J Mol Evol (2003) 57:562–573
DOI: 10.1007/s00239-003-2510-x

JOURNAL OF MOLECULAR EVOLUTION

Molecular Phylogeny of the *Drosophila melanogaster* Species Subgroup

Wen-Ya Ko, Ryan M. David,* Hiroshi Akashi

Institute of Molecular Evolutionary Genetics and Department of Biology, The Pennsylvania State University, University Park, PA 16802, USA

Received: 8 November 2002 / Accepted: 2 June 2003

Abstract. Although molecular and phenotypic evolution have been studied extensively in *Drosophila melanogaster* and its close relatives, phylogenetic relationships within the *D. melanogaster* species subgroup remain unresolved. In particular, recent molecular studies have not converged on the branching orders of the *D. yakuba–D. teissieri* and *D. erecta–D. orena* species pairs relative to the *D. melanogaster–D. simulans–D. mauritiana–D. sechellia* species complex. Here, we reconstruct the phylogeny of the *melanogaster* species subgroup using DNA sequence data from four nuclear genes. We have employed "vectorette PCR" to obtain sequence data for orthologous regions of the *Alcohol dehydrogenase* (*Adh*), *Alcohol dehydrogenase related* (*Adhr*), *Glucose dehydrogenase* (*Gld*), and *rosy* (*ry*) genes (totaling 7164 bp) from six *melanogaster* subgroup species (*D. melanogaster*, *D. simulans*, *D. teissieri*, *D. yakuba*, *D. erecta*, and *D. orena*) and three species from subgroups outside the *melanogaster* species subgroup [*D. eugracilis* (*eugracilis* subgroup), *D. mimetica* (*suzukii* subgroup), and *D. lutescens* (*takahashii* subgroup)]. Relationships within the *D. simulans* complex are not addressed. Phylogenetic analyses employing maximum parsimony, neighbor-joining, and maximum likelihood methods strongly support a *D. yakuba–D. teissieri* and *D. erecta–D. orena* clade within the *melanogaster* species subgroup. *D. eugracilis* is grouped closer to the *melanogaster* subgroup than a *D. mimetica–D. lutescens* clade. This tree topology is supported by reconstructions employing simple (single parameter) and more complex (nonreversible) substitution models.

Key words: *Drosophila melanogaster* species subgroup — Phylogeny — Alcohol dehydrogenase — Alcohol dehydrogenase related — Glucose dehydrogenase — *rosy*

Introduction

Comparative studies of molecular and phenotypic evolution in *D. melanogaster* and its close relatives often rely on knowledge of phylogenetic relationships (Powell and DeSalle 1995; Sullivan et al. 2000). Recent studies of vertical and horizontal transfers of transposable elements (Jordan and McDonald 1998; Terzian et al. 2000; Gentile et al. 2001), evolution of gene expression patterns (Ross et al. 1994; Bonneton and Wegnez 1995), gene duplication and divergence (Bettencourt and Feder 2001; Parsch et al. 2001), and lineage-specific molecular evolution (Akashi 1995, 1996; Harr et al. 2000; Takano-Shimizu 1999, 2001) have not assumed the same topology among *melanogaster* subgroup species.

The *melanogaster* species group contains over 150 species classified into 10 or 11 species subgroups (Ashburner 1989; Powell 1997; Schawaroch 2002). Early morphological, chromosomal, and molecular studies assigned these subgroups to three major clades, (1) *ananassae*, (2) *montium*, and (3) *suzukki–takahashii–*

**Present address:* University of Texas Health Science Center at San Antonio, 7703 Floyd Curl Drive, San Antonio, TX 78284, USA
Correspondence to: H. Akashi, 208 Mueller Lab, Department of Biology, The Pennsylvania State University, University Park, PA 16802, USA; *email*: akashi@psu.edu

ficusphila–melanogaster–elegans–eugracilis (*melanogaster* + oriental subgroups) (Ashburner et al. 1984; Lemeunier et al. 1986; Pélandakis et al. 1991; Toda 1991; Pélandakis and Solignac 1993). The *melanogaster* species subgroup consists of nine species that appear to be of Afrotropical origin. The human commensals, *D. melanogaster* and *D. simulans*, are cosmopolitan in their distributions. *D. sechellia* and *D. mauritiana* are endemic island species closely related to *D. simulans* (Caccone et al. 1996). *D. teissieri* and *D. yakuba* have similar geographic ranges spreading from northwest to southeast Africa. *D. teissieri*, mainly a forest species, is more western distributed, while *D. yakuba*, an open field species, is found more often in eastern regions. *D. erecta* and *D. orena* are restricted to west central Africa (Lachaise et al. 1988). Finally, *D. santomea*, a close relative of *D. yakuba*, was recently discovered on São Tomé island in the Gulf of Guinea in West-equatorial Africa (Lachaise et al. 2000). Species within this subgroup can be distinguished from male genital structures and courtship songs (Tsaca and Bocquet 1976; Cowling and Burnet 1981).

Molecular phylogenetic studies have not converged on a tree topology for species in the *melanogaster* subgroup. In particular, the branching orders of the *D. yakuba–D. teissieri* and *D. erecta–D. orena* species pairs remain unresolved. Some studies group the *D. yakuba–D. teissieri* pair closest to the *melanogaster* species complex, which includes *D. melanogaster*, *D. simulans*, *D. sechellia*, *and D. mauritiana* (referred to as tree topology I) (Santamaria 1975; Eisses et al. 1979; Ohnishi et al. 1983; Solignac et al. 1986; Cariou 1987; Daïnou et al. 1987; Caccone et al. 1988; Jeffs et al. 1994; Russo et al. 1995; Shibata and Yamazaki 1995). Other studies support a *D. yakuba–D. teissieri* + *D. erecta–D. orena* clade (tree topology II) (Tsaca and Bocquet 1976; Barnes et al. 1978; Cowling and Burnet 1981; Lemeunier and Ashburner 1976, 1984). Lachaise et al. (1988) favored topology I, citing that most evidence supporting this topology reject topology II, while the evidence supporting topology II does not strongly refute topology I. Jeffs et al. (1994) obtained moderate bootstrap support (86%) for topology I using maximum likelihood analysis of *Adh* sequences. This appears to have led to widespread acceptance of this topology (Powell 1997; Takano-Shimizu 1999, 2001; Terzian et al. 2000; Bettencourt and Feder 2001). However, Schlötterer et al.'s (1994) maximum parsimony analyses of an internal transcribed spacer (ITS) region of ribosomal RNA gene placed *D. orena* closest to the *melanogaster* complex (denoted tree topology III) with 100% bootstrap support.

In this study, phylogenetic relationships were reconstructed among six *melanogaster* subgroup species (*D. melanogaster*, *D. simulans*, *D. teissieri*, *D. yakuba*, *D. erecta*, and *D. orena*) and three species outside the *melanogaster* subgroup, *D. eugracilis* (*eugracilis* subgroup), *D. mimetica* (*suzukii* subgroup), and *D. lutescens* (*takahashii* subgroup). We employed a "vectorette PCR"-based method to obtain sequence data for four genes (*Adh*, *Adhr*, *Gld*, *ry*; total, 7164 bp) from these nine species.

Sequences from *D. pseudoobscura* and *D. subobscura* (available from the public databases) were employed as outgroups. With these data, maximum parsimony, neighbor-joining, and maximum likelihood methods all strongly support a *D. yakuba–D. teissieri* + *D. erecta–D. orena* clade (tree topology II) within the *D. melanogaster* subgroup. Our analyses also support the placement of *D. eugracilis* closer to the *melanogaster* subgroup than a *D. mimetica–D. lutescens* clade.

Materials and Methods

Drosophila *Strains and DNA Sequences*

Strains of *D. simulans* (251.6), *D. teissieri* (257.0), *D. yakuba* (261.0), *D. orena* (245.0), *D. eugracilis* (451.3), *D. lutescens* (271.1), and *D. mimetica* (341.0), were obtained from The National *Drosophila* Species Resource Center (Department of Biological Sciences, Bowling Green State University) and the Tucson Stock Center (University of Arizona). *D. erecta* (S-18) was kindly provided by Michael Ashburner (Department of Genetics, University of Cambridge, Cambridge, UK).

