Molecular Phylogenetics and Evolution
Vol. 16, No. 2, August, pp. 296–307, 2000
doi:10.1006/mpev.2000.0824, available online at http://www.idealibrary.com on IDEAL®

Phylogenetic Analysis of the *repleta* Species Group of the Genus *Drosophila* Using Multiple Sources of Characters

Celeste M. Durando,*,† Richard H. Baker,†,‡,[1] William J. Etges,§ William B. Heed,¶ Marvin Wasserman,* and Rob DeSalle†

*Department of Biology, City University of New York, New York, New York 10036; †Department of Entomology, American Museum of Natural History, 79th Street at Central Park West, New York, New York 10024; ‡Department of Biology, Yale University, New Haven, Connecticut 06511; §Department of Biological Sciences, University of Arkansas, Fayetteville, Arkansas 72701; and ¶Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 05721

Received September 14, 1999, revised April 7, 2000

The species in the *repleta* group of the genus *Drosophila* have been placed into five subgroups—the *mulleri, hydei, mercatorum, repleta,* and *fasciola* subgroups. Each subgroup has been further subdivided into complexes and clusters. Extensive morphological and cytological analyses of the members of this species group have formed the foundation for the proposed relationships among the members of the *repleta* species group. Fifty-four taxa, including 46 taxa belonging to the *repleta* species group, were sequenced for fragments of four genes—16S ribosomal DNA (16S), cytochrome oxidase II (COII), and nitrogen dehydrogenase 1 (ND1) of the mitochondrial genome and a region of the *hunchback* (*hb*) nuclear gene. We also generated a partial data set of elongation factor 1-alpha (Ef1α) sequences for a subset of taxa. Our analysis used both DNA characters and chromosomal inversion data. The phylogenetic hypothesis we obtained supports many of the traditionally accepted clades within the *mulleri* subgroup, but the monophyly of taxonomic groups outside of this subgroup appears not to be supported. Phylogenetic analysis revealed one well-supported, highly resolved clade that consists of closely related members of the *mulleri* and *buzzatii* complexes. The remaining taxa, a wide assortment of taxonomic groups, ranging from members of other species groups to members of several subgroups and members of three species complexes from the *mulleri* subgroup are found in poorly supported arrangements at the base of the tree.

INTRODUCTION

The *Drosophila repleta* species group is among the largest of all species groups in the genus *Drosophila*. Part of the *virilis–repleta* radiation (Throckmorton, 1982a), it is considered one of the most important and successful radiations in the genus *Drosophila*. For the most part the members of the *repleta* species group are found in the arid or semiarid deserts of the New World where they live on various species of cactus (Wasserman, 1992). The species in this group have been placed into five subgroups—the *D. mulleri* subgroup, the *D. hydei* subgroup, the *D. mercatorum* subgroup, the *D. repleta* subgroup, and the *D. fasciola* subgroup—and each subgroup has been further subdivided into complexes, clusters, and subclusters.

Current knowledge of the phylogeny of this species group is based on the morphological work of Throckmorton (1982a) and Vilela (1983) and the cytological work of Wasserman (1982, 1992 for reviews). In addition, several allozyme studies (Zouros, 1973; Richardson *et al.*, 1975; Richardson and Smouse, 1976; Richardson *et al.*, 1977; Heed *et al.*, 1990) and some molecular studies (Sullivan *et al.*, 1990; Russo *et al.*, 1995; Spicer, 1995, 1996) have contributed to our knowledge of relationships among members of subsets of this species group. However, none of the allozyme or molecular studies has attempted to address the phylogenetic relationships of the *repleta* species group as a whole.

Detailed polytene chromosome maps have been constructed for 70 of the 91 species in this group, and more than 296 inversions have been mapped, of which 118 constitute fixed differences between species. Ninety-four of these 118 fixed chromosomal inversions are autapomorphic and diagnostic for various single species. Several of the chromosomal inversions that have been studied so far are variable among closely related species indicating the possible utility of inversions as phylogenetic tools at this level, yet the degree of resolution from the inversion data within species complexes and clusters is low (Wasserman, 1992). In addition, although these data have been used to infer relationships among species, they tell us little about

[1] Present address: The Galton Laboratory, University College London, 4 Stephenson Way, London NW1 2HE, UK.

1055-7903/00 $35.00

the direction of evolution because there is no way to determine which inversions are primitive and which are derived. Therefore, it is necessary to place the inversion data into a cladistic framework along with other sources of characters.

Chromosomal information (Wasserman, 1982, 1992) has been an excellent base for understanding of phylogeny in this species group. However, a more detailed phylogenetic understanding of the group will most likely be obtained from DNA sequence information, which is the focus of this study. In particular, there are several questions that remain open in the phylogenetics of this species group that can be addressed by addition of molecular information. These include (1) the phylogenetic relationships of the various subgroups to one another; (2) the monophyly of the various subgroups, (3) the phylogenetic relationships of species within species complexes, especially the *mulleri* complex; and (4) the phylogenetic placement of the miscellaneous species, such as *D. pegasa* and *D. hamatofila*, into complexes.

For this study we have sequenced four genes: mt 16S rDNA, mt ND1, mt COII (Simon *et al.,* 1994) and the nuclear gene *hunchback* (Treier *et al.,* 1989). We have assessed the relative contribution of each gene partition, as well as the chromosomal inversion data, to the phylogenetic hypothesis we generated for this study. Previous studies, such as those done on Hawaiian *Drosophila* (Carson, 1972; Gillespie, 1996), the *Drosophila melanogaster* species group (Lemeunier *et al.,* 1986), the *Drosophila virilis* species group (Throckmorton, 1982b), and the *Drosophila repleta* species group (Wasserman, 1982, 1992), have indicated that *Drosophila* inversion character data provide valuable information. As more data accrue for various groups of *Drosophila,* this notion of high utility of inversion data can be examined empirically.

MATERIALS AND METHODS

Flies and DNA sequences. Table 1 lists all flies used in this study. We chose eight outgroup taxa to root our phylogenetic trees—*D. melanogaster,* a member of the subgenus *Sophophora; D. cyrtoloma* and *D. longipodis,* members of the Hawaiian *Drosophila* group; *D. virilis; D. pavani* and *D. gaucha,* members of the *D. mesophragmatica* group; *D. canalinea* of the *D. canalinea* species group; and *D. camargoi,* a member of the *D. dreyfusi* group. *D. melanogaster* was chosen as a distant outgroup as it is a member of another subgenus within the genus *Drosophila,* and the Hawaiian *Drosophila* were chosen as a closer outgroup as they are one of the potential sister groups to the *virilis–repleta* radiation (Remsen and DeSalle, 1998; Kwiatowski and Ayala, 1999). We also chose to examine the placement of *D. virilis* in relation to the *repleta* species group flies as well as the relationship of the species in the *D. mesophragmatica* species group and the *D. canalinea* and *D. dreyfusi* species groups because these species have long been considered the closest sister taxa to the *D. repleta* group. DNA was isolated from single and multiple flies using the methods outlined in Vogler *et al.* (1993). The polymerase chain reaction (PCR) was used to amplify fragments for sequencing. Primers used are described in Baker and DeSalle (1997) and Baker *et al.* (1998). PCR products were cleaned using Geneclean kits (BIO 101) and sequenced directly using either manual or automated sequencing methods. Manual sequencing was accomplished using Sequenase (U.S. Biochemicals) and S^{35} labeling. Autoradiograms were used to visualize the sequences, and all sequences were generated in both directions. Automated sequencing was accomplished using the ABI dye terminator system and sequenced on an ABI 373 machine. Inspection and correction of automated sequences were accomplished using SEQUENCHER software (Gene Codes Corp., 1995).

