
ROB DESALLE AND ANDREW V. Z. BROWER

Department of Entomology, American Museum of Natural History, Central Park West at 79th Street, New York, New York 10024-5192, USA; E-mail: desalle@amnh.org (R.D.)

As data from multiple sources (morphology, different genes) become available for inferring relationships among taxa, the problem of how these data should be analyzed has emerged as a topic of renewed theoretical discussion (Miyamoto, 1985; Kluge, 1989; Barrett et al., 1991; Bull et al., 1993; de Queiroz, 1993; Lanyon, 1993; Chippindale and Wiens, 1994; de Queiroz et al., 1995; Miyamoto and Fitch, 1995; Brower et al., 1996; Hedges and Maxson, 1996; Huelsenbeck et al., 1996; Nixon and Carpenter, 1996; Page, 1996). Philosophically, the problem is straightforward. Phylogenetic inference depends on evidence from characters that imply hierarchical patterns of grouping. Not all features of organisms are of equal evidential value for inferring relationships, so some form of discrimination between "good" and "bad" characters is necessary. However, such differential character weighting requires ad hoc assumptions about the nature of empirical evidence. Many of the current disputes in systematics revolve around the advantages and disadvantages of alternative weighting schemes (even the debate between advocates of maximum likelihood and advocates of cladistic parsimony can be viewed as a difference in willingness to make assumptions about differential weighting of character types and character state transformations). In our view, the goal of systematics is to produce phylogenetic hypotheses with the greatest explanatory power (in the Popperian sense). This explanatory power is reduced by the hypothesis' contingency on supernumerary background assumptions. Although no systematic method is free of assumptions, a scheme that minimizes ad hoc assumptions about character weighting is preferred on philosophical grounds (Farris, 1983; Kluge, 1989; Brower et al., 1996).

In this paper, we develop three theoretical arguments against the use in systematics of topological consensus methods as a means of combining data from different sources, with particular emphasis on sequence data from different genes. Because treatment of data from different sources as belonging to evidentially different classes implies a crude and poorly justified differential weighting scheme whereas combining all characters in a single analysis does not, the latter is preferable based on the criterion of greater explanatory power. Of course, one may be interested in aspects of data structure other than its evidential value in phylogenetic inference, in which case exploratory data partitioning is very important (most of the analyses described in this paper are analyses of partitioned data). Although comparing results of separate analyses of subsets of the data may be fundamental to understanding character evolution, the best hypothesis of relationships is derived from simultaneous analysis of all relevant data.

We illustrate some of the salient features of this problem by analyzing the relationships among four closely related species of Hawaiian Drosophila, as implied by data from DNA-DNA hybridization, mitochondrial DNA (mtDNA) sequences and restriction fragment length polymorphisms, and sequences from six nuclear gene regions (Table 1).

The Study System

The species quartet of Drosophila differens (D), D. heteroneura (H), D. planitibia (P), and
TABLE 1. Drosophila data subsets examined in this study. The approximate number of bases sequenced or restriction sites scored and the number of phylogenetically informative sites among these is given.

<table>
<thead>
<tr>
<th>Data source</th>
<th>Abbreviation</th>
<th>No. characters</th>
<th>Informative sites</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>mtDNA restriction sites</td>
<td>RS</td>
<td>61</td>
<td>36</td>
<td>DeSalle and Giddings, 1986</td>
</tr>
<tr>
<td>Cytochrome c oxidase II</td>
<td>COIL</td>
<td>446</td>
<td>5</td>
<td>this study</td>
</tr>
<tr>
<td>Alcohol dehydrogenase</td>
<td>ADH</td>
<td>1,632</td>
<td>33</td>
<td>Rowan and Hunt, 1991</td>
</tr>
<tr>
<td>Yolk protein 1</td>
<td>ypl</td>
<td>1,350</td>
<td>10</td>
<td>Kambysellis et al., 1995</td>
</tr>
<tr>
<td>Vestigial</td>
<td>vg</td>
<td>473</td>
<td>5</td>
<td>this study</td>
</tr>
<tr>
<td>Acetylcholinesterase</td>
<td>ACHE</td>
<td>347</td>
<td>3</td>
<td>this study</td>
</tr>
<tr>
<td>Wingless</td>
<td>wg</td>
<td>450</td>
<td>0</td>
<td>this study</td>
</tr>
<tr>
<td>Hunchback</td>
<td>hb</td>
<td>350</td>
<td>0</td>
<td>this study</td>
</tr>
</tbody>
</table>

D. silvestris (S) is a closely related clade of flies in the Hawaiian Drosophila picture-winged species group. Drosophila silvestris and D. heteroneura are found only on the Big Island of Hawaii, and D. planitibia and D. differens are endemic to Maui and to Molokai, respectively. The age of Molokai gives an upper limit to the age of this clade at 1.5 million years, but the actual divergence of the species may have been considerably more recent (DeSalle et al., 1987).