The coding regions of *Alcohol dehydrogenase* (*Adh*; 771 bp, complete; *D. melanogaster* cytogenetic map, 35B3), *Alcohol dehydrogenase related* (*Adhr*; 840 bp, complete; 35B3), *rosy* (*ry*; 4005 bp, approx. 98% of the total coding sequence; 87D9), and *Glucose dehydrogenase* (*Gld*; 1548 bp, approx. 84% of the total coding sequence; 84D3) were employed in phylogenetic analyses (Fig. 1). Accession numbers for sequences obtained from the GenBank/EMBL/DDBJ nucleic acid sequence data banks are Z00030 (*Adh* + *Adhr*, *D. melanogaster*), X00607 (*Adh* + *Adhr*, *D. simulans*), X54118 (*Adh* + *Adhr*, *D. teissieri*), X54120 (*Adh*, partial, *D. yakuba*), X54116 (*Adh* + *Adhr*, partial, *D. erecta*), Z00032 (*Adh*, *D. orena*), M29298 (*Gld*, *D. melanogaster*), U63325 (*Gld*, *D. simulans*), Y00308 (*ry*, *D. melanogaster*), Y00602 (*Adh* + *Adhr*, *D. pseudoobscura*), M29299 (*Gld*, *D. pseudoobscura*), M33977 (*ry*, *D. pseudoobscura*), M55545 (*Adh* + *Adhr*, *D. subobscura*), AF025811 (*Gld*, *D. subobscura*), and AF058976 and AF058977 (*ry*, *D. subobscura*). Sequences obtained for this study can be found under accession numbers AY279322–AY279343.

Genomic DNA Extraction and Vectorette DNA Library Construction

We employed vectorette PCR to sequence orthologs of *D. melanogaster* genes in species from the *melanogaster* group. Vectorette PCR was designed for amplification of DNA regions adjacent to known regions and has been employed to amplify 5′ and 3′ ends of exons as well as long introns (Riley et al. 1990; Arnold and Hodgson 1991). To construct vectorette libraries, genomic DNA digestion with a restriction enzyme is followed by ligation of annealed synthetic oligonucleotides (vectorettes) (Fig. 2A). PCR amplification from these libraries employs a specific primer that recognizes a known, or conserved, region of interest and one primer that anneals to the vectorette. Vectorettes contain a central mismatched region, and vectorette PCR primers are designed to anneal to the *complement* of this region, which is

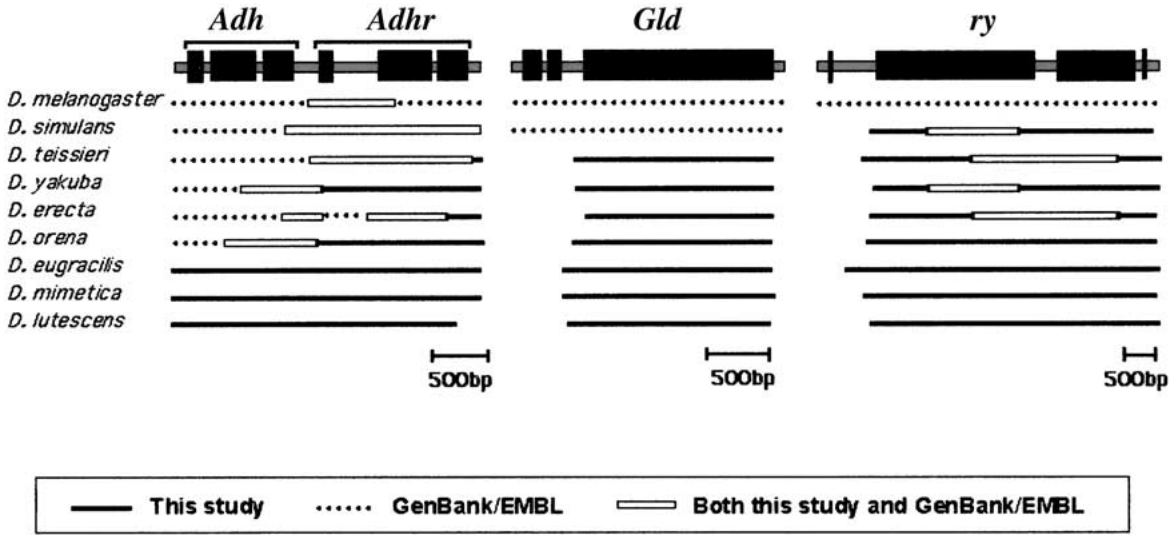

Fig. 1. Diagrams of the *Alcohol dehydrogenase* (*Adh*), *Alcohol dehydrogenase related* (*Adhr*), *Glucose dehydrogenase* (*Gld*), and rosy (*ry*) gene regions in *Drosophila melanogaster*. Sequences obtained for this study are shown.

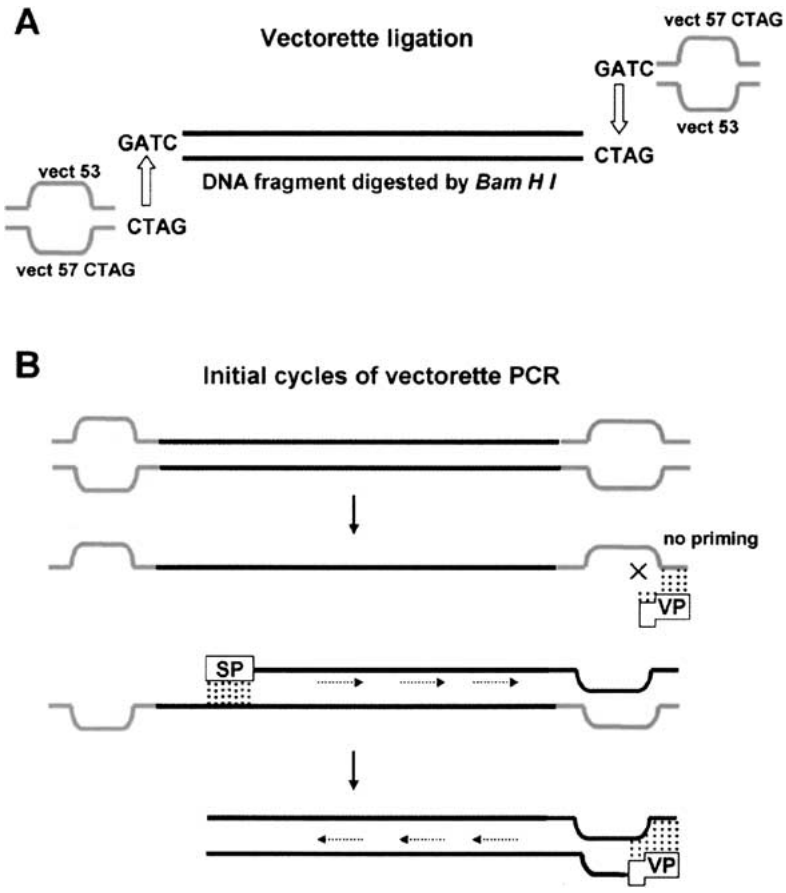

Fig. 2. Ligation of vectorette to digested genomic DNA (**A**). In the first round of amplification, primer extension only proceeds from the specific primer (SP) that hybridizes to the known region. Amplification from vectorettes ligated to nontarget regions does not occur because the vectorette primer only anneals to the *complement* of the bottom strand of the vectorette. In the second and subsequent rounds of PCR, amplification proceeds from both the specific primer (SP) and the vectorette primer (VP) (**B**).

synthesized from specific primers in the initial cycles of PCR (Fig. 2B) (Arnold and Hodgson 1991).

Vectorette sequences and library construction were modified from a protocol from the Botstein laboratory web site at Stanford University (http://genome-www.stanford.edu/group/botlab/index.html). We constructed 15 vectorette libraries for each species using a different restriction enzyme for each library. First, genomic DNA was extracted from ~100 adult flies. Files were homogenized in a 100 m*M*–Tris-HCl (pH 8.0), 50 m*M* EDTA, 100 m*M* NaCl, and 2% sodium dodecyl sulfate (SDS) solution with 40 µl of proteinase K (20 mg/ml; Qiagen). The solution was incubated at 55°C for 3 h followed by the addition of 8 µl of RNase (100 mg/ml; Qiagen) and further incubation at 55°C for 1 h. Phenol/chloroform and chloroform extractions were followed by ethanol precipitation, and genomic DNA was dried and resuspended with T.E. (10 m*M* Tris–Cl, pH 8.0, 0.1 m*M* EDTA).

Genomic DNA was then digested with restriction enzymes. Twenty micrograms of genomic DNA (in 200 µl T.E.) was digested

with 20 units of restriction enzyme according to the manufacturer's instructions [New England Bio Labs, Inc. (NEB)] for 10–16 h (10–16× overdigestion). After digestion, the enzyme was inactivated according to the manufacturer's instructions.

Vectorettes were constructed by annealing two partially complementary vectorette oligonucleotides (see supplementary material for sequences; and http://www.bio.psu.edu/people/faculty/akashi). A mixture of 0.02 μmol each of two oligonucleotides in 10 μl T.E. was incubated at 65°C for 5 min. $MgCl_2$ was added to a final concentration of 2 m*M*, and the vectorette mix was incubated at 65°C for another 5 min. After incubation, the mix was cooled slowly to room temperature to allow vectorette annealing. T.E. was added to bring the concentration of vectorettes to 1 m*M*.