Data matrices and phylogenetic analysis. All sequences were compiled into NEXUS files after alignment of individual gene partitions. We explored the alignment space using the methods outlined in Gatesy *et al.* (1994) in which alignments are obtained for several gap:change costs and various multiple alignments are examined for regions of "ambiguity." We removed ("culled") regions in our alignments that exhibited ambiguity as defined by Gatesy *et al.* (1994). COII alignments were trivial as no indels occurred in this gene region for all of the taxa in our matrix. The mt 16S rDNA sequences required alignment, and approximately 60–70 bases were removed in the culling process. *Hunchback* (*hb*) alignments were accomplished by first translating the DNA sequences into amino acid sequences and performing alignments on the amino acid sequences using ClustalX (Gibson *et al.,* 1994). Areas of alignment ambiguity were removed as described above and the corresponding DNA sequences were substituted for the amino acids. This process was necessary due to the existence of large stretches of polyglutamine repeats in the *hb* sequences that produced several regions of alignment ambiguity. Chromosomal inversion data were also coded as presence or absence of particular inversions and included in the data matrix. Phylogenetic analysis was accomplished using PAUP 4.01b (Swofford, 1999). We analyzed each of the four genes and the inversions separately and in combination in order to explore the interaction of the various gene regions in phylogenetic analysis. In addition, we examined the congruence of the various gene partitions and inversions using the incongruence length difference (ILD; Farris *et al.,* 1994, 1995) and

TABLE 1

Taxonomic List of *Drosophila* Species Used in This Study

Species group	Subgroup	Complex	Cluster	Species	Source
D. repleta	*D. mulleri*	*D. mulleri*	*D. mojavensis*	*D. mojavensis**	Vallecito, CA
				D. arizonae	Tomatlan, Jalisco, Mexico
				D. navojoa	Las Bocas, Sonora, Mexico
			D. mulleri	*D. wheeleri**	Catalina Island, CA
				D. aldrichi	Zapotitlan, Puebla, Mexico
				D. mulleri	Big Pine Key, FL
				D. nigrodumosa	Merida, Venezuela
				D. huaylasi	WBH
				D. mayaguana	Grand Inagua Island
				*D. straubae**	Navassa Island
				D. parisiena	Fond Parisien, Haiti
			D. longicornis	*D. longicornis*	Zapotitlan, Puebla, Mexico
				D. propachuca	MW
				*D. pachuca**	MW
				D. mainlandi	Catalina Island, CA
				D. hexastigma	Zapotitlan, Puebla, Mexico
				D. spenceri	Guayamas, Sonora, Mexico
			D. ritae	*D. desertorum*	MW
				D. ritae	1471.2
		D. buzzatii	*D. martensis*	*D. martensis**	MW
				D. starmeri	La Palmares, Venezuela
				D. uniseta	MW
				D. venezolana	MW
			D. buzzatii	*D. buzzatii*	1291.1
				D. borborema	MW
				*D. serido**	1431.2
				D. koepferae	Vipos, Argentina
			D. stalkeri	*D. stalkeri*	Big Pine Key, FL
				D. richardsoni	Spanish Point, Monserrat
		D. eremophila		*D. eremophila**	Guayalejo, Tamaulipas, Mexico
				D. mettleri	MW
				*D. micromettleri**	Skyline Drive, Cuba
		D. meridiana		*D. meridiana*	1342.0
				*D. meridionalis**	MW
		D. anceps		*D. anceps**	MW
				*D. leonis**	1395.0
				*D. nigrospiracula**	1503.0
		Miscellaneous		*D. hamatofila*	MW
				*D. pegasa**	Oaxaca, Oaxaca, Mexico
	D. hydei	*D. bifurca*		*D. nigrohydei*	MW
		D. hydei		*D. hydei*	Tequila, Jalisco, Mexico
	D. mercatorum			*D. paranaensis**	MW
				*D. mercatorum**	MW
	D. repleta	*D. fulvimacula*		*D. fulvimacula**	MW
		D. repleta		*D. neorepleta**	1611.2
	D. fasciola	Miscellaneous		*D. ellisoni*	MW
D. canalinea				*D. canalinea**	1221.1
D. dreyfusi				*D. camargoi*	1221.2
D. mesophragmatica				*D. gaucha**	1231.0
				D. pavani	1241.0
D. virilis				*D. virilis**	See Baker and DeSalle (1997)
Hawaiian *Drosophila*				*D. cyrtoloma*	See Baker and DeSalle (1997)
				D. longipodis	See Baker and DeSalle (1997)
D. melanogaster				*D. melanogaster**	See Baker and DeSalle (1997)

Note. Taxa marked with an asterisk are those belonging to the subset of taxa used to perform the analysis with Ef1α. Numbers refer to Bowling Green Stock Center numbers. Localities refer to locations where M.W., W.B.H., and W.J.E. collected specimens. M.W. and W.B.H. refer to specimens archived in the labs or Wasserman or Heed.

the associated statistical test for congruence implemented in PAUP 4.01b. Bootstrap values (Felsenstein, 1985) were generated using PAUP 4.01b. Bremer supports (Bremer, 1988, 1994) were calculated using the AUTODECAY program (Eriksson, 1997). Partitioned Bremer supports for the various character partitions

were calculated using the methods outlined in Baker and DeSalle (1997) and Baker *et al.* (1998).

Higher level versus lower level analyses. Initially our phylogenetic analysis was accomplished using a broad sampling of species in the *repleta* species group as well as the outgroup and sister taxa mentioned above. Our initial analysis using the four gene regions listed above resulted in a lack of resolution at the base of the tree where relationships between the species subgroups would be most evident. To examine relationships of the species subgroups we generated sequences for a subset of taxa for Ef1α (Table 1). We sequenced Ef1α for only a subset because at this level we were interested in the overall relationships of the species subgroups, and there is a relatively low level of variability in Ef1α (Cho *et al.*, 1995). We analyzed the Ef1α data in a combined analysis by pruning our data matrix to only those taxa that had complete sequences for all gene partitions. Alignments for Ef1α were trivial as no indels were apparent in these sequences for the taxa we examined. All phylogenetic analyses using this higher level matrix were accomplished as described above.