These four taxa have been the target of intense morphological, behavioral, and genetic research over the past 30 years (reviewed by Carson, 1982, 1987; Kaneshiro, 1983; DeSalle and Hunt, 1987; Kaneshiro and Boake, 1987; Grimaldi and Fenster, 1989). There are no fixed phylogenetically informative chromosomal or allozyme differences among the four species (Craddock and Johnson, 1979; Carson, 1982, 1987), and population-level analysis of mtDNA restriction sites suggests that D. silvestris is paraphyletic with respect to D. heteroneura, although both species are phylogenetically diagnosable (DeSalle and Vogler, 1994).

Using mtDNA restriction site data at the species level, DeSalle and Giddings (1986) found the pattern of relationship (D(P(H, S))) among these four species to be topologically incongruent with that from the nuclear DNA–DNA hybridization data ((D, P)(H, S)) (Hunt and Carson, 1983). Based on an unsubstantiated evolutionary scenario, DeSalle and Giddings (1986:6906) suggested that a more parsimonious alternative to explain the nuclear and mtDNA discordance is that the mtDNA phylogeny represents the more accurate evolutionary history of these species, and the nuclear DNA phylogeny is indicative of a lack of differentiation of nuclear genetic components of D. differens and D. planitibia.

However, the invocation of ad hoc arguments to justify the mtDNA topology over the DNA–DNA hybridization topology was unnecessary. DNA–DNA hybridization provides no character information but instead gives pairwise measures of similarity that may or may not have anything to do with hierarchical patterns of relationship among taxa (Cracraft, 1987; Mindell, 1992). DNA–DNA hybridization dendrograms should be ignored if hypotheses based on character-based data exist (Brown et al., 1996).

The absence of data from morphology or other molecules that could be compared with the mtDNA topology led to provisional acceptance of the cladogram of DeSalle and Giddings (1986) for the four Drosophila species. However, more recent sequence information from the ADH locus (Rowan and Hunt, 1991; Thomas and Hunt, 1993) and the vitellogenin gene ypl (Ho, 1994; Kambysellis et al., 1995) supports the ((D, P)(H, S)) relationship implied by DNA–DNA hybridization and contradicts the mtDNA restriction site topology. Now that there are additional data implying different relationships among the taxa, a thorough examination of congruence and support for alternative topologies is warranted. In this study, we added new
characters from mtDNA and four nuclear genes to the already published data. Here, we discuss the implications of each of these sources for inferring the pattern of phylogenetic relationships.

**Materials and Methods**

**Flies**

Individual lines used for the quartet of species examined in this study are the same for all of the gene partitions that we report. We used the Waikamoi *D. planitibia* line (U84Y), the Hanalilolilo *D. differens* line (U43V1), the Hilo side *D. heteroneura* line (Q71G12), and the Kona side *D. silvestris* line (U26B9), all provided by K. Kaneshiro. For the new sequences, our outgroup is the *D. hemipeza* line (W33J) provided by M. Kambysellis. The *ADH* (Rowan and Hunt, 1991) and *ypl* (Kambysellis et al., 1995) gene sequences were obtained from GenBank. For *ADH* we used the published *D. picticornis* sequence as an outgroup. Given the morphological and mtDNA diagnosability of putative sister species *D. heteroneura* and *D. silvestris* (DeSalle and Vogler, 1994), we were willing to draw phylogenetic inferences using single lines as exemplars of their respective species.

**DNA Isolation and Manipulation**

DNA was isolated from single flies using the small scale DNA prep described by DeSalle et al. (1993). PCR of fly template DNA was performed using primers designed for the genes listed in Table 1. Primers for mtDNA were published by Brower (1994b); those for new nuclear genes were designed by Baker and DeSalle (1997). Sequences from the PCR products were generated in three different ways: (1) the PCR products were cloned using the TA cloning vector (Invitrogen) and directly sequenced from double-stranded (ds) plasmid DNA using a dideoxy sequencing kit (US Biochemical) and 35S; (2) ds PCR product was directly sequenced with 35S; and (3) the ds PCR product was sequenced using the automated dideoxy sequencing system on an ABI 373. Manual sequences were read and verified with MacVector 3.5 (International Biotechnologies). The automated sequences were edited using MacClade (Maddison and Maddison, 1992) and visual inspection of the chromatographs that resulted from the sequencing runs.

