Data Sets, Partitions, and Characters: Philosophies and Procedures for Analyzing Multiple Data Sets

J. William O. Ballard,1,2 Margaret K. Thayer,1 Alfred F. Newton, Jr.,1 and Elizabeth R. Grismer3,4

1Department of Zoology, Field Museum, Roosevelt Road at Lake Shore Drive, Chicago, Illinois 60605, USA
2Department of Zoology, Field Museum, Roosevelt Road at Lake Shore Drive, Chicago, Illinois 60605, USA; E-mail: ballard@fmnh.org
3Biochemistry Laboratory, Field Museum, Roosevelt Road at Lake Shore Drive, Chicago, Illinois 60605, USA
4Stratagene Cloning Systems, Protein Dept. 11011 N. Torrey Pines, La Jolla, California 92037, USA

Abstract.—We compared four approaches for analyzing three data sets derived from staphylinoid beetles, a superfamily whose known species diversity is roughly comparable to that of vertebrates. One data set is derived from adult morphology and the two molecular data sets are from 12S ribosomal RNA and cytochrome b mitochondrial DNA. We found that taxonomic congruence following conditional data combination, herein called compatible evidence (CE), resolved more nodes compatible with an initial conservative hypothesis than did total evidence (TE), conditional data combination (CDC), or taxonomic congruence (TC). CE sets a base of nodes obtained by CDC analysis and then investigates what further agreement may arise in a universe where these nodes are accepted as given. We suggest that CE75-75 may be appropriate for future studies that aim to both generate a well-corroborated tree and investigate conflicts between data sets, partitions, and characters. CE75-75 is a 75% bootstrap consensus CDC tree followed by combinable-component consensus of a 75% bootstrap consensus of each homogeneous set of partitions having hierarchical structure. [Beetles; compatible evidence; conditional data combination; Staphylinidae; taxonomic congruence; total evidence.]

The growth of systematics and the availability of independent data sets raises the question of how these data sets should be analyzed. Kluge (1989) suggested that all data sets should be analyzed simultaneously according to a total evidence (TE) or simultaneous analysis (Nixon and Carpenter, 1993) procedure to maximize informativeness. This approach seeks a single best-fitting hypothesis that, in cladistics, involves maximizing character congruence. However, each data set may have different levels of hierarchical structure (Faith and Cranston, 1991), be subject to distinct biological forces (Bull et al., 1993), and be informative at different divergence levels (Friedlander et al., 1994). As a result, no single criterion or model may be applicable to all sets of evolutionary processes. Chippindale and Wiens (1994) argued that differences in character evolution can be accommodated by differential character weighting in a combined analysis of all data. They acknowledged, however, that the appropriate weighting scheme is as unknown as the true phylogeny.

A step toward the resolution of this quandary is to define data partitions a priori. It may be argued that partitions recognized by evolutionary biologists do not conform to natural sets of characters but are merely artifacts of tradition and technology. The naturalness of different partitions can be defended by relating their existence to the concept of evolutionary processes defined by Bull et al. (1993). These authors defined process partitions as subsets of characters that are evolving under demonstrably different rules. There are at least three reasons why partitions may appear to evolve differently: (1) The signal in one or more partitions may be swamped by homoplasy, (2) distinct and conflicting processes may be operating, or (3) methods for investigating whether process partitions should be combined may be inadequate.

Faith and Cranston (1991) developed the combined hierarchical evidence (CHE) procedure (Faith, personal communication) to address the potential problem of signal being swamped by homoplasy. In CHE the data are partitioned, each independent partition is tested for structure using the permutation probability (PTP) test, and a tree is constructed from partitions with hierarchical structure.
Employing this procedure, Faith and Cranston (1991) showed that excluding data sets without PTP structure may resolve a higher number of nodes than including data sets both with and without PTP structure.