RESULTS AND DISCUSSION

A Phylogenetic Hypothesis for Relationships of Species in the repleta Species Group

Two parsimony trees were obtained by analysis of the 46 ingroup taxa and 8 outgroup taxa (two Hawaiian *Drosophila* species and *D. melanogaster*). There was a total of 501 characters that were phylogenetically informative in the combined character matrix. The strict consensus of the two most parsimonious trees (CI = 0.31, RI = 0.52, and a total length of 2491 steps) is shown in Fig. 1. This consensus tree contains one major clade showing a high degree of resolution and a second set of taxa at the base of the tree that shows a low degree of resolution. The clade that is highly resolved contains the *mulleri* and *buzzatii* complexes as well as a single "miscellaneous" species, *D. hamatofila*, all from the *mulleri* subgroup. The less resolved group of taxa contains a mixture of all five species subgroups, including three *mulleri* subgroup complexes and species from three of the species groups allied to the *repleta* group (*mesophragmatica, canalinea,* and *dreyfusi* species groups). The *meridiana* complex, a member of the *mulleri* subgroup, is shown as the most basal clade in the consensus tree. While this unresolved group of flies is shown as a monophyletic group in Fig. 1, support for this hypothesis is low. In addition, within this group only relationships at the tips of the tree are well resolved, with relationships among the species subgroups and complexes depicted as a polytomy.

The high degree of resolution in the *mulleri* and *buzzatii* complexes can be contrasted with the lack of significant resolution in the rest of the taxa in the analysis. Of the five subgroups (*hydei, repleta, fasciola, mulleri,* and *mercatorum*) only the *mercatorum* and *hydei* subgroups are seen as monophyletic in the consensus tree in Fig. 1. The species subgroup for which we have the best sampling, the *mulleri* subgroup, is polyphyletic and the representatives of the *repleta* subgroup are paraphyletic with respect to the *mercatorum* subgroup flies. Of the five species complexes in the *mulleri* subgroup, four are clearly monophyletic with moderate to high character support (see Table 2). The fifth complex is embedded in the unresolved clade described above. The fact that the ten species comprising the *buzzatii* complex are recovered as a monophyletic group supports the conclusions of the cytological reanalysis of these species performed by Ruiz and Wasserman (1993). Our consensus tree shows that the *buzzatii* complex is sister to the *mulleri* complex and that the *buzzatii* complex (BS = 18 and BP = 100%) is monophyletic. Within the *mulleri* complex there are four so-called clusters (*mojavensis, ritae, longicornis,* and *mulleri*). The *mulleri* cluster and the *mojavensis* cluster are intermixed in Fig. 1 as four of the *mulleri* cluster species (*aldrichi, wheeleri, mulleri,* and *nigrodumosa*) are observed as sister to the *mojavensis* cluster (*mojavensis, navojoa,* and *arizonae*) with the remaining *mulleri* cluster species (*parisiena, mayaguana,* and *straubae*) as sister to these. Support for these relationships is relatively strong as the node defining the break between the *mulleri* cluster flies (node 31) has BS = 3 and BP = 71%. It is interesting to note that *D. huaylasi,* which had been placed into the *mulleri* cluster based on the fact that it is homosequential with the other species in that cluster, is recovered as a member of the *mojavensis* cluster in Fig. 1. The males in the *mojavensis* cluster possess a characteristic penis, and when we went back to the original description of *D. huaylasi* (Fontevila *et al.*, 1990), we found that it too possessed the *mojavensis* cluster penis. The *longicornis* and *ritae* clusters are also intermixed, but the relationships of the flies in these two clusters is not robust as BS and BP values are relatively low at the base of this clade.

We also estimated the number of steps that need to be added to the parsimony tree in order to make the groups in Fig. 1 monophyletic. These trees indicate that although a group is not monophyletic, it could easily become monophyletic with the addition of more data, as evidenced by the low number of extra steps required to make it monophyletic (*repleta* group, *repleta* subgroup, *anceps* complex). On the other hand, the lack of monophyly of several clusters (*mojavensis, ritae, longicornis,* and *mulleri*) and the *mulleri* subgroup is more strongly supported by our data, as a larger number of extra steps are necessary to make each of them monophyletic (Table 2).

In general, resolution at the base of the tree is ex-

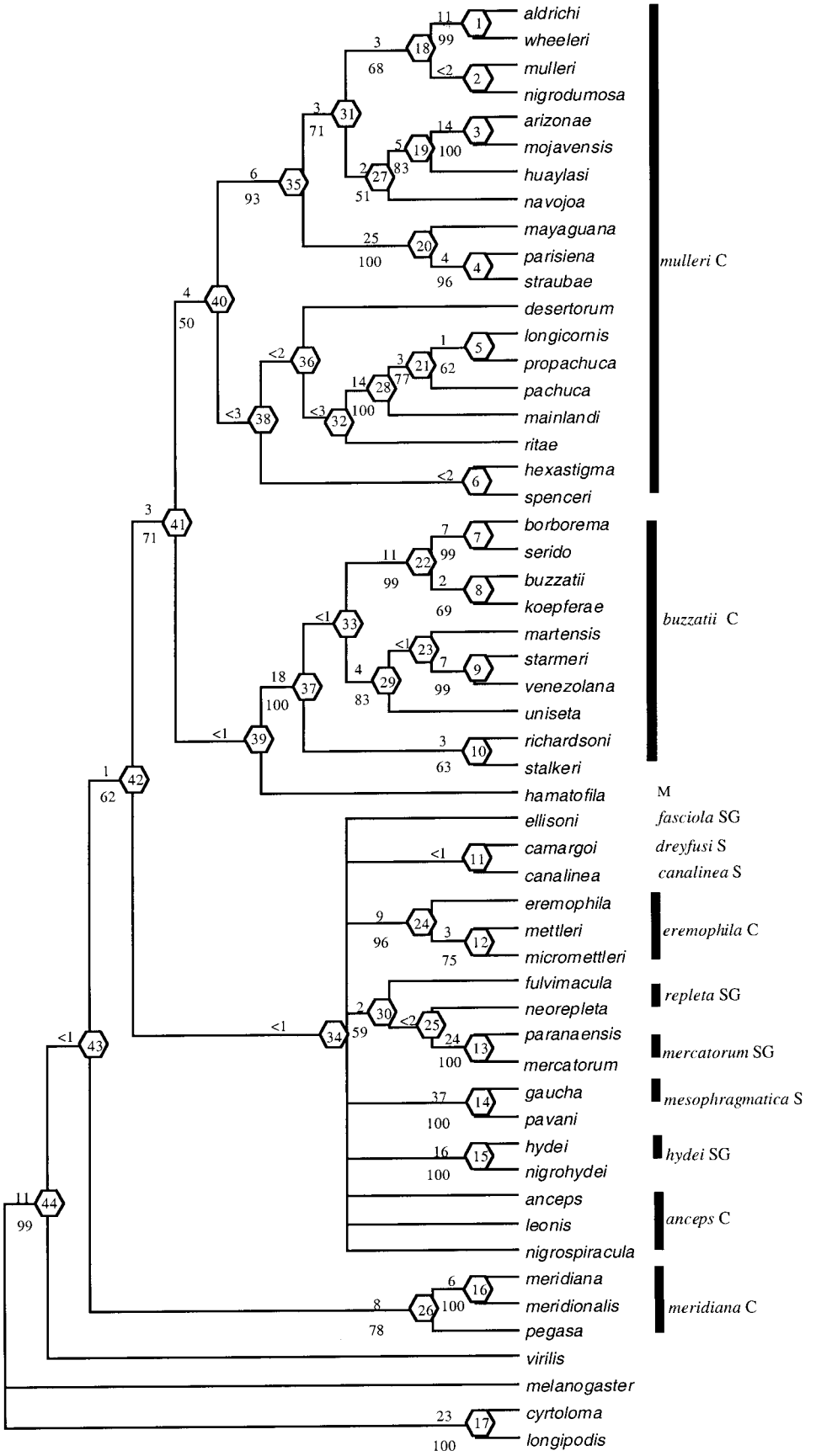

FIG. 1. Strict consensus of two most parsimonious trees from combined analysis of all the data equally weighted. Numbers in hexagons are node designations. Bremer support values are provided above each node and bootstrap values below each node. SG, subgroup; C, complex; S, species group.