**Alignment**

New sequences were aligned with the Malign software (Wheeler and Gladstein, 1994). Alignments for mtDNA COII, *wg*, and *hb* were trivial because the regions sequenced contained no indels. The *vg* and *ACHE* regions were aligned with gap: change costs of 2 and 4, which yielded a single stable alignment. In *vg*, none of the inferred indels are phylogenetically informative, but the inferred *ACHE* indels fall in a single small region. We replaced this area in the analysis with a single multistate gap character at the end of the data matrix. Published alignments were used for the *ADH* and *ypl* sequences (Rowan and Hunt, 1991; Ho, 1994). We removed two phylogenetically informative gap regions in the *ADH* sequence and recoded them as single multistate characters. These gap codes are presented in Figure 1. All aligned sequences are available in NEXUS format on the AMNH Web page (http://research.amnh.org/molecular).

**Phylogenetic Analyses**

All characters were weighted equally. All phylogenetic analyses were performed using exhaustive searches in PAUP 3.1 (Swofford, 1993). Branch support (Bremer, 1988, 1994) was calculated by subsparsimonious search, and bootstrap proportions (based on 100 replications) were calculated using PAUP. Unambiguous character support for branches was estimated for ACC-TRAN trees based on the numbers of characters supporting the branches with rescaled consistency indices of 1.0. The degree of incongruence among different data subsets was analyzed using ARN (Farris et al., 1994, 1995) with 999 replications. Incongruence length differences (ILDs) were calculated as discussed by Mickevich and Farris (1981) and Farris et al. (1994, 1995).
Results
Phylogenetic Information in the Various Data Subsets

Table 1 lists the gene regions studied and the number of characters in each. The data that existed prior to this study were the mtDNA restriction site data (DeSalle and Giddings, 1986), the ADH data (Rowan and Hunt, 1991), and the ypl data (Ho, 1994; Kambysellis et al., 1995). Given prior hypotheses of close relationship among the four Drosophila species, we targeted sequences from gene regions thought to be rapidly evolving and thus potentially yielding useful phylogenetic information (Brower and DeSalle, 1994). We hypothesized that the intron region of ACHE (Fournier et al., 1994), the hypervariable regions of wg (Rijsewijk et al., 1987; Kraft and Jackie, 1994) and hb (Tautz et al., 1987), the 5′ regulatory region of vg (Williams et al., 1994), and the rapidly evolving mitochondrial gene COII (Beckenbach et al., 1993; Brower, 1994a) would contain variable sites that might be phylogenetically informative. However, the recent divergence of these four taxa is reflected by the small number of variable sites detected among these five genes. The regions of wg and hb we sequenced contain no phylogenetically informative sites at all. The vg and ACHE regions we sequenced showed limited variation, with five and three informative positions, respectively (Table 1). In the nuclear genes sequenced, most of the variation and all of the phylogenetic information is found in noncoding regions. In the mtDNA sequence, there are 23 variable sites, all at third positions. Eighteen of these sites are transitions, and the remainder are A-T transversions. Five of these sites are phylogenetically informative. Overall, the raw ratio of phylogenetically informative positions to the number of bases sequenced is 56:4,806 (1/85).

Congruence among Cladograms from Individual Genes

For four taxa, 15 bifurcating rooted cladograms are possible. Separate phylo-
genetic analysis of the various data subsets recovered two of these cladograms that differ only in the placement of *D. differens* (Fig. 2), the M (mitochondrial) topology (recovered by restriction sites and COII) and the N (nuclear) topology (recovered by ADH, *yp1*, and *vg* and by distance analyses of allozyme and DNA–DNA hybridization data [Craddock and Johnson, 1979; Hunt and Carson, 1983]). The M and N topologies were equally parsimonious in the analysis of the *ACHE* data. Table 2 lists tree statistics for cladograms from the different gene regions.

Brower et al. (1996) classified systematic incongruence into four types and presented a flow chart for evaluating incongruence among different data sets. The first two types are trivial, involving cases of claimed incongruence in which no phylogenetic analysis has been conducted (type I) or in which data are not amenable to comparison (type II). Type III is topological incongruence based on weak character support, which can be detected by statistical testing of the ILD (or alternative methods, e.g., Larson, 1994; Huelsenbeck and Bull, 1996). Filtering out these types of pseudo-incongruence leaves the fourth type, which is significant character incompatibility between alternative data subsets (sensu Farris, 1971; our type IV incongruence).