Annealed vectorettes were constructed with four different overhanging sequences for ligation to digested genomic DNA. The four vectorette types share the vect 53 oligonucleotide but differ in the oligonucleotide that contains the complement of the 5′ overhangs left by the restriction enzymes. Vect 57 CTAG matches the overhang created by *Bam*HI, *Bcl*I, and *Bgl*II. Vect 57 GATC matches *Xba*I, *Nhe*I, and *Spe*I, vect 57 AATT matches *Eco*RI, *Apo*I, and *Mfe*I, and vect 57 GC matches *Bsa*HI, *Bst*BI, *Cla*I, *Taq*$^{\alpha}$I, *Hpa*II, and *Nar*I (see supplementary material; and http://www.bio.psu.edu/people/faculty/akashi/melsub_phy/suppl).

The appropriate vectorettes were ligated to digested genomic DNA with T4 DNA ligase (NEB). Vectorettes were annealed to genomic DNA by incubating a combination of 20 μl of 1 m*M* annealed vectorette mix and 20 μg of digested genomic DNA at 65°C for 5 min. The mixture was slowly cooled to room temperature. Then 800 U ligase, ligase buffer (NEB), and ATP (Sigma) were added to final concentrations of 1× buffer and 1 m*M* ATP and the ligation reaction was incubated at 16°C for 16 h. The ligase was then heat inactivated and the reaction mix was extracted with phenol/chloroform and chloroform. After ethanol precipitation, the DNA was dried and dissolved in T.E to final concentrations of 20 ng DNA/μl. Fifteen different vectorette DNA libraries were constructed for each of nine *Drosophila* species.

Vectorette PCR Amplification and DNA Sequencing

PCR primers were designed to recognize conserved regions between *D. melanogaster* and *D. pseudoobscura* sequences using the PrimerSelect program in the DNASTAR package (DNASTAR, Inc.). For a given specific primer, all 15 vectorette libraries plus 1 negative control (PCR reaction mix + primers without DNA) were employed in a single run of PCR amplification with a vectorette primer (see supplementary material; and http://www.bio.psu.edu/people/faculty/akashi/melsub_phy/suppl) for a given species. Each reaction was performed at three annealing temperatures. PCR cycles consisted of 1 cycle (95°C for 14 min) [to activate HotStarTaq (Qiagen)], 5 cycles (95°C for 1 min, 63–72°C for 1 min, 72°C for 2 min), 5 cycles (95°C for 1 min, 59–68°C for 1 min, 72°C for 2 min), 15 cycles (95°C for 45 s, 55–64% for 1 min, 72°C for 2 min), 15 cycles (95°C for 45 s, 51–60°C for 1 min, 72°C for 2 min), and 1 cycle (72°C for 10 min) [RoboCycler Gradient temperature cycler (Stratagene)]. For each reaction tube, 20 ng of vectorette DNA was added to a mixture containing 1× PCR buffer, 0.5 U HotStarTaq DNA polymerase (Qiagen), 200 μ*M* each dNTP (Sigma), and 0.25 μ*M* each primer (final volume, 20 μl). Initial PCR cycles employed higher annealing temperatures [approximately 10°C above the melting temperature (T_m) of the primer–template complexes] and the annealing temperature was gradually lowered in steps of 4°C. Early PCR cycles with high annealing temperatures are designed for specificity, while later cycles with reduced annealing temperatures produce greater yields. This is a modified form of "touchdown" PCR referred to as "step-down" PCR (Hecker and Roux 1996).

To obtain sufficient quantities of PCR fragments for DNA sequencing, some PCR products were reamplified using the initial specific primer and a second vectorette primer (see supplementary material; and http://www.bio.psu.edu/people/faculty/akashi/melsub_phy/suppl). Nonspecific binding can be reduced by using a vectorette primer internal to that used in the initial amplification. PCR products were purified using the Qiagen PCR purification and gel extraction kits. One hundred picomoles of amplification products was sequenced with either a specific primer or a vectorette primer using a Beckman CEQ 2000 automated DNA sequencer. All sequences were obtained on both strands. Protocols for vectorette library construction and PCR methods are available at our laboratory website (http://www.bio.psu.edu/people/faculty/akashi/vectpcr).

Table 1. GC content of gene sequences analyzed[a]

	GC content (%)	
	All sites	3rd codon
D. melanogaster	55.04	66.35
D. simulans	55.47	67.50
D. teissieri	56.59	70.44
D. yakuba	56.18	69.25
D. erecta	56.10	69.29
D. orena	56.37	70.36
D. eugracilis	51.23	57.24
D. mimetica	55.86	68.91
D. lutescens	56.40	70.83
D. pseudoobscura	58.77	74.96
D. subobscura	58.29	73.64

[a] Data are shown for the concatenated sequence of *Adh* + *Adhr* + *Gld* + *ry* (7164 bp).

Phylogenetic Analyses

Sequence alignment was carried out using the CLUSTAL algorithm (Higgins and Sharp 1988) in the MegAlign application (DNASTAR software package) and modified by eye. Phylogenetic trees were reconstructed using unweighted maximum parsimony (MP) (Fitch, 1971) with a branch and bound search, neighbor-joining distance (NJ) (Saitou and Nei 1987) with the Tamura–Nei (1993) model (TN), and maximum likelihood (ML) with the general time reversible (GTR) model (Felsenstein 1981). PAUP 4.0b08 (Swofford 2000) was used with 1000 bootstrap replicate tests (Felsenstein 1985) for each method. For the maximum likelihood method, parameters for nucleotide frequencies and proportions of invariant sites (*I*) were estimated from the data. Rates of substitution at variable sites were assumed to follow a discrete gamma distribution model (*G*) with four rate categories (Yang 1994) and the shape parameter (α) was estimated from the data.

For the maximum parsimony method, Bremer support (BS) (Bremer 1988, 1994) and partitioned Bremer support (PBS) (Baker and DeSalle 1997) were determined for each node in the phylogenetic tree. To calculate PBS, each data partition (gene) is first used to compute the tree length of the tree of interest (i.e., topology II, the maximum parsimony tree for the combined data set). Then the tree length is computed for the shortest tree constrained not to contain a node of interest. PBS values for each data partition are the numbers of extra steps in the length of the latter tree. The BS for each node is the sum of the PBS values. The TREEROT program (Sorenson 1999) was combined with PAUP for these analyses.

Phylogenetic inference under the TN and GTR + *G* + *I* models assumes stationarity of base composition among lineages. Analyses were also conducted using Galtier and Gouy's (1995, 1998) more complex substitution model that allows nonstationarity of base composition (see also Galtier et al. 1999; Tarrío et al. 2001). In our data, the GC content of *D. eugracilis* differs from that of other lineages in the *melanogaster* species group (Table 1). Galtier

