. Although the new model increases the fit to the data, the consensus topology returned was the same as that from the conventional single model using either MrBayes or the ratchet likelihood search (save two nodes with low posterior probabilities). Importantly and consistent with our earlier results, the two matrix model did not distinguish between the two genes; sites that were fit better by the one of the other model were distributed across both genes.

We reiterate that the best supported relationship (*obscura*, (*ananassae*, (*montium*, (*melanogaster*, oriental group)))) is not significantly better than, e.g., the HYP2 alternative. The a posteriori support of 0.82 for this node is consistent with the recent suggestion that Bayesian support may be prone to Type 1 errors (Erixon et al., 2003; see also Simmons et al., 2004).

In this phylogenetic evaluation, we have provided additional evidence regarding the controversial phylogeny of the *melanogaster* species group using a large combined dataset in a likelihood framework. We hope the framework presented in Fig. 1. will serve as a template for future investigations.

Acknowledgments

We are very grateful for the use of the BEOWOLF Cluster at SFU for the analyses of our datasets and we thank Rutger Vos for direction (and navigation) through these analyses. We thank the FAB*-lab at SFU for discussion, two anonymous referees for very constructive comments, and especially Artyom Kropp for alerting us to the mislabeling of the Tucson Stock. This research was funded by an NSERC Canada grant to A.Ø.M.

Appendix A. Sophophora subgroup taxonomy and Sequence Accession Numbers

Species group	Subgroup	Species	COI accession	Reference	COII accession	Reference
melanogaster	*ananassae*	*ananassae*	AY098462.1 AF050748.1	Kopp and True (2002) Stenico and Nigro (1998)	AF474077.1	O'Grady and Kidwell (2002)

(*continued on next page*)

Appendix A (*continued*)

Species group	Subgroup	Species	COI accession	Reference	COII accession	Reference
		bipectinata	AY757287	This study	AY757275	This study
		malerkotliana	AY757289	This study	AY757277	This study
		pallidosa			AF461280.1	Schawaroch (2002)
		parabipectinata	AY757282	This study	AY757270	This study
		phaeopleura	AY757291	This study	AY757279	This study
		pseudoananassae	AY757280	This study	AY757268	This study
	elegans	*elegans*	AY098458.1	Kopp and True (2002)	AF461307.1	Schawaroch (2002)
		gunngcola	AB032129.1	Goto and Kimura (2001)		
	eugracilis	*eugracilis*	AY098461.1	Kopp and True (2002)	AF474079.1	O'Grady and Kidwell (2002)
			X58913.1	Nigro et al. (1991)		
	ficusphila	*ficusphila*	AB032133.1	Goto and Kimura (2001)	AF474080.1	O'Grady and Kidwell (2002)
	melanogaster	*mauritiana*	M57912.1	Satta and Takahata (1990)	AF474081.1	O'Grady and Kidwell (2002)
		melanogaster	AJ400907.1	Azou and Bregliano (2001)	AJ400907.1	Azou and Bregliano (2001)
		orena	AY757281	This study	AY757269	This study
		sechellia	M57908.1	Satta and Takahata (1990)		
		simulans	M57909.1	Satta and Takahata (1990)	AF474082.1	O'Grady and Kidwell (2002)
		tessieri	U51618.1	Gleason et al. (1997)	AF461282.1	Schawaroch (2002)
			AF050743.1	Stenico and Nigro (1998)		
		yakuba	X03240.1	Clary and Wolstenholme (1985)	X03240.1	Clary and Wolstenholme, 1985
	montium	*biauraria*	AB027259.1	Goto et al. (2000)	AF474084.1	O'Grady and Kidwell (2002)
		jambulina	AY757284	This study	AY757272	This study
		kikkawai	AF050746.1	Stenico and Nigro (1998)	AF461293.1	Schawaroch (2002)
		lacteicornis	AY757286	This study	AY757274	This study
		greeni	AY757288	This study	AY757276	This study
		rufa	AB027265.1	Goto et al. (2000)	AF461305.1	Schawaroch (2002)
		triauraria	AB027262.1	Goto et al. (2000)	AF474087.1	O'Grady and Kidwell (2002)
	rhopaloa	*fuyamai*	AY098460.1	Kopp and True (2002)		
	suzukii	*biarmipes*	AY098456.1	Kopp and True (2002)	AF474094.1	O'Grady and Kidwell (2002)
		lucipennis	AY098459.1	Kopp and True (2002)	AF461272.1	Schawaroch (2002)
		mimetica	AY098454.1	Kopp and True (2002)	AF474092.1	O'Grady and Kidwell (2002)
		pulchrella			AF474093.1	O'Grady and Kidwell (2002)
		suzukii	AB032128.1	Goto and Kimura (2001)		
	takahashii	*lutescens*	AB027267.1	Goto et al. (2000)	AF474090.1	O'Grady and Kidwell (2002)
		paralutea	AY757290	This study	AY757278	This study
		prostipennis	AB027266.1	Goto et al. (2000)	AF474091.1	O'Grady and Kidwell (2002)
		pseudotakahashii	AY757283	This study	AY757271	This study
		takahashii	X58915.1	Nigro et al. (1991)	AF474089.1	O'Grady and Kidwell (2002)
		trilutea	AB027261.1	Goto et al. (2000)		
obscura	*affinis*	*affinis*	AF519410.1	Perlman et al. (2003)	AF519346.1	Perlman et al. (2003)
			U51604.1	Gleason et al. (1997)		
	obscura	*obscura*	U51614.1	Gleason et al. (1997)	AF081356.1	O'Grady (1999)
	pseduoobscura	*pseudoobscura*	AF519412.1	Perlman et al. (2003)	AF519348.1	Perlman et al. (2003)
			U51607.1	Gleason et al. (1997)		
willistoni	*willistoni*	*tropicalis*	U51601.1	Gleason et al. (1997)	AF474103.1	O'Grady and Kidwell (2002)
	willistoni	*willistoni*	U51590.1	Gleason et al. (1997)	AF474104.1	O'Grady and Kidwell (2002)