We used this scheme to compare the various Hawaiian *Drosophila* data subsets. The behavioral data (Kaneshiro, 1983) were not described in a manner amenable to recoding as explicit characters for cladistic analysis, so the information from that source failed the type I incongruence assessment and was not considered further. Topologies based on allozyme frequencies (Craddock and Johnson, 1979) and DNA–DNA hybridization values (Hunt and Carson, 1983) are not comparable at the character level to the rest of the data (type II incongruence) and were therefore not useful for inferring patterns of hierarchical relationship. All the sequence data and the mtDNA restriction site data are comparable sets of char-

---

**Figure 2.** Cladograms from separate analyses of different *Drosophila* data subsets. Branch support is indicated on internal nodes. Mitochondrial (M) or nuclear (N) topologies are indicated. P = *D. planitibia*; D = *D. differens*; H = *D. heteromeura*; S = *D. silvestris*; O = outgroup.
TABLE 2. Branch support (BR), unambiguous character support (U), and bootstrap proportions (BS) for the two Drosophila hypotheses (M and N topologies, Fig. 2). The H + S clade appears in both the M and the N topologies. There is no branch support for the P + D clade in the gene partitions that support the M topology because that clade does not occur in that hypothesis. Likewise, there is no branch support for the H + S + P clade in the gene partitions that support the N topology.

<table>
<thead>
<tr>
<th>Node</th>
<th>All mtDNA</th>
<th>RS</th>
<th>COII</th>
<th>All nuclear</th>
<th>ADH</th>
<th>ACHF</th>
<th>vg</th>
<th>ypl</th>
<th>All data</th>
</tr>
</thead>
<tbody>
<tr>
<td>H + S</td>
<td>BR</td>
<td>9</td>
<td>6</td>
<td>12</td>
<td>15</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>U</td>
<td>15</td>
<td>10</td>
<td>6</td>
<td>19</td>
<td>1</td>
<td>4</td>
<td>7</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>BS</td>
<td>98</td>
<td>99</td>
<td>63</td>
<td>100</td>
<td>72</td>
<td>100</td>
<td>98</td>
<td>100</td>
</tr>
<tr>
<td>H + S + P</td>
<td>BR</td>
<td>7</td>
<td>6</td>
<td>1</td>
<td>NR*</td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>U</td>
<td>17</td>
<td>12</td>
<td>5</td>
<td>NR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BS</td>
<td>94</td>
<td>92</td>
<td>63</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P + D</td>
<td>BR</td>
<td>6</td>
<td>4</td>
<td>NR</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>U</td>
<td>13</td>
<td>9</td>
<td>NR</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BS</td>
<td>95</td>
<td>93</td>
<td>NR</td>
<td>67</td>
<td>83</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Topology</th>
<th>M</th>
<th>M</th>
<th>M</th>
<th>N</th>
<th>N</th>
<th>M, N</th>
<th>N</th>
<th>N</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>CI</td>
<td>0.71</td>
<td>0.72</td>
<td>0.70</td>
<td>0.86</td>
<td>0.86</td>
<td>0.80</td>
<td>1.00</td>
<td>0.85</td>
<td>0.75</td>
</tr>
<tr>
<td>RI</td>
<td>0.56</td>
<td>0.61</td>
<td>0.48</td>
<td>0.82</td>
<td>0.82</td>
<td>0.67</td>
<td>1.00</td>
<td>0.80</td>
<td>0.63</td>
</tr>
<tr>
<td>Length</td>
<td>90</td>
<td>50</td>
<td>40</td>
<td>66</td>
<td>43</td>
<td>5</td>
<td>5</td>
<td>13</td>
<td>162</td>
</tr>
</tbody>
</table>

* NR = not resolved.
CI = consistency index.
RI = retention index.

Actors. The three published studies, mtDNA restriction analysis (DeSalle and Giddings, 1986), ADH sequences (Rowan and Hunt, 1991), and the ypl sequences (Kambysellis et al., 1995), all presented explicit data matrices and both cladistic and phenetic analyses, and all these data were reanalyzed cladistically here.

We tested the data from the various sources for significant (type IV) incongruence using the ILD (Mickevich and Farris, 1981), and statistical significance of the ILDs was calculated using the ARN program described by Farris et al. (1994, 1995). Table 3 summarizes the results of these tests. Only the two data subsets with the strongest support for alternative hypotheses show significant incongruence (ADH vs. mtDNA restriction sites; \( P < 0.05 \)). The ILD of the six gene regions with informative characters was also assessed in a simultaneous, six-way test, but this ARN score is not significant (\( P > 0.05 \)). When the six nuclear genes or the mitochondrial restriction site and sequence data are grouped together, a single tree is found for each of the two subsets (Fig. 2; Table 2). The ILD for the lumped mitochondrial versus lumped nuclear data sets is 6, and the ARN test suggests that this value is significant (\( P < 0.05 \)).