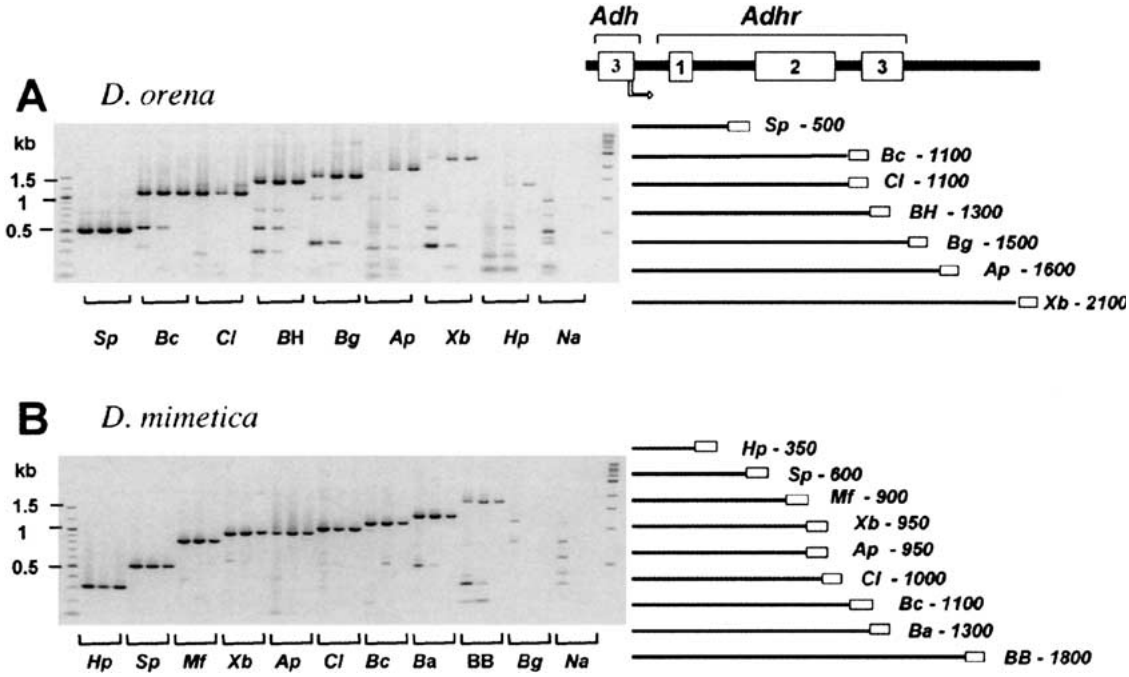

Fig. 3. Vectorette PCR of the *Adhr* gene region in *D. orena* and *D. mimetica*. Gel images show vectorette PCR fragments from amplifications using a specific primer that recognizes the end of the last exon of *Adh*. Amplifications were performed at three different annealing temperatures (shown from left to right for annealing temperatures from low to high for each vectorette library). Abbreviations for restriction enzymes: *Apo*I (Ap), *Bam*HI (Ba), *Bcl*I (Bc), *Bgl*II (Bg), *Bsa*H1 (BH), *Bst*B1 (BB), *Cla*I (Cl), *Eco*RI (ER), *Hpa*II (Hp), *Mfe*I (Mf), *Nar*I (Na), *Nhe*I (Nh), *Spe*I (Sp), *Taq*^{a}I (Ta), and *Xba*I (Xb). Size standards are shown in the right- and leftmost lanes of each gel. Vectorette PCR products are shown in increasing order of size and the corresponding regions of the *Adhr* region are depicted to the right of each gel image. The *open box* and the number at the right end of each PCR product represent the vectorette sequence and the size of PCR product, respectively. Results are also shown for two libraries that were not expected to yield fragments (Hp and Na in *D. orena* and Bg and Na in *D. mimetica*).

and Gouy's method is based on the Tamura model (1992), which has two parameters: a Ts/Tv ratio (κ) and an equilibrium GC content (θ). By assigning a different value of θ to each branch, changes in base composition can be taken into account. The distance matrix (assuming Galtier and Gouy's model and a gamma distribution of rates among sites) and NJ trees were generated using the GGG95 and SK programs. The shape parameter, α, of the gamma distribution was estimated using the EVAL_NH program from the NHML package (Galtier and Gouy 1995, 1998; Galtier et al. 1999; Galtier 2001). GGG95 and SK programs, authored by Dr. Nicolas Tourasse, were kindly provided by Dr. Francisco Rodríguez-Trelles. We used the SEQBOOT program in the PHYLIP package to produce 1000 bootstrap sequence data sets and consensus trees were obtained using the CONSENSE program in the PHYLIP package (Felsenstein 1993).

The SOWH test (Goldman et al. 2000), a likelihood-based parametric bootstrap test, was performed to determine if the maximum likelihood tree (topology II) obtained from our study is better supported than alternative topologies (i.e., topologies I and III and the second-best ML tree). Three independent tests were performed for evaluating alternative trees. For each test, the null hypothesis (H_0) = "the alternative tree is the true topology," while H_A = "the ML tree (tree topology II) obtained from our data is the true topology." One thousand simulated data sets were generated by parametric bootstrapping based on the H_0 tree topology and the ML estimates of the parameters for the H_0 tree (i.e., parameters for nucleotide frequencies, GTR + G + I model, and branch lengths). SEQ-GEN (Rambaut and Grassly 1997) was employed for this process. For each simulated dataset, likelihood scores of the H_0 tree (L_0) and the ML tree (L_{ml}) were estimated to calculate the test statistic, $\delta = L_{ml} - L_0$, using the partial optimization version of the SOWH test ("posPpud" of Goldman et al. [2000]). The δ values for 1000 simulated data sets were employed as a null distribution to determine the probability of the δ value calculated from the original data.

Supplementary material for the gene specific phylogenetic trees, vectorette sequences, and primers and a list of primers used in this study are available at our laboratory web site (http://www.bio.psu.edu/people/faculty/akashi). The sequence alignment files for phylogenetic analyses will be provided upon request.

Results

Vectorette PCR

Figure 3 shows products of vectorette PCR from the *Adhr* locus in *D. orena* and *D. mimetica*. Subsequent DNA sequencing confirmed that vectorette PCR successfully amplified all expected products within 1.5 kb of the specific primer. Desired and spurious PCR products can be distinguished because the ratio of target to spurious fragments increases with annealing temperature. In Fig. 3, the *Bcl*I, *Bsa*HI, *Bgl*II, *Apo*I, and *Xba*I lanes in *D. orena* and the *Bam*HI and *Bst*BI lanes in *D. mimetica* show multiple bands at lower annealing temperatures. At higher annealing temperatures, spurious PCR products decrease in intensity, while target PCR products remain visible.

Table 2. Bootstrap supports for tree topologies of the *melanogaster* species subgroup[a]

	bp	Topology I (((*m*, *s*), (*t*, *y*)), (*e*, *o*))	Topology II (((*e*, *o*), (*t*, *y*)), (*m*, *s*))	Topology III (((*m*, *s*), (*e*, *o*)), (*t*, *y*))
Adh	761/771			62:73:62
Adhr	819/840		86:89:95	
Gld	1541/1548		92:88:76	
ry	3939/4005		100:100:100	
Adh + *Adhr*	1580/1611		85:63:93	
Adh + *Gld*	2302/2319		87:66:78	
Adh + *ry*	4700/4776		100:100:100	
Adhr + *Gld*	2360/2388		98:97:99	
Adhr + *ry*	4758/4845		100:100:100	
Gld + *ry*	5480/5553		100:100:100	
Adh + *Adhr* + *Gld*	3121/3159		98: 89:98	
Adh + *Adhr* + *ry*	5519/5616		100:100:100	
Adh + *Gld* + *ry*	6241/6324		100:100:99	
Adhr + *Gld* + *ry*	6299/6369		100:100:100	
Adh + *Adhr* + *Gld* + *ry*	7060/7164		100:100:100	

[a] bp: the numbers of nucleotides used in phylogenetic analyses/numbers of aligned nucleotide positions (aligned regions containing gaps were not included in the analyses). Percentage bootstrap support (1000 replicates) for each method (maximum parsimony:neighbor joining:maximum likelihood) is shown. Abbreviations for species: *D. melanogaster* (*m*), *D. simulans* (*s*), *D. teissieri* (*t*), *D. yakuba* (*y*), *D. erecta* (*e*), and *D. orena* (*o*).

Vectorette PCR with multiple libraries has several advantages over conventional genomic PCR for amplification of homologous regions from different species. A single specific primer can yield overlapping PCR products ranging in size from 200 bp to 2 kb. Since one end of each PCR fragment is the vectorette region (Fig. 3), an internal vectorette primer designed to anneal to this region can be used as a common sequencing primer to obtain up to 2 kb of sequences from one strand. Other sequencing primers can then be designed to obtain the sequence of the complementary strand. 5′ and 3′ ends of exons, and intron regions that are too large to span by PCR between adjacent exons, can be readily sequenced using this approach. This method can greatly reduce the numbers of primers required to sequence homologous regions of DNA and may be especially useful in molecular evolutionary and systematics studies that require sequences from multiple genes from the same group of species. Designing PCR primers that recognize unsequenced regions can be a limiting factor in such studies. However, recognition of a unique sequence in the genome by the single "specific" primer is critical for a successful vectorette PCR. This method may be less applicable in organisms whose genomes contain large numbers of close paralogs.

Phylogenetic Reconstruction

D. melanogaster Subgroup

For *Adhr* (840 bp) and *Gld* (1548 bp), tree topology II, which groups the *D. teissieri*–*D. yakuba* and *D. erecta*–*D. orena* species pairs together as a species complex, was supported by all phylogenetic algorithms (Table 2). Tree topology II is supported by *ry* (4005 bp) with high bootstrap scores (100% for all three methods) (Table 2). *Adh* sequence data supported topology III with relatively low bootstrap scores (62, 73, and 62% for MP, NJ, and ML, respectively) (Table 2).

For a concatenated sequence [including *Adh*, *Adhr*, *Gld*, and *ry* (7164 bp)], all three methods support topology II with 100% bootstrap support for each branch within the *melanogaster* species subgroup (Fig. 4A). Exclusion of *ry*, which contributes over half of the concatenated sequence, has little effect on the results; tree topology II remains strongly supported (98, 89, and 98% for MP, NJ, and ML, respectively) (Table 2). To detect alternative tree topologies that might be supported by subsets of the concatenated genes, we performed phylogenetic analyses using all combinations of the four genes. For all combined data sets, tree topology II was supported using all three methods (Table 2).