TABLE 2

Support for Previously Established Species Groups, Subgroups, Complexes, and Clusters

Previously established group	Fig. 1 (BS, BP)	Fig. 4a (BP)	Fig. 4b (BP)	MONO
Group				
repleta	N	Y (<50)	N	2
mesophragmatica	Y (37, 100)	Y (100)	Y (88)	
Subgroup				
mulleri	N	N	N	7
hydei	Y (16, 100)	Y (100)	Y (100)	
repleta	N	N	N	1
mercatorum	Y (24, 100)	Y (100)	Y (100)	
Complex				
mulleri	Y (4, 50)	Y (78)	Y (73)	
buzzatii	Y (18, 100)	Y (100)	Y (100)	
meridiana	Y (8, 78)	Y (100)	Y (100)	
anceps	N	Y (63)	N	3
eremophila	Y (9, 96)	Y (100)	Y (100)	
Cluster				
mojavensis	N	Y (56)	N	5
ritae	N	N	N	6
longicornis	N	N	N	8
mulleri	N	N	N	5
martensis	Y (4, 83)	Y (95)	Y (90)	
buzzatii	Y (11, 99)	Y (92)	Y (100)	
stalkeri	Y (3, 63)	Y (59)	Y (78)	

Note. See also Fig. 1. BS, Bremer support; BP, bootstrap; MONO, number of extra steps required to make a nonmonophyletic group monophyletic.

tremely poor. For instance, there are 19 nodes in the tree (Fig. 1) with bootstrap values greater than 80%. Of these 19 strongly supported nodes only 2 unite species from different clusters, and both of these occur within the *mulleri* subgroup.

Separate Analyses and Conflicting Signals among Molecular Partitions

To assess the relative contribution of the various gene regions to the simultaneous analysis (SA) hypothesis we analyzed each gene separately and calculated both partitioned Bremer support and incongruence length differences. The general pattern which emerges from these analyses indicates significant disagreement between, on the one hand, COII and ND1 and, on the other hand, *hb*. All of the separate analyses are characterized by low consistency (Table 3) and a low degree of resolution (Fig. 2). As with the SA trees, the individual gene trees exhibit strongly supported relationships only among closely related taxa. All of the genes show substantial topological disagreement with each other and with the SA hypothesis (Fig. 1). Only 4 nodes on the SA tree (numbers 3, 13, 14, and 16 in Fig. 1) appear in each of the separate analyses (Fig. 2). The *hb* tree shows the greatest topological similarity to the SA tree, with 17 nodes in common, followed by COII (13), ND1 (10), and 16S (9). The separate analyses of each of the three mitochondrial genes are particularly divergent from the SA topology in several areas. For instance, within the COII strict consensus (Fig. 2), *D. virilis* has a highly derived placement within the ingroup, *D. uniseta* is strongly separated from the other *buzzatii* complex species, and *D. hydei* and *D. micromettleri* are sister taxa. In the ND1 consensus tree (Fig. 2), *D. virilis* again has a highly derived placement within the ingroup, and members of the *anceps* and *eremophila* complexes are united with the most distantly related outgroup taxa. The 16S consensus tree (Fig. 2) places the outgroup taxa at several derived positions, and representatives from the *hydei* subgroup, *repleta* subgroup, and *anceps* complex are widely separated.

Because of the poor resolution in the separate analyses it is difficult to establish to what degree individual genes support relationships that emerge in the combined analysis. Partitioned Bremer support provides one means for assessing the contribution of different genes to the SA topology. For this data set, the genes differ significantly in the extent to which they contribute to the total Bremer support of the SA hypothesis. Relative to its size, 16S provides over twice the support of any of the other genes, whereas ND1 provides virtually no support (Table 4). Differences in support provided by the various genes may result from either differences in internal homoplasy within each gene or conflicting signals among the genes. An incongruence

TABLE 3

Tree Statistics for Individual and Combined Data Partitions

	Tot. No. chars.	No. PI	No. trees	Steps	CI	RI
Inv	119	40	4	80	0.89	0.93
CO II	442	154	72	972	0.25	0.47
ND1	129	40	60	217	0.29	0.60
16S	521	71	550	257	0.37	0.58
hb	527	196	1710	848	0.40	0.61
Total	1738	501	2	2491	0.31	0.52

Note. Tot. No. chars., total number of characters in the data partition; No. PI, number of phylogenetically informative characters in the data partition; No. trees, number of most parsimonious trees obtained in the analysis; Steps, length of most parsimonious trees; CI, consistency index; RI, retention index.