Consensus and Simultaneous Analysis

Given the significant ARN result between the mitochondrial and nuclear subsets, it is possible to take four alternative phylogenetic approaches (Huelsenbeck et al., 1996). Perhaps the most conservative way to summarize the data is in a taxonomic congruence framework (Miyamoto and Fitch, 1995), as represented by a strict consensus of the two fundamental parsimony trees obtained from analysis of data subsets (Fig. 3a). Alternatively, the combinable-component (CC) consensus (Bremer, 1990) can be used to represent the union...
of data sets, retaining clades that do not contradict one another. For this data set, the CC and strict consensus trees are the same. Another alternative is the Bull et al. (1993) view that because there is significant incongruence between the mitochondrial and nuclear partition, the two incongruent topologies (Fig. 2) should be kept separate, with the ambiguity of the group's systematic status reflected by retaining both hypotheses of relationship. A fourth approach, simultaneous analysis of all the characters (total evidence; Miyamoto, 1985; Kluge, 1989; Nixon and Carpenter, 1996), produces the cladogram in Figure 3b.

**DISCUSSION**

Although our small sample of taxa and the correspondingly small number of possible topologies suggest that inferring relationships among these four *Drosophila* species should be a simple problem, our data place that problem at the crux of the total evidence issue. If the individual gene trees were all congruent, that would be satisfying taxonomically but not very interesting from a theoretical standpoint. However, the data are not congruent, and the mtDNA topology is favored in simultaneous analysis of all data. Because we have no independent empirical basis for dismissing the mtDNA evidence as a manifestation of hypothetical ancestral polymorphism or lineage sorting, as is frequently done in the literature (reviewed by Avise, 1994; Brower et al., 1996), we accept it as our best hypothesis of relationships, even though several ostensibly independent nuclear gene regions support a single alternative topology.

This acceptance of the mitochondrial topology may seem counterintuitive. Brower et al. (1996) advocated the alternative N topology ((D, P)(H, S)) as the best hypothesis of relationships for the group based on just the restriction site, *ADH*, and *ypt* data. The contingency of the outcome on the character subsets included is obvious. We could conclude that phylogenetic relationships of these four taxa are not resolvable and that accepting the total evidence hypothesis is an inferior approach because it may be giving us the "wrong" tree. One of the three other alternatives (strict consensus, combinable consensus, conditional combination) could be viewed as a more accurate representation of the state of our knowledge.

Here, we outline our reasons for preferring the total evidence hypothesis and offer critiques of arguments supporting the alternatives. Our discussion focuses on three issues: the validity of process partitions, the independence of data subsets, and the interpretation of consensus trees as summaries of competing hypotheses of relationships. Our views are based on a substantial foundation of prior studies (e.g., Farris, 1983; Kluge, 1989; Swofford, 1991; Eernisse and Kluge, 1993; Wenzel and Carpenter, 1994; Nixon and Carpenter, 1996), but the continued controversy surrounding data partitions suggests that the logical basis of simultaneous analysis is in need of further discussion.

**Process Partitions**

An effective critique of our choice of topologies would demonstrate that partitioned analyses are logically and practically superior to the simultaneous analysis approach, that something is gained by dividing data into process partitions for phylogenetic analysis. How might this be accomplished? Bull et al. (1993:392) defined process partitions as "a division of char-

**FIGURE 3.** *Drosophila* combined data trees. P = *D. planitibia*; D = *D. differens*; H = *D. heteroneura*; S = *D. silvestris*; O = outgroup. (a) Taxonomic congruence tree, based on strict consensus of cladograms in Figure 2. The combinable component consensus has the same topology. If data from *wg* and *hb* were included, the strict consensus would be completely unresolved. (b) Single most-parsimonious cladogram from simultaneous analysis of all data. Branch support is indicated; see Table 2 for other summary statistics.
acters into two or more subsets such that characters in each subset have evolved according to rules that are demonstrably different from those in other subsets." They demonstrated these rules by statistical comparison of topologies derived from different parts of the data matrix: different process partitions support different trees. Among the types of possible process partitions Bull et al. identified are different genes, protein-coding versus noncoding regions within genes, and first and second versus third positions within protein-coding regions. A flaw in the Bull et al. concept is the lack of criteria for recognizing the number and boundaries of process partitions (Kluge and Wolfe, 1993). For example, does our *Drosophila* data set represent two process partitions (mtDNA vs. nuclear), or three (restriction sites vs. protein-coding sequences vs. noncoding sequences), or eight (the separate gene regions), or more, if we consider within-gene differences in evolutionary process such as silent versus nonsynonymous sites? Once data partitioning is started, it is not logical to stop seeking significantly different partitions until all of them are discovered, yet the lower limit of possible data subdivision seems set only by the minimum number of characters per partition needed for partitions to differ statistically from one another (Miyamoto and Fitch, 1995). One could resort to the ad hoc procedure of identifying process partitions as an outcome of phylogenetic analysis or some statistical test (i.e., process partitions are those groups of characters that support a topology other than the most-parsimonious one), but that approach would be similar to clade analysis, which has already been soundly critiqued (Farris and Kluge, 1979, 1985) and abandoned by most systematists. Given these ambiguities, the process partition concept as defined by Bull et al. seems confusing and nonoperational.