Phylogenetic analyses of the concatenated sequence was also conducted separately for each of the three codon positions (Table 3). Tree topology II is supported by the first, first-plus-second, and third codon positions. Interestingly, topology I is supported by the second codon position with 71, 86, and 78% bootstrap support using MP, NJ, and ML, respectively. However, the number of parsimony-informative sites at the second codon position is small (107 of 2358 sites). Maximum parsimony and neighbor-joining applied to the concatenated protein sequences support topology II (90 and 91% bootstrap support for MP and NJ, respectively).

D. eugracilis, *D. mimetica* (suzukii Subgroup), and *D. lutescens* (takahashii Subgroup)

Phylogenetic inferences among *D. eugracilis*, *D. mimetica*, *D. lutescens*, and the *melanogaster* sub-

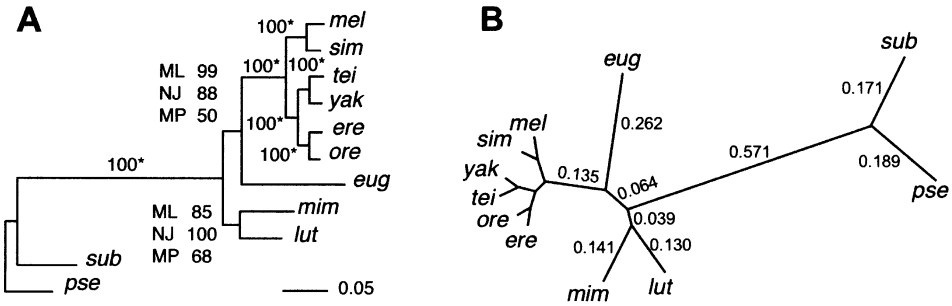

Fig. 4. Maximum likelihood trees of species in the *melanogaster* group inferred from a concatenated sequence of *Adh* + *Adhr* + *Gld* + *ry* genes (7164 bp). Abbreviations for species: *D. melanogaster* (*mel*), *D. simulans* (*sim*), *D. teissieri* (*tei*), *D. yakuba* (*yak*), *D. erecta* (*ere*), *D. orena* (*ore*), *D. eugracilis* (*eug*), *D. mimetica* (*mim*), *D. lutescens* (*lut*), *D. pseudoobscura* (*pse*), and *D. subobscura* (*sub*). **A** Bootstrap consensus tree. Bootstrap values (1000 replicates) on each node are shown and represent, from top to bottom, maximum likelihood (ML), neighbor-joining distance (NJ), and maximum parsimony (MP) values. *indicates identical bootstrap values for ML, NJ, and MP. **B** An unrooted maximum likelihood distance tree based on the 3rd codon positions. Branch lengths (numbers of substitution per site) are shown except for the *melanogaster* subgroup species: *mel* (0.052), *sim* (0.041), *tei* (0.047), *yak* (0.035), *ere* (0.041), *ore* (0.035), *mel-sim* (0.051), *tei-yak* (0.04), *ere-ore* (0.038), and *tei-yak-ere-ore* (0.031). Lineages are named according to the most recent node (*i.e.*, *mel-sim* refers to the lineage from the common ancestor of the subgroup to the common ancestor of *D. melanogaster* and *simulans*).

Table 3. Phylogenetic analyses using different codon positions[a]

	Topology I (((*m*, *s*), (*t*, *y*)), (*e*, *o*))	Topology II (((*e*, *o*), (*t*, *y*)), (*m*, *s*))	Topology III (((*m*, *s*), (*e*, *o*)), (*t*, *y*))	PIS/Tot
1st Codon		69:73:76		257/2354
2nd codon	71:86:78			107/2358
1st + 2nd codon		64:54:62		364/4712
3rd codon		100:100:100		1203/2348
Protein sequence		90:91:—		225/2336

[a] Results are from the concatenated sequence of *Adh* + *Adhr* + *Gld* + *ry* (7164 bp). Percentage bootstrap support (1000 replicates) for each method (maximum parsimony:neighbor joining:maximum likelihood) is shown. Abbreviations for species: *D. melanogaster* (*m*), *D. simulans* (*s*), *D. teissieri* (*t*), *D. yakuba* (*y*), *D. erecta* (*e*), and *D. orena* (*o*). The maximum likelihood method was not applied to protein sequence data. PIS/Tot is the number of parsimony-informative sites/total sequence length.

group vary with the reconstruction method. Under maximum parsimony, different genes appear to support different tree topologies (Table 4). *Adh* supports ((*melanogaster* subgroup, *D. eugracilis*), (*D. lutescens–D. mimetica*)) while *Adhr* and *Gld* support ((*melanogaster* subgroup, *D. lutescens–D. mimetica*), *D. eugracilis*). *ry* supports (((*melanogaster* subgroup, *D. eugracilis*), *D. lutescens*,), *D. mimetica*). When the four genes were concatenated a ((*melanogaster* subgroup, *D. eugracilis*), *D. lutescens–D. mimetica*) topology was obtained with 50% bootstrap support (Table 4 and Fig. 4A). In contrast, neighbor-joining with TN, a relative complex substitution model, gave a common tree topology, ((*melanogaster* subgroup, *D. eugracilis*), *D. lutescens–D. mimetica*), for all genes and combined data sets except *Gld* and *Adhr* + *Gld* (Table 4). Under a more complex maximum likelihood model (GTR + *G* + *I*), the same tree topology was supported by all genes (Table 4). Analyses of the concatenated sequence (*Adh* + *Adhr* + *Gld* + *ry*; 7164 bp) support *D. eugracilis* as the closest relative to the *melanogaster* subgroup with 88 and 99% bootstrap support for NJ and ML, respectively (Fig. 4A).

Since maximum parsimony gives inconsistent tree topologies among *D. eugracilis*, *D. lutescens*, *D. mimetica*, and the *melanogaster* subgroup for different data sets, parsimony-based Bremer support (BS) and partitioned Bremer support (PBS) were determined on each node of the tree (Fig. 4A) for distance and maximum likelihood methods. Table 5 shows that most nodes are supported by each gene (PBS > 0). However, data from *Adh* are incongruent with the nodes grouping *D. teissieri–D. yakuba* + *D. erecta–D. orena* (PBS = −4) and *D. lutescens* and *D. mimetica* (PBS = −2). The node grouping *D. eugracilis* and the *melanogaster* subgroup is neither supported nor rejected by any of the four genes (PBS = 0 for each gene). The *ry* data show a strong disagreement with grouping *D. mimetica* and *D. lutescens* as sister species (PBS = −19), while the total score is still positive (BS = 6) (Table 5).

Phylogenetic Reconstructions Under Simple and Complex Substitution Models

Most substitution models used for reconstructing gene trees assume stationarity of nucleotide frequen-

Table 4. Bootstrap supports for tree topologies of the *melanogaster* species group[a]

	bp	((*melsub*, *u*), (*l*, *i*))	(((*melsub*, *u*), *l*), *i*)	((*melsub*, (*l*, *i*), *u*)
Adh	761/771	91:88:73		
Adhr	819/840	:59:69		92: :
Gld	1541/1548	:—:56		78: :
ry	3939/4005	:75:86	90: :	
Adh + *Adhr*	1580/1611	:92:94		57: :
Adh + *Gld*	2302/2319	:59:88		54: :
Adh + *ry*	4700/4776	:87:96	94: :	
Adhr + *Gld*	2360/2388	:—:72		95: :
Adhr + *ry*	4758/4845	:82:93	65: :	
Gld + *ry*	5480/5553	:69:86	64: :	
Adh + *Adhr* + *Gld*	3121/3159	:69:93		84: :
Adh + *Adhr* + *ry*	5519/5616	:90:98	76: :	
Adh + *Gld* + *ry*	6241/6324	:80:95	75: :	
Adhr + *Gld* + *ry*	6299/6369	:74:93		59: :
Adh + *Adhr* + *Gld* + *ry*	7060/7164	50:88:99		

[a] Numbers of nucleotides used in phylogenetic analyses and the numbers in sequence alignments are shown (aligned regions containing gaps were not included in the analyses). Percentage bootstrap support (1000 replicates) for each method (maximum parsimony:neighbor joining: maximum likelihood) are shown. Bootstrap values less than 50% are shown as . Abbreviations: the *melanogaster* subgroup (*melsub*), *D. eugracillis* (*u*), *D. mimetic* (*i*), and *D. lutescens* (*l*).