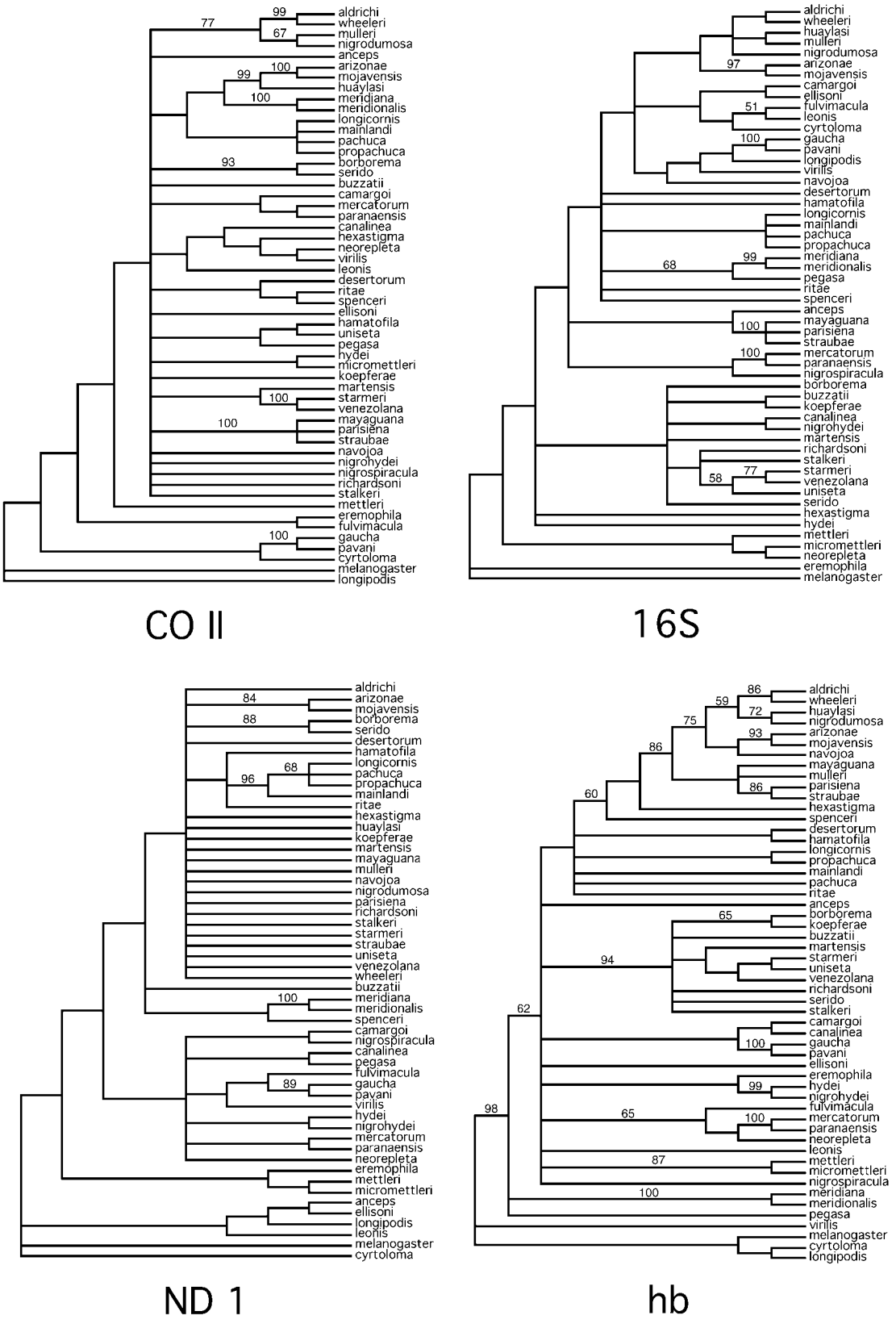

FIG. 2. Strict consensus of most parsimonious trees from analyses of individual gene partitions. Bootstrap values are provided above each node.

TABLE 4

Partitioned Bremer Support Scores Summed across the Simultaneous Analysis Trees (Fig. 1) for Each of the Gene Partitions and Standardized by the Minimum Number of Steps for Each Partition

Gene	Summed PBS	Min steps	Summed PBS/Min steps
COII	86.05	243	0.35
16S	79.39	96	0.82
ND1	6.08	63	0.10
hb	111.75	336	0.33

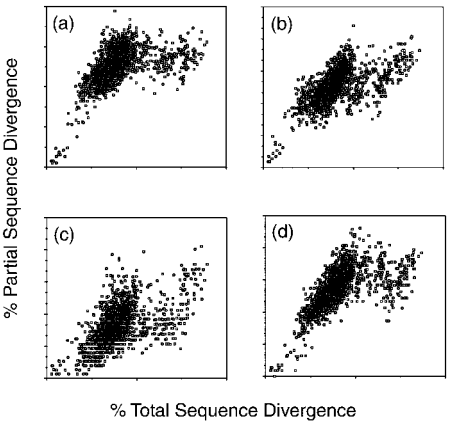

FIG. 3. Saturation plots of percentage total sequence divergence on the *X*-axis versus percentage partial sequence divergence for third positions (a), transitions (b), first and second positions (c), and transversions (d). Spreadsheets were created in Microsoft Excel with the appropriate variables from PAUP 4.01b (Swofford, 1998) and graphed in Excel.

length test for all the genes combined indicates significant disagreement (standardized ILD = 0.047, $P = 0.01$). Table 5 shows the results of ILD tests for all pairwise gene comparisons and for each gene against the rest of the data combined. The pattern of ILD scores from Table 5 suggests that the major area of conflict is between the mitochondrial protein coding genes (COII-ND1) and *hb*. The *hb* gene is significantly incongruent with both mitochondrial genes, which are not different from each other. A breakdown of the partitioned Bremer support values at each node also indicates substantial conflict between these partitions. Of the 42 resolved nodes on the SA consensus tree, 25 have conflicting PBS values (i.e., negative values for one partition, positive values for the other) for COII-ND1 and *hb*.

Given the conflict between COII-ND1 and *hb*, it is important to identify whether this disagreement is isolated to specific characters or taxa. Several studies have demonstrated differences in rates of evolutionary change among different classes of molecular data (Brown *et al.*, 1982; DeSalle *et al.*, 1987; Helm-Bychowski and Cracraft, 1993; Knight and Mindell, 1993), and this result is often used as a criterion for downweighting these characters in phylogenetic analysis (Martin, 1995; Yoder *et al.*, 1996; Murphy and Collier, 1997; Bloomer and Crowe, 1998; Danforth and Ji, 1998; Martin and Bermingham, 1998; Simons and Mayden, 1998). Saturation curves have generally been used as a means for assessing relative rates of change for different types of character data, although the utility of these plots has recently been questioned (Zang, 1998; Allard *et al.*, 1999; Baker *et al.*, submitted). A comparison of pairwise sequence distances for *hb* and COII-ND1 (Fig. 3) shows a marked contrast in the pattern of divergences between the two partitions. Divergences for *hb* are linear with respect to the total amount of sequence change, whereas COII-ND1 divergences clearly asymptote at a point approximately corresponding to ingroup–outgroup comparisons. Figure 4 breaks down the saturation in COII-ND1 with respect to codon position and transitions/transversions. Figure 4a suggests that third position sites are changing at slightly faster rates than first and second position sites because the most closely related taxa have the highest ratio of third position distances to first and second position distances. A similar pattern exists for transitions relative to transversions (Fig. 4b). This type of pattern of sequence change is often used as a justification for downweighting the effects of third po-

TABLE 5

ILD (Incongruence Length Difference) Values and Significance of Value for Pairwise Data Partition Combinations

	CO II	16S	ND1	hb	Rest
CO II	—	NS	NS	0.01	0.01
16S	0.038	—	NS	NS	NS
ND1	0.035	0.067	—	0.02	NS
hb	0.036	0.028	0.034	—	0.01
Rest	0.031	0.013	0.014	0.020	—

Note. The upper half of the matrix provides the significance values (NS, not significant) and the lower half of the matrix provides the ILDs standardized by the length of the most parsimonious tree(s).

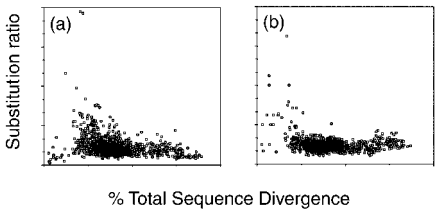

FIG. 4. Saturation plots of percentage total sequence divergence on the *X*-axis versus substitution ratio on the *Y*-axis for third positions divided by first and second position (a) and for transitions divided by transversions (b). Spreadsheets were created in Microsoft Excel with the appropriate variables from PAUP 4.01b (Swofford, 1998) and graphed in Excel.

sition sites and transitions (Martin, 1995; Martin and Bermingham, 1998; Simons and Mayden, 1998).