Much of the controversy surrounding data partitioning has sprung from attempts to rescue molecular data that provide inadequate resolution of the clades being examined or that produce "incorrect" reconstructions of a "known" phylogeny (Miyamoto and Cracraft, 1991; de Queiroz, 1993; Huelsenbeck et al., 1996; Sullivan, 1996). Adding auxiliary hypotheses to rein in homoplastic or otherwise intractable characters after the fact reduces the explanatory power of the data (Popper, 1959:145). To avoid such unparsimonious rationalizations, we recommend judicious a priori selection of characters (gene regions), which we have referred to as stage 1 sieving (Brower and DeSalle, 1994). From a practical perspective, sources of characters we suspect of not providing enough phylogenetic information to be worth the effort or of containing confounding levels of homoplastic variation within terminals should be avoided at the outset (a procedure that is routine for morphologists). Once characters are admitted, the status of primary homologies in the data matrix (de Pinna, 1991), they may not be capriciously rejected because they fail to support preconceived notions of relationship (Brady, 1983).

Miyamoto and Fitch (1995) applied the notion of process partitions in a slightly different way, to articulate the "naturalness" of subsets of data from separate sources, such as the different genes we have employed in this study. They claimed (1995:67) that "most systematists agree that their data sets conform to natural divisions among characters, thereby leaving taxonomic congruence as a viable alternative to character congruence." This view of data partitions is completely different from that of Bull et al. (1993), which is based on inferred evolutionary processes. For example, Bull et al. might consider different nucleotide positions in the same gene to represent different process partitions, whereas Miyamoto and Fitch's view implies that a protein-coding DNA sequence and its corresponding amino acid sequence do not represent separate "natural classes" of characters. We support Miyamoto and Fitch's view and have referred to the different gene regions we studied as alternative data subsets. We recognize that these subsets are not random groups of characters but that they represent practical partitions of the data that result from dif-
ferent sampling procedures. However, this distinction does not address the question of whether the practical recognition of data from different genes as separate implies anything about the independence of these genes as separate lines of phylogenetic evidence.

The Independence of Data Subsets

It is generally held that the more characters support a clade, the more plausible is the hypothesis that the clade represents a natural group. But can the same be said of groups of characters or “independent data sets”? Miyamoto and Fitch (1995) and many others (e.g., Mickevich, 1978; Wheeler, 1991; Lanyon, 1993; Brower, 1994a) have suggested that there is a relative increase in the probability of a tree being “true” if separate hypotheses of phylogeny from different data are congruent with one another. In blocked experimental designs, there is a gain in statistical power associated with increasing the number of replicates because of the increased degrees of freedom and the greater proportion of the total variance being ascribable to error. Thus, 10 replicates of 100 observations may provide greater confidence in a particular pattern in the data than 2 replicates of 500 observations. Perhaps it is an analogue to this increase in statistical power that Lanyon (1993:47) sought when he argued that “when a phylogenetic hypothesis is supported by several independent lines of evidence, we gain confidence in it as an estimate of phylogenetic history” and that Miyamoto and Fitch (1995:70) identified as “the ‘something more’ that systematists seek when they analyze their data sets separately.”

A troubling corollary to the view that data partitions are independent is the implication of nonindependence of data within partitions. Most systematic methods, and certainly those that rely on statistics, depend upon the independence of characters (Farris, 1983; Felsenstein, 1983). For data partitions to exhibit independence from one another as partitions, they must be reified, thus subjugating their component characters as parts or aspects of the larger entity. Characters become nonindependent as parts of independent partitions. Following this line of reasoning leads to the conclusion that the more character support resides within a partition, the less weight each of those characters should add to a hypothesis of relationships because those observations are tainted by character correlation simply by virtue of being parts of an integrated whole. Such a view is obviously counterproductive for producing well-resolved phylogenetic hypotheses.