Table 5. Partitioned Bremer support (PBS) for the phylogenetic tree of the concatenated sequence[a]

	Tree node							
	(*m*, *s*)	(*t*, *y*)	(*e*, *o*)	(*t-y*, *e-o*)	(*m-s*, (*t-y*, *e-o*))	(*melsub*, *u*)	(*i*, *l*)	(*melsub-u*, *i-l*)
Adh	1	11	12	−4	11	0	−2	37
Adhr	12	10	5	10	25	0	13	70
Gld	28	20	23	11	38	0	14	131
ry	28	25	32	19	54	0	−19	243
BS	69	61	72	36	128	0	6	481

[a] Partitioned Bremer support is shown at each node for each of the genes analyzed. Bremer support (BS) is the sum of the partitioned Bremer support of each gene at the corresponding node. Abbreviations for species: *D. melanogaster* (*m*), *D. simulans* (*s*), *D. teissieri* (*t*), *D. yakuba* (*y*), *D. erecta* (*e*), *D. orena*(*o*), *D. melanogaster* subgroup (*melsub*), *D. eugracilis* (*u*), *D. mimetic* (*i*), and *D. lutescens* (*l*). Tree node (*m*, *s*) represents the node grouping *m* and *s* and (*t-y*, *e-o*) represents the node grouping (*t-y*) and (*e-o*).

cies. Departures from stationarity can mislead tree reconstruction; sequences with similar base composition will tend to be grouped together regardless of their true phylogenetic relationships (Saccone et al. 1989; Weisburg et al. 1989; Loomis and Smith 1990; Hasegawa et al. 1993; Lockhart et al. 1992, 1994; Galtier and Gouy 1995, 1998; Tarrío et al. 2001). In our concatenated sequence, the GC content of *D. eugracilis* is ~8.5% lower than the average of all sequences and ~17% lower at third codon positions. To test base composition stationarity, we employed Rzhetsky and Nei's *I* test (1995), which takes phylogenetic correlations among sequences into account. The stationarity of base composition was rejected for each gene and for the concatenated sequence both for third codon positions ($p < 0.0001$) and for all sites ($p < 0.0001$). In addition, stationarity was rejected at first plus second codon sites for the concatenated sequence and individually for the *ry* gene ($p < 0.0001$).

We tested the impact of nonstationary base composition on our phylogenetic analyses by comparing a ML tree under the Jukes and Cantor (1969) model (JC), which assumes equal base frequencies, and a NJ tree under Galtier and Gouy's model (T92 + *G* + GC%), a nonreversible substitution model that allows GC content to vary among lineages.

Figure 5 shows that the same tree topologies are supported under JC, GTR + *G* + *I*, and T92 + *G* + GC% models. Within the *melanogaster* species subgroup, the *D. teissieri*–*D. yakuba* + *D. erecta*–*D. orena* cluster is strongly supported with very high bootstrap values (100, 100, and 98% for JC, GTR + *G* + *I* and T92 + *G* + GC%, respectively). *D. eugracilis* was grouped closest to the *melanogaster* species subgroup with strong bootstrap support (96, 99, and 99% for JC, GTR + *G* + *I*, and T92 + *G* + GC%, respectively). Bootstrap values supporting the *D. mimetic*–*D. lutescens* species pair increase with the complexity of the model (68, 85, and 91% for JC, GTR + *G* + *I*, and T92 + *G* + GC%, respectively). The tree topology ((*melanogaster* subgroup, *D. eugracilis*), (*D. lutescens*, *D. mimetica*)) remains identical under all three models and is most strongly supported by Galtier and Gouy's method.

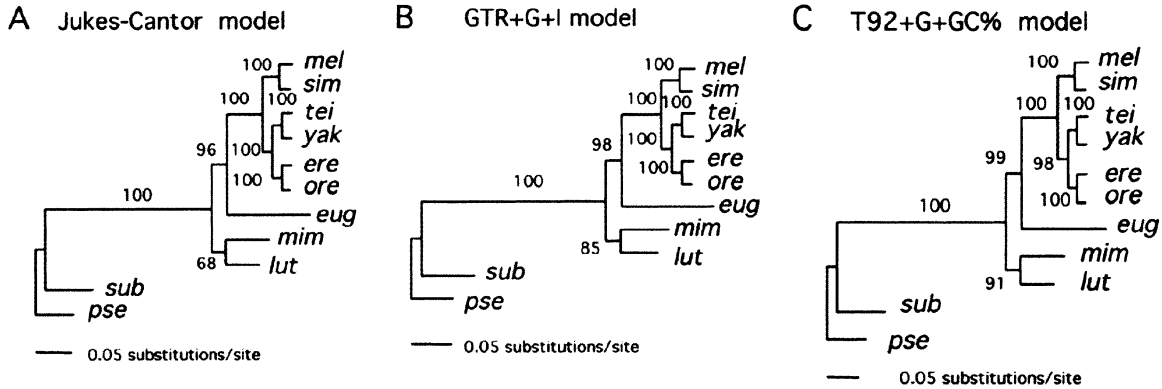

Fig. 5. Phylogenetic trees for concatenated sequences of *Adh* + *Adhr* + *Gld* + *ry* genes (7164 bp) supported by different substitution models. **A** Maximum likelihood tree under the Jukes and Cantor model. **B** Maximum likelihood tree under a general-time reversible model with rate variation among sites and a proportion of invariant sites (GTR + *G* + *I*). **C** Neighbor-joining tree under Galtier and Gouy's model using Tamura's (1992) distance with rate variation among sites and GC content variation among lineages (T92 + *G* + GC%). Abbreviations for species: *D. melanogaster* (*mel*), *D. simulans* (*sim*), *D. teissieri* (*tei*), *D. yakuba* (*yak*), *D. erecta* (*ere*), *D. orena* (*ore*), *D. eugracilis* (*eug*), *D. mimetica* (*mim*), *D. lutescens* (*lut*), *D. pseudoobscura* (*pse*), and *D. subobscura* (*sub*). Bootstrap values (1000 replicates) are shown on each node.

Independent parametric bootstrap tests (SOWH test [Goldman et al. 2000]) indicate that our maximum likelihood tree topology (Fig. 4A) is better supported than the three alternative hypotheses (i.e., topologies I and III and the second-best ML tree) ($p < 0.001$ for each test).

Phylogenetic inference from gene trees assumes orthology of sequences. Comparisons among paralogous sequences that arose from duplications in the lineage ancestral to the *melanogaster* subgroup could mislead our phylogenetic inference. BLAST searches (Altschul et al. 1990) against the *D. melanogaster* genome resulted in low identities for the closest matching protein sequences for the four loci studied (≤38% for *Adh* and *Adhr*, ≤44% for *Gld*, and ≤31% for *ry*). If duplicated genes have been lost, or are highly diverged, in the *melanogaster* species subgroup, different genes should give conflicting tree topologies. Consistent support for topology II from genes on different chromosomes (chromosome 2L for *Adhr*; chromosome 3R for *Gld* and *ry*) suggest that the result is not due to paralogy.

Discussion

Resolution of phylogenetic relationships within the *D. melanogaster* subgroup will be important for comparative genomic (Rifkin et al. 2003) as well as molecular evolutionary studies among these closely related species. Tree topology I, which groups *D. teissieri–D. yakuba* close to *D. melanogaster* species complex, is supported primarily by molecular studies employing allozyme 2d gel electrophoresis (Eisses et al. 1979; Ohnishi et al. 1983; Daïnou et al. 1986; Cariou 1987), DNA–DNA hybridization of scnDNA and mtDNA (Solignac et al. 1986), and restriction map studies of mtDNA (Caccone et al. 1988). DNA sequence evidence favoring topology I comes mainly from studies of the *Adh* gene (Jeffs et al. 1994; Russo et al. 1995). Our analyses of *Adh* supports tree topology III in agreement with Moriyama and Gojobori's (1992) study, which considered only synonymous sites. In the Jeffs et al. (1994) and Russo et al. (1995) analyses, *D. tsacasi* (*montium* subgroup) was employed as the outgroup for tree reconstruction while *D. eugracilis*, *D. mimetica*, and *D. lutescens* were used in our analyses. We included the *D. tsacasi Adh* sequence to reanalyze *Adh* and found that each of the three tree topologies can be obtained using different combinations of outgroups and algorithms, albeit with relatively low bootstrap support (< 76%) (data not shown). This may reflect the short internal branch differentiating the three tree topologies; the branch length supporting topology III is only 0.005 substitution per site (3.85 nucleotide substitutions) for *Adh* data.