A more detailed analysis of the incongruence patterns, however, suggests that the disagreement between *hb* and COII-ND1 is not simply caused by the faster evolving characters. This is supported by several results. First, within COII and ND1, the phylogenetic behavior from the faster evolving characters does not differ substantially from that of the slower evolving characters. Relative to the rest of the data, first and second positions sites and transversions for COII-ND1 show a pattern of saturation similar to third position sites and transitions (Fig. 3). There is also no significant conflict, within COII and ND1, between the first and second positions sites and third position sites (ILD = 31, $P = 0.94$). Second, the removal of third position characters or transitions does not eliminate the conflict with *hb*. COII-ND1 partitions including only first and second positions sites (ILD = 38, $P = 0.01$) or transversions (ILD = 56, $P = 0.01$) are significantly incongruent with *hb*. Third, the disagreement between COII-ND1 and *hb* is not isolated to the most divergent taxa. The pairwise distance plots for COII-ND1 (Fig. 3) asymptote approximately at the point of ingroup–outgroup comparisons. Saturation curves for the ingroup taxa alone exhibit a linear pattern of change. Therefore, if saturated characters are the primary cause of conflict between COII-ND1 and *hb*, we would expect agreement between the partitions for analyses including only ingroup taxa. To examine this possibility, we conducted ILD tests for both the repleta group species and the *mulleri* complex species (node 40 in Fig. 1). For both of these subtaxa analyses, there is still significant disagreement between COII-ND1 and *hb* (*repleta:* ILD = 39, $P = 0.01$; *mulleri:* ILD = 26, $P = 0.01$). Overall, the saturated characters in COII-ND1 provide only a limited explanation for the disagreement with *hb*, and there is little evidence that removing these characters improves phylogenetic estimation.

Inversions and Congruence

We found that the inversion data conflict with the simultaneous analysis hypothesis at two nodes (nodes 19 and 26) while the combined molecular partitions conflict with the simultaneous analysis hypothesis at three nodes (nodes 29, 40, and 41). The total Bremer support from inversions at all nodes in the simultaneous analysis tree is 29 and for the molecular partition is 270. If these values are standardized by dividing the total Bremer support by the minimum steps for each partition (Baker *et al.*, 1998), both the inversion partition and the molecular partition contribute 0.37 Bremer support units each per phylogenetically informative character. In addition, the consistency indices of the inversion data, whether analyzed alone or forced onto the SA hypothesis, are three times higher than the consistency indices for the molecular data [CI (for inversions) = 0.89, CI (for molecules) = 0.30 for data analyzed alone; CI (for inversions) = 0.89, CI (for molecules) = 0.29 for data forced on the SA tree]. Because a consistency index of one is an indication of no homoplasy in a data set, the CI (for inversions) = 0.89 appears to suggest some conflict with the basic assumption of Wasserman (1992) that chromosomal inversions are unique events. However, we may conclude that the inversion data set contains a greater amount of information than the molecular data set as it possesses a far lesser degree of homoplasy than the mo-

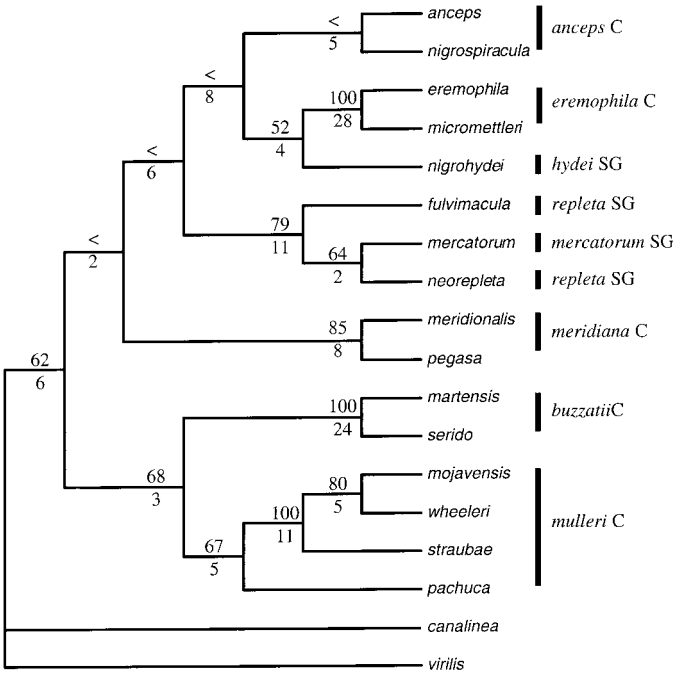

FIG. 5. Single most parsimonious tree from combined analysis including Ef1α using the pruned data set. Bootstrap values are provided above each node and Bremer support values below each node. SG, subgroup; C, complex.