References

Adams, M.D., Celniker, S.E., Holt, R.A., et al. (195 co-authors), 2000. The genome sequence of *Drosophila melanogaster*. Science 287, 2185–2195.

Altekar, G., Dwarkadas, S., Huelsenbeck, J.P., Ronquist, F., 2004. Parallel metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. Bioinformatics 20, 407–415.

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

Azou, Y., Bregliano, J.C., 2001. I-R system of hybrid dysgenesis in *Drosophila melanogaster*: analysis of the mitochondrial DNA in reactive strains exhibiting different potentials or I factor transposition. Heredity 86, 110–116.

Beckenbach, A.T., Wei, Y.W., Liu, H., 1993. Relationship in the *Drosophila obscura* species group, inferred from mitochondrial *cytocrome oxidase II* sequences. Mol. Biol. Evol. 10, 619–634.

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

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

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

Clary, D.O., Wolstenholme, D.R., 1985. The mitochondrial DNA molecular of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. J. Mol. Evol. 22, 252–271.

Dukas, R., Mooers, A.O., 2003. Environmental enrichment improves mating success in fruit flies. Anim. Behav. 66, 741–49.

Erixon, P., Svennblad, B., Britton, T., Oxelman, B., 2003. Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. Syst. Biol. 52, 665–673.

Felsenstein, J., 1985. Confidence limits of phylogenies: an approach using the bootstrap. Evolution 39, 783–791.

Felsenstein, J., 2004. Inferring Phylogenies. Sinauer Associates, Sunderland, Massachusetts.

Friberg, U., Arnqvist, G., 2003. Fitness effects of female mate choice: Preferred males are detrimental for *Drosophila melanogaster* females. J. Evol. Biol. 16, 797–811.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2004. Bayesian Data Analysis. Chapman and Hall, New York.

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

Gleason, J.M., Griffith, E.C., Powel, J., 1998. A molecular phylogeny of the *Drosophila willistoni* group: conflicts between species concepts?. Evolution 52, 1093–1103.

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

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

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

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

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 Hominidea. J. Mol. Evol. 29, 170–179.

Kluge, A.G., 1989. A concern for evidence and a phylogenetic hypothesis of relationships among Epicrates (Boidae, Serpentes). Syst. Zool. 38, 7–25.