Schlötterer et al. (1994) obtained tree topology III with 100% bootstrap support using an internal transcribed spacer (ITS) region of a ribosomal RNA gene (average sequence length of 620 bp). However, this result is sensitive to gene alignments. Schlötterer et al. employed the PILEUP program in the GCG package (UWGCG package, version 7.0, University of Wisconsin Genetics Computer Group) with manual adjustments to minimize gaps. The same data aligned with CLUSTAL using default parameters and without subsequent modification support topology II (82, 94, and 96% bootstrap supports for MP, NJ, and ML, respectively); 53% of sites in the CLUSTAL alignments and 42% in the Schlötterer et al. (1994) alignment contain a gap in at least one of the eight species.

Recent studies have yielded ambiguous results for the branching orders of the *D. melanogaster–D. sim-*

ulans, *D. teissieri–D. yakuba*, and *D. erecta–D. orena* lineages (Shibata and Yamazaki 1995; Okuyama et al. 1996; Inomata et al. 1997; Munté et al. 2001). Different tree topologies were obtained using different genes in the *Amy* multigene family (Shibata and Yamazaki 1995; Okuyama et al. 1996; Inomata et al. 1997). High bootstrap support for different trees were obtained using the nucleotide and amino acid sequences of the *yellow* gene (Munté et al. 2001). In our study, trees reconstructed using *Adhr*, *Gld*, and *ry* sequences individually, and those reconstructed using different combinations of genes, all support a *D. teissieri–D. yakuba* + *D. erecta–D. orena* species clade within the *melanogaster* subgroup. In all combined data of >4000 bp, tree topology II is favored with 99–100% bootstrap support (Table 2). This topology was also supported by analyses of mitochondrial DNA (Nigro et al. 1991), and the *fruitless* gene (Gailey et al. 2000), although bootstrap supports were not given in these studies. Our results are also consistent with phylogenetic inferences using the *H3* gene (Matsuo 2000), the *Cu*, *Zn SOD* gene (Arhontaki et al. 2002), and the *janB* protein sequence (Parsch et al. 2001). Kopp and True (2002) obtained topology II with high bootstrap values using combined data from mitochondrial DNA and nuclear genes. However, support for this topology was strongly dependent on one of the six genes analyzed. Tree topology II is also consistent with evidence from polytene chromosome banding patterns (Lemeunier and Ashburner 1976, 1984), male genital structures (Tsaca and Bocquet 1976), and acoustic characteristics of courtship songs (Cowling and Burnet 1981).

Lachaise et al. (1988) proposed a historical biogeographic scenario for the *melanogaster* subgroup species according to topology I. However, the biogeographic evidence does not refute tree topology II. According to our evidence, an initial separation occurred between the *D. erecta–D. orena–D. teissieri–D. yakuba* lineage and the ancestral lineage of the *melanogaster* species complex. This may have been related to fragmentation of the Congolese forest. The northwest population of the Congolese forest led to the *D. erecta–D. orena–D. teissieri–D. yakuba* lineage, while the northeast population led to the *melanogaster* species complex.

Although this study has focused on relationships within the *melanogaster* subgroup, our analyses may also shed light on phylogenetic relationship among sister subgroups. The *takahashii* subgroup is thought to be closely related to the *suzukii* subgroup (Pélandakis et al. 1991; Pélandakis and Solignac 1993; Goto and Kimura 2001; Kopp and True 2002; Schawaroch 2002), although the latter may be polyphyletic (Schawaroch 2002; Kopp and True 2002).

The phylogenetic position of *D. eugracilis* has been difficult to resolve. Pélandakis et al. (1991) and Pélandakis and Solignac (1993) have inferred a close relationship between *eugracilis* and the *takahashii* + *suzukii* clade. Other studies support different topologies: Schawaroch (2002) grouped *D. eugracilis* closest to the *melanogaster* subgroup, while Kopp and True (2002) suggested a basal position of *D. eugracilis* in the *melanogaster* + oriental lineages.

In our analyses, maximum parsimony gave inconsistent results for phylogenetic relationships among *D. eugracilis*, *D. mimetica*, *D. lutescens*, and the *melanogaster* subgroup for different data sets (Table 4). PBS analyses indicate that individually, these four genes provide little information about the phylogenetic position of *D. eugracilis* (PBS = 0 for each gene) and give conflicting results for grouping *D. lutescens* and *D. mimetica* as sister species (Table 5). This inconsistency may reflect a combination of long branch lengths (Fig. 4B) and a difference in GC content between *D. eugracilis* and other *melanogaster* group species (Table 1). Galtier and Gouy (1995) have shown that the efficiency of maximum parsimony is sensitive to GC content differences among lineages. On the other hand, with more complex substitution models, neighbor-joining and maximum likelihood converge on the same topology ((*melanogaster* subgroup, *D. eugracilis*), *D. lutescens–D. mimetica*) for each gene and for different combined data sets. Galtier and Gouy's method applied to the concatenated sequences (7164 bp) gives strong bootstrap support for this topology (Fig. 5C). Goldman et al.'s (2000) likelihood-based tree topology test also shows that this tree topology is better supported than alternative topologies. Sequence data from closely related outgroups may help to establish relationships among these subgroups more rigorously.

Acknowledgments. We are grateful to C. dePamphilis for allowing us use of his automated sequencer and to S. Hauswaldt for technical assistance. We also thank R. DeSalle and K. Montooth for comments on the manuscript and F. Rodríguez-Trelles for advice on use of the GGG95 and SK programs. M. Ashburner kindly provided the *D. erecta* strain. This research was supported by grants from the Alfred P. Sloan Foundation and the National Science Foundation (DEB-0235472) to HA.

References

Akashi H (1995) Inferring weak selection from patterns of polymorphism and divergence at "silent" sites in *Drosophila* DNA. Genetics 139:1067–1076

Akashi H (1996) Molecular evolution between *Drosophila melanogaster* and *D. simulans*: Reduced codon bias, faster rates of amino acid substitution, and larger proteins in *D. melanogaster*. Genetics 144:1297–1307

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Arhontaki K, Eliopoulos E, Goulielmos G, Kastanis P, Tsakas S, Loukas M, Ayala F (2002) Functional constraints of the Cu, Zn

superoxide dismutase in species of the *Drosophila melanogaster* subgroup and phylogenetic analysis. J Mol Evol 55:745–756

Arnold C, Hodgson IJ (1991) Vectorette PCR: A novel approach to genomic walking. PCR Methods Appl 1:39–42

Ashburner M (1989) *Drosophila:* A laboratory handbook. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

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

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

Barnes SR, Webb DA, Dover G (1978) The distribution of satellite and main-band DNA components in the *melanogaster* species subgroup of *Drosophila* I. Fractionation of DNA in actinomycin D and distamycin A density gradients. Chromosoma 67:341–363

Bettencourt BR, Feder ME (2001) *Hsp70* duplication in the *Drosophila melanogaster* species group: How and when did two become five? Mol Biol Evol 18:1272–1282

Bonneton F, Wegnez M (1995) Developmental variability of metallothionein *Mtn* gene expression in the species of the *Drosophila melanogaster* subgroup. Dev Genet 16:253–263

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

Bremer K (1994) Branch support and tree stability. Cladistics 10:295–304

Caccone A, Moriyama EN, Gleason JM, Nigro L, Powell JR (1996) A molecular phylogeny for the *Drosophila melanogaster* subgroup and the problem of polymorphism data. Mol Biol Evol 13:1224–1232

Caccone A, Amato GD, Powell JR (1988) Rates and patterns of scnDNA and mtDNA divergence within the *Drosophila melanogaster* subgroup. Genetics 118:671–683

Cariou ML (1987) Biochemical phylogeny of the eight species in the *Drosophila melanogaster* subgroup, including *D. sechellia* and *D. orena*. Genet Res 50:181–185

Cowling DE, Burnet B (1981) Courtship songs and genetic control of their acoustic characteristics in sibling species of *Drosophila melanogaster* subgroup. Anim Behav 29:924–935

Daïnou O, Cariou ML, David JR, Hickey D (1987) Amylase gene duplication: An ancestral trait in the *Drosophila melanogaster* species subgroup. Heredity 59(Pt 2):245–251

Eisses KT, Van Dijk H, Van Delden W (1979) Genetic differentiation within *melanogaster* species group of the genus *Drosophila* (*Sophophora*). Evolution 33:1063–1068

Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol 17:368–376

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

Felsenstein J (1993) PHYLIP, phylogenetic inference package and documentation. Distributed by the author, Department of Genetics, University of Washington, Seattle

Fitch WM (1971) Towards defining the course of evolution: Minimum change for a specific tree topology. Syst Zool 20:406–416

Gailey DA, Ho SK, Ohshima S, Liu JH, Eyassu M, Washington MA, Yamamoto D, Davis T (2000) A phylogeny of the Drosophilidae using the sex-behaviour gene fruitless. Hereditas 133:81–83