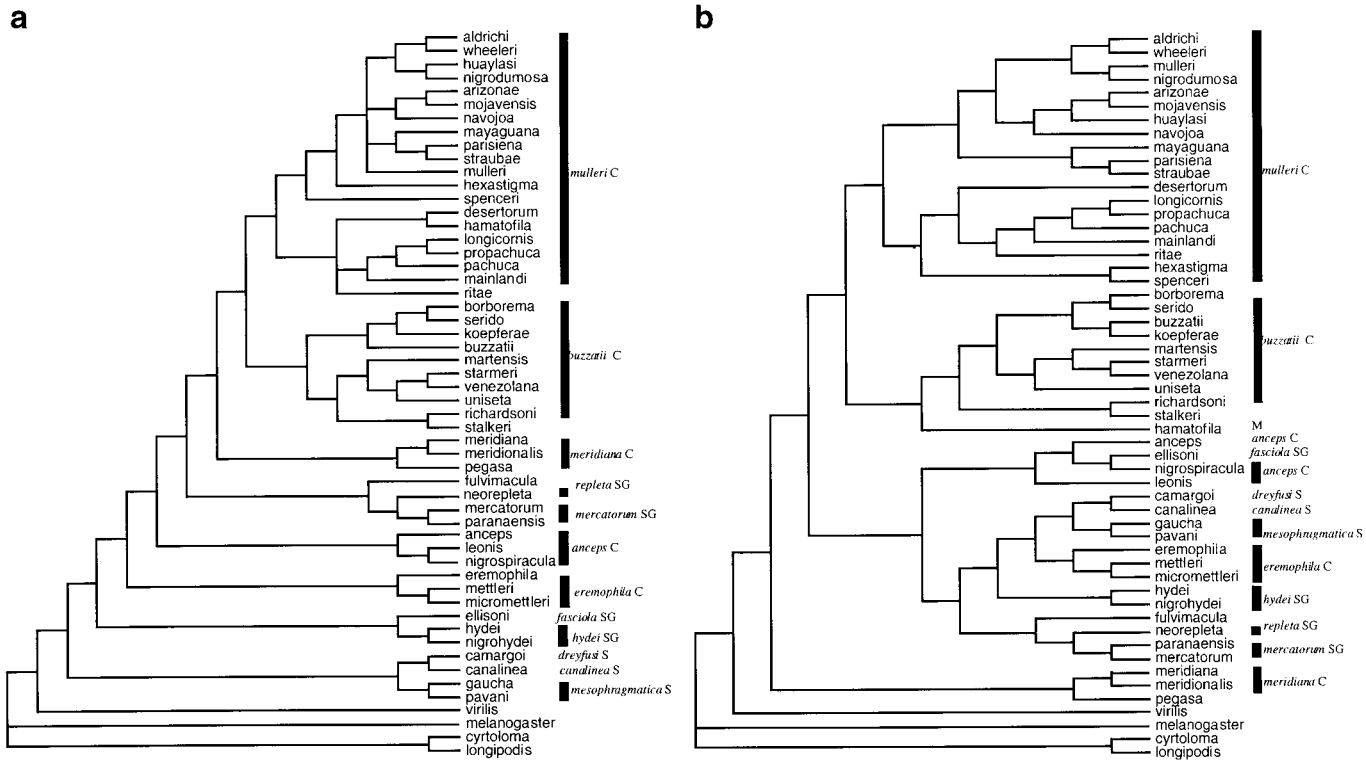

FIG. 6. (a) Tree generated by removing third position and all transition changes from the COII partition. (b) Tree generated using successive approximation weightings based on the retention index.

lecular partition. Finally, we note that there is no significant incongruence between the molecular and inversion partitions (ILD = 0.011, NS).

A Phylogenetic Hypothesis for the Relationships of Species Subgroups in the repleta Species Group Using Ef1α

Due to the lack of resolution at the base of our tree and the lack of inference on the species subgroup relationships, we constructed a second data matrix by adding Ef1α (Cho *et al.*, 1995) sequences. Because of the slow rate of evolution of Ef1α sequences, we decided to sequence only a subset of taxa from our original sampling to see if the Ef1α sequences could add resolution to the cladogram. We used representatives from four of the five species subgroups. Because Fig. 1 suggests that the *mulleri* subgroup is not monophyletic, we included representatives of all five species complexes in this subgroup. Table 1 shows the taxa for which we obtained Ef1α sequences. Phylogenetic analysis of this pruned data set resulted in a single parsimony tree (CI = 0.42, RI = 0.37, steps = 1568) shown in Fig. 5. This cladogram demonstrates the utility of the Ef1α sequences in resolving relationships in this problematic clade as all nodes are resolved and have relatively high Bremer support, although bootstrap values were low.

The cladogram shows the *mulleri* subgroup splitting into two major clades as observed in Fig. 1. The *mulleri* and *buzzatii* complexes are sisters in one of the monophyletic groups and the *meridiana*, *anceps*, and *eremophila* complexes reside in the second monophyletic group. The *hydei* subgroup representative is observed as the sister taxon to the *eremophila* complex of the *mulleri* subgroup (BS = 4 and BP = 57%). The *mercatorum* and *repleta* subgroups are observed as members of a well-supported clade (BS = 11 and BP = 79%) that is sister to the *hydei, eremophila, anceps* clade (BS = 6 and BP = 50%). The *meridiana* complex is the most basal representative of this clade. While Ef1α adds to the resolution of the relationships of taxa in this problematic clade, other data will probably be required to resolve the relationships of these species subgroups.

CONCLUSIONS

The present study suggests that the *repleta* species group is not monophyletic. However, this inference may be the result of poor resolution at the base of the phylogenetic tree we obtained (Fig. 1). Other existing taxonomic groups, such as subgroups, complexes, and clusters (Table 1) in the *repleta* lineage, were also examined for monophyly. In general, these other taxonomic groupings are upheld by our analyses (Table 2). A notable exception, however, is the monophyly of the

mulleri species subgroup, which is paraphyletic at best and most likely polyphyletic (see Figs. 1 and 6). However, this could be explained by the fact that Wasserman used the *mulleri* subgroup as a "waste paper basket" to include several complexes and clusters which he could not place elsewhere. At the level of relationships of species subgroups, our results indicate close affinity of the *repleta* and *mercatorum* subgroups. A well-resolved clade composed of the *mulleri* and *buzzatii* species complexes and a poorly resolved mixed taxonomic group composed of a mixture of *mulleri* subgroup species complexes and all other species subgroups, as well as flies from other species groups (*dreyfusi*, *canalinea*, and *mesophragmatica*), arise from our study. This poorly resolved grouping (Fig. 1) is the result of the consensus of two rather different arrangements of the taxa in this clade at the base of our tree. Although character weighting of the COII partition slightly increases recovery of previously accepted monophyletic groups (Fig. 6), successive weighting does not. Resolution in this area of the tree in an equal weighting framework will come from addition of more characters.

Analyses of the relative utility of the different partitions we used in this study indicate extreme saturation in the protein coding third positions and in transition changes, especially in the mitochondrial genes. These results suggest that mitochondrial protein coding regions are not adequate to resolve these basal relationships in an equal weighting analysis. On the other hand, addition of nuclear protein coding genes such as *hb* and Ef1α should be useful in deciphering relationships at the base of the phylogeny of this important group of drosophilids.

REFERENCES

Allard, M. W., Farris, J. S., and Carpenter, J. M. (1999). Congruence among mammalian mitochondrial genes. *Cladistics* **15:** 75–84.

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

Baker, R. H., Yu, X., and DeSalle, R. (1998). Assessing the relative contribution of molecular and morphological characters in simultaneous analysis trees. *Mol. Phylogenet. Evol.* **9:** 427–436.

Bloomer, P., and Crowe, T. M. (1998). Francolin phylogenetics: Molecular, morphobehavioral, and combined evidence. *Mol. Phylogenet. Evol.* **9:** 236–254.

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.

Brower, A. V. Z., DeSalle, R., and Vogler, A. (1996). Gene trees, species trees, and systematics: A perspective. *Annu. Rev. Ecol. Syst.* **27:** 423–450.

Cho, S., Mitchell, A., Regier, J. C., Mitter, C., Poole, R. W., Friedlander, T. P., and Zhao, S. (1995). A highly conserved nuclear gene for low-level phylogenetics: Elongation factor-1a recovers a morphology-based tree for Heliothine moths. *Mol. Biol. Evol.* **12:** 650–656.

Danforth, B. N., and Ji, S. (1998). Elongation factor-1α occurs as two copies in bees: Implications for phylogenetic analysis of EF-1α sequences in insects. *Mol. Biol. Evol.* **15:** 225–235.

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

DeSalle, R., Williams, A. K., and George, M. (1993). Isolation and characterization of animal mitochondrial DNA *In* "Methods in Enzymology" (E. A. Zimmer, T. J. White, R. L. Cann, and A. C. Wilson, Eds.), Vol. 24, pp. 176–204. Academic Press, New York.

Eriksson, T. (1997). AutoDecay Version 2.9.8 (Hypercard stock disk distributed by the author). Botoniska Institutionen, Stockholm Univ., Stockholm.