Second, distinct and conflicting patterns may result from such phenomena as introgression or selection. Bull et al. (1993) developed the conditional data combination (CDC) procedure to address this issue. Chippindale and Wiens (1994) refer to this approach as prior agreement. In CDC the data are partitioned into process partitions, each independent partition is tested for homogeneity, and combinable partitions are pooled and a tree constructed (Bull et al., 1993; de Queiroz, 1993; Rodrigo et al., 1993). Consider the nonmonophyly of the two mitochondrial haplotypes of *Drosophila mauritiana* with respect to the *D. simulans* mitochondrial haplotypes (Satta and Takahata, 1990). Alternative explanations for the observed nonmonophyly are the retention of ancestral polymorphisms or introgression. Inspection of Satta and Takahata’s (1990) sequence data show *D. mauritiana maI* differs from the *D. simulans siIII* haplotype by a single substitution. Moreover, the mtDNA of *D. mauritiana maI* and that of *D. simulans siIII* have no restriction site length polymorphism differences (Solignac and Monnerot, 1986). This suggests that there has been introgression of *D. simulans siIII* mtDNA into *D. mauritiana*. Laboratory crosses of *D. mauritiana* (males) and *D. simulans* (females) result in sterile males and fertile females (Robertson, 1983), supporting the feasibility of introgression.

Third, methods for investigating whether process partitions should be combined may be inadequate. Sullivan (1996) showed that, in deer mice and grasshopper mice, portions of 12S ribosomal RNA and cytochrome *b* have conflicting signals but a combined parsimony analysis with equal weights recovered a well-corroborated hypothesis. One explanation for this apparent contradiction is that the phylogenetic signal from the different loci with different evolutionary processes was additive when it was analyzed under a homogeneous model.

An alternative method for analyzing multiple data sets, taxonomic congruence (TC), was proposed by Mickevich (1978). This method involves partitioning data into discrete process partitions, analyzing the characters in each partition separately, and constructing a consensus tree that summarizes the topological features shared by trees from the separate analyses (Jones et al., 1993). The TC approach has many methodological advantages because each data set is treated independently, separate analyses can corroborate specific assemblages, and conflicts between process partitions can be directly investigated (Lanyon, 1985; Miyamoto and Fitch, 1995). However, consensus methods may endorse trees that contradict the tree obtained from the pooled data. Barrett et al. (1991) suggested that these situations “may not be rare” and in such circumstances the consensus trees should not be regarded as the “best” inference to make from the available data.

We investigated a new approach to the analysis of multiple data sets, herein called compatible evidence (CE). This procedure can be broken down into six steps and involves a series of familiar elements of systematic analysis (Fig. 1). CE sets a base of nodes obtained by CDC analysis and then investigates what further agreement may arise in a universe where these nodes are treated as given. The process is consistent with the tenets of meta-analysis and combines the philosophy of CDC with the methodology of TC. Meta-analysis originally was defined as statistical analysis of a large collection of individual data sets for the purpose of integrating the results (Glass, 1976). However, the term is often used more restrictively (e.g., Hedges and Olkin 1985; Dickersin and Berlin, 1992). In restrictive meta-analysis, a weighted combined analysis of all data from all studies is statistically compared to the separate analysis of each individual study. Consistent with meta-analysis, Hillis (1987) suggested that the best estimate of a phylogeny is obtained from a combined analysis, but that congruence among data sets is evidence that the underlying phylogeny is correctly estimated.

We constructed an initial conservative hypothesis and compared the TE, TC, CDC, and CE approaches by analyzing three independently derived data sets from staphylinoid beetles. One data set is derived from adult morphology and the other two data sets are from...
12S ribosomal RNA (12S rRNA) and cytochrome $b$ mitochondrial DNA (mtDNA). In this study we compare the number of well-supported nodes generated by the four techniques and prefer the method that supports the highest number of such nodes that are congruent with the a priori hypothesis (Miyamoto et al., 1994; Wheeler, 1995). CE resolved a higher number of nodes compatible with the initial conservative hypothesis than did any other procedure. Further, CE facilitates investigation of conflicts between data sets, partitions, and characters. However, CE may accept a tree that is not the most parsimonious and may generate a biased estimator of nodal support.

**MATERIALS AND METHODS**

**Current Taxonomy**

Staphylinoid beetles comprise a large component of the suborder Polyphaga, which is the largest and most diverse suborder within the Coleoptera. Polyphaga contain more than 90% of the known 350,000 species and 166 families of beetles.