Galtier N (2001) Maximum-likelihood phylogenetic analysis under a covarion-like model. Mol Biol Evol 18:866–873

Galtier N, Gouy M (1995) Inferring phylogenies from DNA sequences of unequal base compositions. Proc Natl Acad Sci USA 92:11317–11321

Galtier N, Gouy M (1998) Inferring pattern and process: Maximum-likelihood implementation of a nonhomogeneous model of DNA sequence evolution for phylogenetic analysis. Mol Biol Evol 15:871–879

Galtier N, Tourasse N, Gouy M (1999) A nonhyperthermophilic common ancestor to extant life forms. Science 283:220–221

Gentile KL, Burke WD, Eickbush TH (2001) Multiple lineages of R1 retrotransposable elements can coexist in the rDNA loci of *Drosophila*. Mol Biol Evol 18:235–245

Goldman N, Anderson JP, Rodrigo AG (2000) Likelihood-based tests of topologies in phylogenetics. Syst Biol 49:652–670

Goto SG, Kimura MT (2001) Phylogenetic utility of mitochondrial *COI* and nuclear *Gpdh* genes in *Drosophila*. Mol Phylogenet Evol 18:404–422

Harr B, Zangerl B, Schlotterer C (2000) Removal of microsatellite interruptions by DNA replication slippage: phylogenetic evidence from *Drosophila*. Mol Biol Evol 17:1001–1009

Hasegawa M, Hashimoto T, Adachi J, Iwabe N, Miyata T (1993) Early branchings in the evolution of eukaryotes: Ancient divergence of entamoeba that lacks mitochondria revealed by protein sequence data. J Mol Evol 36:380–388

Hecker KH, Roux KH (1996) High and low annealing temperatures increase both specificity and yield in touchdown and stepdown PCR. Biotechniques 20:478–485

Higgins DG, Sharp PM (1988) CLUSTAL: A package for performing multiple sequence alignment on a microcomputer. Gene 73:237–244

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

Jeffs PS, Holmes EC, Ashburner M (1994) The molecular evolution of the alcohol dehydrogenase and alcohol dehydrogenase-related genes in the *Drosophila melanogaster* species subgroup. Mol Biol Evol 11:287–304

Jordan IK, McDonald JF (1998) Evolution of the copia retrotransposon in the *Drosophila melanogaster* species subgroup. Mol Biol Evol 15:1160–1171

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) Mammalian protein metabolism. Academic Press, New York, pp 21–132

Kopp A, True JR (2002) Phylogeny of the Oriental *Drosophila melanogaster* species group: A multilocus reconstruction. Syst Biol 51:786–805

Lachaise D, Cariou ML, David JR, Lemeunier F, Tsacas L, Ashburner M (1988) Historical Biogeography of the *Drosophila melanogaster* species subgroup. Evol Biol 22:159–225

Lachaise D, Harry M, Solignac M, Lemeunier F, Benassi V, Cariou ML (2000) Evolutionary novelties in islands: *Drosophila santomea*, a new *melanogaster* sister species from Sao Tome. Proc R Soc Lond B Biol Sci 267:1487–1495

Lemeunier F, Ashburner M (1976) Relationships within the *melanogaster* species subgroup of the genus *Drosophila* (*Sophophora*). II. Phylogenetic relationships between six species based upon polytene chromosome banding sequences. Proc R Soc Lond B Biol Sci 193:275–294

Lemeunier F, Ashburner M (1984) Relationships within the *melanogaster* species subgroup of the genus *Drosophila* (*Sophophora*) IV. The chromosomes of two new species. Chromosoma (Berl) 89:343–351

Lemeunier F, David JR, Tsacas L, Ashburner M (1986) The *melanogaster* species group. In: Ashburner M, Carson HL, Thompson JN (eds) The genetics and biology of *Drosophila*. London, pp 147–256

Lockhart PJ, Howe CJ, Bryant DA, Beanland TJ, Larkum AW (1992) Substitutional bias confounds inference of cyanelle origins from sequence data. J Mol Evol 34: 153–162

Lockhart PJ, Stell MA, Hendy MD, Penny D (1994) Recovering evolutionary tress under a more realistic model of sequence evolution. Mol Biol Evol 11:605–612

Loomis WF, Smith DW (1990) Molecular phylogeny of *Dictyostelium discoideum* by protein sequence comparison. Proc Natl Acad Sci USA 87:9093–9097

Matsuo Y (2000) Molecular evolution of the histone 3 multigene family in the *Drosophila melanogaster* species subgroup. Mol Phylogenet Evol 16:339–343

Moriyama EN, Gojobori T (1992) Rates of synonymous substitution and base composition of nuclear genes in *Drosophila*. Genetics 130:855–864

Munté A, Aguadé M, Segarra C (2001) Changes in the recombinational environment affect divergence in the yellow gene of *Drosophila*. Mol Biol Evol 18:1045–1056

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

Ohnishi S, Kawanishi M, Watanabe TK (1983) Biochemical phylogenies of *Drosophila*: Protein differences detected by two-dimensional electrophoresis. Genetica 6:55–63

Okuyama E, Shibata H, Tachida H, Yamazaki T (1996) Molecular evolution of the 5′-flanking regions of the duplicated Amy genes in *Drosophila melanogaster* species subgroup. Mol Biol Evol 13:574–583

Parsch J, Meiklejohn CD, Hauschteck-Jungen E, Hunziker P, Hartl DL (2001) Molecular evolution of the *ocnus* and *janus* genes in the *Drosophila melanogaster* species subgroup. Mol Biol Evol 18:801–811

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

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

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 28:87–138

Rambaut A, Grassly NC (1997) Seq-Gen: An application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. Comput Appl Biosci 13:235–238

Rifkin SA, Kim J, White KP (2003) Evolution of gene expression in the *Drosophila melanogaster* subgroup. Nat Genet 33:138–144

Riley J, Butler R, Ogilvie D, Finniear R, Jenner D, Powell S, Anand R, Smith JC, Markham AF (1990) A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. Nucleic Acids Res 18:2887–2890

Ross JL, Fong PP, Cavener DR (1994) Correlated evolution of the cis-acting regulatory elements and developmental expression of the *Drosophila Gld* gene in seven species from the subgroup *melanogaster*. Dev Genet 15:38–50

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

Rzhetsky A, Nei M (1995) Tests of applicability of several substitution models for DNA sequence data. Mol Biol Evol 12:131–151

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

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

Santamaria P (1975) Transplantation of nuclei between eggs of different species of *Drosophila*. Wilhelm Roux Arch 178:89–98

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

Schlötterer C, Hauser MT, von Haeseler A, Tautz D (1994) Comparative evolutionary analysis of rDNA ITS regions in *Drosophila*. Mol Biol Evol 11:513–522

Shibata H, Yamazaki T (1995) Molecular evolution of the duplicated *Amy* locus in the *Drosophila melanogaster* species subgroup: Concerted evolution only in the coding region and an excess of nonsynonymous substitutions in speciation. Genetics 141:223–236

Solignac M, Monnerot M, Mounolou JC (1986) Mitochondrial DNA evolution in the *melanogaster* species subgroup of *Drosophila*. J Mol Evol 23:31–40

Sorenson MD (1999) TreeRot, version 2. Boston University, Boston, MA

Sullivan W, Ashburner M, Hawley R (2000) *Drosophila* protocols. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Swofford DL (2000) PAUP*. Phylogenetic analysis using parsimony (* and other methods). Version 4. Sinauer Associates, Sunderland, MA

Takano-Shimizu T (1999) Local recombination and mutation effects on molecular evolution in *Drosophila*. Genetics 153:1285–1296

Takano-Shimizu T (2001) Local changes in GC/AT substitution biases and in crossover frequencies on *Drosophila* chromosomes. Mol Biol Evol 18:606–619

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, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10:512–526

Tarrio R, Rodriguez-Trelles F, Ayala FJ (2001) Shared nucleotide composition biases among species and their impact on phylogenetic reconstructions of the Drosophilidae. Mol Biol Evol 18:1464–1473

Terzian C, Ferraz C, Demaille J, Bucheton A (2000) Evolution of the Gypsy endogenous retrovirus in the *Drosophila melanogaster* subgroup. Mol Biol Evol 17:908–914

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

Tsacas L, Bocquet C (1976) L'espèce chez les *Drosophilidae*. Mém Soc Zool Fr 39:203–247

Weisburg WG, Giovannoni SJ, Woese CR (1989) The Deinococcus-Thermus phylum and the effect of rRNA composition on phylogenetic tree construction. Syst Appl Microbiol 11:128–134

Yang Z (1994) Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. J Mol Evol 39:306–314