The idea that there is one natural hierarchy, or history, is the central postulate of modern comparative biology (Brady, 1985, 1994; Panchen, 1992; Bull et al., 1993). This is why the various tree-building algorithms are designed to discover hierarchical patterns. Although different characters may be drawn from physically different sources (in terms of how they are observed), the data can no longer be divided into independent data sets for phylogenetic inference because all bits of evidence are a priori assumed to be parts of a whole, i.e., manifestations of the same hierarchical pattern that was used to infer the existence of evolutionary processes in the first place. Tests such as ARN, Larson’s (1994) homoplasy contingency test, or the likelihood ratio test of Huelsenbeck and Bull (1996) are useful for exploring patterns in data but cannot serve as criteria for determining if (and which) evidence should be deleted or downweighted. In our Drosophila study, for example, the data are not significantly incongruent in an eight-partition comparison but are incongruent if the six nuclear genes and two mtDNA data sets are compared as two partitions. Thus, the assumptions of the partitioning strategy and not the evidence itself can be responsible for observed “incongruence” (the circularity of the Bull et al. [1993] method to discover separate process partitions is evident from this example). Furthermore, the discovery of significantly different partitions provides no criterion for favoring one topology over another (Brower, 1996). Once multiple incongruent topologies are admitted as competing hypotheses of relationship, the premise of a single hierar-
chy is violated, and some method other than character congruence becomes necessary to obtain a single working answer.

**Topological Congruence and Confidence**

If these arguments against data partitioning are valid, the rationale for using consensus methods to assess topological congruence between data partitions is eliminated because there is no logical basis for claiming that subsets of characters represent separate lines of evidence. Nevertheless, some might argue that a cladogram from combined data (e.g., Fig. 3a) offers no means to assess the quality or reliability of the result (e.g., Lanyon, 1993). Branch support values (Bremer, 1988, 1994) are provided on our cladograms to at least partly address this problem, i.e., the greater the branch support, the stronger the clade is corroborated by the data and the more robust the clade is to the addition of potentially contradictory data. Thus, the sister-taxon relationship of *D. heteroneura* and *D. silvestris* is corroborated by all the gene regions independently (Fig. 2) and, even more strongly, by the simultaneous analysis of all the data (Fig. 3). Note the synergistic effect of combining the data on support for the H + S node: the sum of branch support values from the individual gene trees is 34, whereas the simultaneous analysis tree has a value of 40. Another popular method used to assess the support for branches is bootstrapping (see Sanderson, 1995). We prefer branch support over bootstrapping for two reasons. First, branch support is based on reanalysis of all the data, whereas bootstrapping is based on reanalysis of subsamples of the data. Second, bootstrap proportions are easily misinterpreted as statistical confidence intervals for clades, whereas branch support values are not.

There are several further points about topological congruence that should be considered. We are not suggesting that strict consensus trees should not be used to summarize the shared data content of multiple equally parsimonious cladograms from a single data analysis. In fact, this is our preferred method for presenting results when more than one most-parsimonious cladogram is obtained. Rather, we object to the topological comparison of results from independent analysis of separate data subsets. There are other circumstances in which topological congruence methods can be heuristically useful, such as the comparison of trees from disparate methods or data sources that are not comparable by character congruence (the severe limitations of distance-based data and quantitative traits for phylogenetic analysis have been discussed at length [e.g., Farris, 1981, 1985; Mindell, 1992; Bookstein, 1994] and will not be reviewed here). Brower et al. (1996) argued that such data are not relevant to phylogenetic hypotheses based on explicit character data and should be employed only as preliminary conjectures of relationship to be tested with subsequent data.

Several authors (e.g., Miyamoto, 1985; Hillis, 1987; Doyle, 1992; Lanyon, 1993; Wägele, 1995) have suggested that large numbers of molecular characters may overwhelm smaller morphological data subsets. The number of nucleotides is not equivalent to the number of cladistic characters, so the potential for bias may not be as severe as it appears. Topological congruence is a crude means to adjust the weights of characters by weighting partitions, instead of characters, equally. Various other weighting schemes to adjust the contribution of alternative data types have been advocated based on the data themselves or on extrinsic models of evolution (e.g., Cracraft and Helm-Bychowski, 1991; Doyle, 1992; Chippindale and Wiens, 1994; Miyamoto et al., 1994; Simon et al., 1994). These approaches suffer from the same compromising ad hoc assumptions about independence as discussed above (Farris, 1983; Brady, 1985; Eernisse and Kluge, 1993; Kluge and Wolf, 1993; Brower and DeSalle, 1994).