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

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

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

Fontevila, A., Wasserman, M., Pla, C., Pilares, L., de Armengol, R., Suyo, M. P., Sanchez, A., Vasquez, J., Ruiz, A., and Garcia, J. L. (1990). Description and evolutionary relationships of two species of the *Drosophila mulleri* cluster (Diptera: Drosophilidae). *Ann. Entomol. Soc. Am.* **83:** 444.

Gatesy, J., DeSalle, R., and Wheeler, W. C. (1994). Alignment-ambiguous nucleotide sites and the exclusion of data. *Mol. Phylogenet. Evol.* **2:** 152–157.

Gibson, T., Higgins, D., and Thompson, J. (1994). ClustalX. Program and documentation available at ftp://ftp.ebi.ac.uk/pub/software/mac/clustalw/clustalx.

Heed, W. B., Sanchez, A., Armengol, R., and Fontdevila, A. (1990). Genetic differentiation among island populations and species of cactophilic *Drosophila* in the West Indies. *In* "Ecological and Evolutionary Genetics of *Drosophila*" (J. S. F. Barker, W. T. Starmer, and R. J. MacIntyre, Eds.), pp. 447–490. Plenum Press, New York.

Helm-Bychowski, K., and Cracraft, J. (1993). Recovering phylogenetic signal from DNA sequences: Relationships within the corvine assemblage (Class Aves) as inferred from complete sequences of the mitochondrial DNA cytochrome-b gene. *Mol. Biol. Evol.* **10:** 1196–1214.

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

Knight, A., and Mindell, D. P. (1993). Substitution bias, weighting of DNA sequence evolution, and the phylogenetic position of fea's viper. *Syst. Biol.* **42:** 18–31.

Kwiatowski, J., and Ayala, F. J. (1999). Phylogeny of *Drosophila* and related genera: Conflict between molecular and anatomical analyses. *Mol. Phylogenet. Evol.* **13:** 319–328.

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

Martin, A. (1995). Mitochondrial DNA sequence evolution in sharks: Rates, patterns, and phylogenetic inferences. *Mol. Biol. Evol.* **12:** 1114–1123.

Martin, A. P., and Bermingham, E. (1998). Systematics and evolution of lower Central American cichlids inferred from analysis of cytochrome b gene sequences. *Mol. Phylogenet. Evol.* **9:** 192–203.

Murphy, W. J., and Collier, G. E. (1997). A molecular phylogeny for Aplocheiloid fishes (Atherinomorpha, Cyprinodontiformes): The role of vicariance and the origins of annualism. *Mol. Biol. Evol.* **18:** 790–799.

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

Richardson, R. H., Richardson, M. E., and Smouse, P. E. (1975). Evolution of electrophoredtic mobility in the *Drosophila mulleri* complex. *In* "Isozymes IV: Genetics and Evolution" (C. L. Markert, Ed.), pp. 533–545. Academic Press, New York.

Richardson, R. H., and Smouse, P. E. (1976). Patterns of molecular variation. I. Interspecific comparison of electromorphs in the *Drosophila mulleri* complex. *Biochem. Genet.* **14:** 447–465.

Richardson, R. H., Smouse, P. E., and Richardson, M. E. (1977). Patterns of molecular variation. II. Associations of electrophoretic mobility and larval substrate within species of the *Drosophila mulleri* complex. *Genetics* **85:** 141–154.

Ruiz, A., and Wasserman, M. (1993). Evolutionary cytogenetics of the *Drosophila buzzatii* species complex. *Heredity* **70:** 582–596.

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

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., and Flook, P. (1994). Evolution, weighting and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Ann. Entomol. Soc. Am.* **87:** 651–710.

Simons, A. M., and Mayden, R. L. (1998). Phylogenetic relationships of the western North American phoxinins (Actinopterygii: Cyprinidae) as inferred from mitochondrial 12S and 16S ribosomal RNA sequences. *Mol. Phylogenet. Evol.* **9:** 308–329.

Spicer, G. S. (1995). Phylogenetic utility of the mitochondrial cytochrome oxidase gene: Molecular evolution of the *Drosophila buzzatii* species complex. *J. Mol. Evol.* **41:** 749–759.

Spicer, G. S., and Pitnick, S. (1996). Molecular systematics of the *Drosophila hydei* subgroup as inferred from mitochondrial DNA sequences. *J. Mol. Evol.* **43:** 281–286.

Sullivan, D. T., Atkinson, P. W., Bayer, C. A., and Menotti-Raymond, M. (1990). The evolution of Adh expression in the *repleta* group of *Drosophila*. *In* "Ecological and Evolutionary Genetics of *Drosophila*" (J. S. F. Barker, W. T. Starmer, and R. J. MacIntyre, Eds.), pp. 407–418. Plenum Press, New York.

Swofford, D. L. (1999). PAUP 4.01: Phylogenetic analysis using parsimony (and other methods), version 4.01. Sinauer.

Throckmorton, L. H. (1982a). Pathways of evolution in the genus *Drosophila* and the founding of the *repleta* group. *In* "Ecological Genetics and Evolution: The Cactus–Yeast–*Drosophila* Model System" (J. S. F. Barker and W. T. Starmer, Eds.), pp. 33–48. Academic Press, New York.

Throckmorton, L. H. (1982b). The *virilis* species group. *In* "The Genetics and Biology of *Drosophila*" (M. Ashburner *et al.,* Eds.), Vol. 3b, pp. 227–296. Academic Press, London and New York.

Treier, M., Pfeifle, C., and Tautz, D. (1989). Comparison of the gap segmentation gene *hunchback* between *Drosophila melanogaster* and *Drosophila virilis* reveals novel modes of evolutionary change. *EMBO J.*

Vilela, C. R. (1983). A revision of the *Drosophila repleta* species group (Diptera, Drosophilidae). *Rev. Brasil. Entomol.* **27:** 1–114.

Vogler, A. P., DeSalle, R., Assmann, T., Knisley, C. B., and Schultz, T. D. (1993). Molecular population genetics of the endangered tiger beetle *Cicindela dorsalis* (Coleoptera: Cicindelidae). *Ann. Entomol. Soc. Am.* **86:** 142–152.

Wasserman, M. (1982). Cytological evolution in the *Drosophila repleta* species group. *In* "Ecological Genetics and Evolution: The Cactus–Yeast–*Drosophila* Model System" (J. S. F. Barker and W. T. Starmer, Eds.), pp. 49–64. Academic Press, New York.

Wasserman, M. (1992). Cytological evolution of the *Drosophila repleta* species group. *In* "*Drosophila* Inversion Polymorphism" (C. B. Krimbas and J. R. Powell, Eds.), pp. 455–552. CRC Press, Boca Raton, FL.

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

Yoder, A. D., Cartmill, M., Ruvolo, M., Smith, K., and Vilgalys, R. (1996). Ancient single origin for Malagasy primates. *Proc. Natl. Acad. Sci. USA* **93:** 5122–5126.

Zouros, E. (1973). Genetic differentiation associated with the early stages of speciation in the *mulleri* subgroup of *Drosophila*. *Evolution* **27:** 601–621.