Kopp, A., True, J.R., 2002. Phylogeny of the oriental *Drosophila melanogaster* species group: a multilocus reconstruction. Syst. Biol. 51, 786–805.

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

Maddison, D.R., Madison, W.P., 2000. MacClade 4: Analysis of Phylogeny and Character Evolution, version 4.0. Sinauer Associates, Sunderland, Massachusetts.

Nigro, L., Solignac, M., Sharp, P.M., 1991. Mitochondrial DNA sequence divergence in the *melanogaster* and oriental species subgroups of *Drosophila*. J. Mol. Evol. 33, 156–162.

Nixon, K.C., 1999. The parsimony ratchet, a new method for rapid parsimony analysis. Cladistics 15, 407–414.

Nylander, J.A.A., Ronquist, F., Huelsenbeck, J.P., Nieves-Aldrey, J.L., 2004. Bayesian phylogenetic analysis of combined data. Syst. Biol. 53, 47–67.

O'Grady, P.M., Clark, J.B., 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.

O'Grady, P.M., 1999. Reevaluation of phylogeny in the *Drosophila obscura* species group based on combined analysis of nucleotide sequences. Mol. Phylogenet. Evol. 12, 124–139.

O'Grady, P.M., Kidwell, M.G., 2002. Phylogeny of the subgenus *Sophophora* (*Diptera*: *drosopilidae*) based on combined analysis of nuclear and mitochondrial sequences. Mol. Phylogenet. Evol 22, 442–453.

Pagel, M., Meade, A., 2004. A mixture model for detecting pattern-heterogeneity in gene sequence or character-state data. Syst. Biol. 53, 571–581.

Péladakis, M., Higgins, D.G., Solignac, M., 1991. Molecular phylogeny of the subgenus *Sophophora* of *Drosophila* derived from the large subunit of ribosomal RNA Sequences. Genetica 84, 87–94.

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

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

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

Presgraves, D.C., Balagopalan, L., Abmayr, S.M., Orr, H.A., 2003. Adaptive evolution drives divergence of a hybrid inviability gene between two species of *Drosophila*. Nature 423, 715–719.

Rambaut, A., 1996. Se-Al: Sequence Alignment Editor. Available from: <http://evolve.zoo.ox.ac.uk/>.

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

Schawaroch, V.A., 2000. Molecular phylogeny of the *Drosophila melanogaster* species group with special emphasis on the *montium* subgroup. Ph. D. Dissertation, City Univ. New York, New York.

Schawaroch, V., 2002. Phylogeny of a paradigm lineage: the *Drosophila melanogaster* species group (*Diptera*: *Drosophilidae*). Biol. J. Linn. Soc. 76, 21–37.

Sempere, L.F., Sokol, N.S., Dubrovsky, E.B., Berger, E.M., Ambros, V., 2003. Temporal regulation of *microRNA* expression *in Drosophila melanogaster* mediated by hormonal signals and broad-complex gene activity. Dev. Biol. 259, 9–18.

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

Sikes, D.S., Lewis, P.O., 2001. PAUPRat: A tool to implement Parsimony Ratchet searches using PAUP*, beta software, version 1.

Simmons, M.P., Pickett, K.M., Miya, M., 2004. How meaningful are Bayesian support values?. Mol. Biol. Evol. 21, 188–199.

Stenico, M., Nigro, L., 1998. Differences in evolutionary rates among amino acid classes in the mitochondrial genes *cytochrome oxidase I* and *cytochrome oxidase II* in *Drosophila*. Mol. Biol. Evol. 15, 777–778.

Swofford, D.L., 2002. PAUP* Phylogenetic analysis using parsimony (* and other methods) version 4.0b10. Sinauer Associates, Sunderland, Massachusetts.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acid Res. 22, 4673–4680.

Vos, R.A., 2003. Accelerated likelihood surface exploration: the likelihood ratchet. Syst. Biol. 52, 368–373.

Watanabe, T.K., Kawanisi, M., 1983. Stasipatric speciation in *Drosophila*. Jpn. J. Genet. 58, 269–274.