Topological congruence is often viewed as a conservative test of data informativeness. For this view to be accurate, strict consensus must be employed (for a review of consensus methods, see Swofford, 1991). This position is contrary to that of Lanyon...
(1993), who argued that strict consensus was an "unduly pessimistic" method for summarizing shared topological structure between cladograms. He advocated a modified version of Bremer's (1990) CC consensus method, which allows retention of nodes for which there is no conflicting resolution among the topologies being compared. In terms of set theory, CC consensus can be thought of as the union of compatible resolved nodes from alternative topologies, whereas strict consensus is the intersection. CC consensus is not a conservative method and leads to serious error when employed as Lanyon used it, to construct meta-consensus trees of consensus trees from separate cladistic analyses that yielded multiple equally parsimonious fundamental cladograms.

This difficulty arises because strict (and CC) consensus trees conflate two types of polytomies: those representing zero-length branches, which have no conflicting character support, and those representing multiple topologies with equally parsimonious but conflicting character support. If strict consensus trees of the latter type are subsequently used to build CC meta-consensus trees (as Lanyon [1993] did in his Me-nidia example), information about character conflict within data partitions is lost, and the resulting "phylogenetic framework" is more resolved than the CC tree of the fundamental cladograms from which it was derived. This result obviously is not acceptable. If the goal of performing consensus manipulations is to obtain a more "robust" topology than the original data yield from simultaneous analysis of all data, then only strict consensus trees should be employed because the absence of information in one data set provides no corroborate of any kind for topologies built from another data set. This approach will inevitably result in the loss of phyllogenetic resolution, as is apparent in the comparison of trees in Figure 3 (for a more dramatic empirical example, see Miller et al., 1997).

The Bottom Line

The fundamental task of systematics is to infer the pattern of hierarchical relationship among taxa. With the growth of molecular data bases, systematists have arrived at the point where they can obtain too much information to make these inferences unambiguously, even with fast computer algorithms. Thus, it is necessary to decide which characters are important and which are not, i.e., to deal with the problem of differential character weighting. We are not opposed to differential weighting in principle (indeed, the initial selection of characters is a weighting procedure in itself), but we think that weights should be evaluated for individual characters based on reasoned hypotheses of homology rather than applied wholesale to poorly substantiated classes of characters based on deduction from general "truths" about the evolutionary process (transversion parsimony is a good example of the latter; see Brower and Desalle, 1994; Allard and Carpenter, 1996). All indications are that the evolutionary process is vastly more complex than the models we have devised to explain it to date (Maynard Smith and Smith, 1996; Philippe et al., 1996), and it is not clear that even the "true" evolutionary model always provides the optimal explanation of the data (Yang, 1996, 1997).

We have used empirical results from Hawaiian Drosophila to illustrate a more general discussion of the controversies that surround analysis of systematic data from heterogeneous sources. From a practical perspective, only one of the methods discussed, simultaneous analysis of all characters, yields a single most-parsimonious cladogram of relationships for the DHPS clade. Of course, this hypothesis is not "robust," and additional data from nuclear genes might sway the weight of the evidence from the (D(P(H, S)) topology to the alternative ((D, P)(H, S)) topology. Then again, they might not.

The message is this: given some set of data from one or more sources, there are an infinite number of ways to perform analyses and establish phylogenetic scenarios of greater or lesser plausibility. The diversity of possible methods is only limited by the imaginations of those who find it useful to invent them. Choosing among
these methods is a philosophical problem that lies outside the realm of systematics or statistics, perhaps between the realms of logic and rhetoric. We present our views here because we hope to convince some we are right and to challenge those who are not convinced to try to show us why we are wrong. It is incumbent upon systematists to debate these issues and reach agreement on a single method of choice, or our Tree of Life will bear very strange and chimeric fruit indeed.

ACKNOWLEDGMENTS

We thank R. Baker, D. Cannatella, C. Durando, B. Farrell, J. Miller, and four anonymous reviewers for their thoughtful comments, which improved the manuscript. The research was supported by postdoctoral fellowships from the American Museum of Natural History and the Smithsonian Institution to A.V.Z.B. and National Science Foundation grant BSR-9220317 to J. S. Miller and R.D.

REFERENCES


MIYAMOTO, M. M., AND J. CRACRAFT. 1991. Phyloge-


Received 20 February 1997; accepted 27 May 1997

Associate Editor: David Cannatella