The earliest fossil definitely attributable to Polyphaga is *Peltosyne triassica* from the Triassic of Central Asia (Arnoldi et al., 1977). However, Crowson (1975) and Ponomarenko (1969) suggested that the Triassic archostematian family Ademosynidae may actually represent basal Polyphaga in which the prothoracic pleuron had not yet become internalized (Lawrence and Britton, 1994). The fossil record of the superfamily Staphylinoidea has been recently extended to the Triassic (unnamed Staphylinoidea from eastern United States: Fraser et al., 1996) from previous estimates of the Middle or Lower Jurassic (Arnoldi et al., 1977; Ryvkin, 1985). Within the Staphylinoidea, the fossil record demonstrates that significant diversification had occurred.
by the upper Jurassic and Lower Cretaceous (e.g., Tikhomirova, 1968; Ryvkin, 1990).

Staphylinoidea, whose known species diversity worldwide (ca. 50,000) is comparable to that of vertebrates, owes its size primarily to several radiations within the Staphylinidae. These radiations have been accompanied by extraordinary habitat diversification, and likely include multiple origins of some morphological adaptations. For example, radiations of several distantly related subfamilies (Pselaphinae, Aleocharinae, Osoriinae, Staphylininae) in forest litter and soil habitats were probably preceded by separate feeding canalization in each of these groups (Leschen, 1993; Newton and Thayer, 1995).

The many classifications of Staphylinidae and Staphylinoidea have been reviewed by Lawrence and Newton (1982, 1995) and Newton and Thayer (1992, 1995). For this study, previous morphological evidence suggests that Agyrtidae is an appropriate outgroup. However, Leiodidae may be more closely related to Agyrtidae than to Staphylinidae. This is left unresolved in all initial analyses.

Initial Hypothesis

In this study, we do not have a known phylogeny. Rather, the initial hypothesis was developed prior to the cladistic analysis we report here to facilitate comparison of the procedures investigated. Our initial hypothesis (Fig. 2) was based on both adult and immature characters (Lawrence and Newton 1982; Ashe and Newton 1993; Newton and Thayer 1995; Newton, unpublished observations; Thayer, unpublished observations). However, only the subset of adult morphological data, without missing data points, was included in our methodological comparisons. The immatures of some taxa are not known, and missing data have the potential to bias our comparisons. Thus, our initial hypothesis is not completely independent of the morphological characters employed in the analysis of multiple data sets. We acknowledge this problem but suggest that the lack of independence is not a priori expected to bias one method over another. To test this prediction we specifically investigate nodes that conflict with the initial hypothesis. We conclude that the initial hypothesis cannot

**Figure 2.** Initial hypothesis of relationships within Staphylinidae for the taxa included in this study (based on Lawrence and Newton, 1982; Newton and Thayer, 1995; Thayer, 1985; Newton, unpublished observations). This is not the result of a quantitative cladistic analysis.
be rejected statistically by currently available data.

**Morphological Data and Voucher Specimens**

Specimens used in the morphological analyses were examined dry, in alcohol, or in cleared and disarticulated slide mounts using dissecting and compound microscopes, as appropriate. In this study we employ the set of adult external morphological characters used by Newton and Thayer (1995:254–268); see that work for a description of the characters and states. Invariant and/or uninterruptable characters are excluded (numbers 11, 18, 73, 84, 109, 111 of Newton and Thayer, 1995). Collection data, storage conditions, and identifications are described in Appendix 1.

The morphological characters are partitioned into two process partitions: head and nonhead characters (Appendix 2). These partitions were chosen because transitions between predation and nonpredation are suspected to have arisen multiple times. Therefore, it was hypothesized that distinct evolutionary forces may be acting in these partitions.

**Molecular Data**

DNA from individual beetles was extracted using the PureGene Kit (Gentra) following the “DNA Isolation From Fixed Tissue” protocol. Specimens had been stored in 95% ethanol for 6 months to 3 years (18 taxa), in 70% ethanol for < 1 year (3 taxa), or in 70% ethanol for 3 years (1 taxon) (Appendix 1, with locality data). Fragments were PCR (polymerase chain reaction) amplified for 35 cycles. Each cycle consisted of 30 sec denaturation at 94°C, 30 sec annealing between 45 and 50°C, and 1 min extension at 72°C. The amplicon was electrophoresed on a 1% agarose gel to verify size and the remainder of the reaction was cleaned and concentrated with Microcon 100 (Amicon). Fifty nanograms of template and 25 ng of primer were added to 4.25 μl of deionized formamide and 50 mM ethylenediamine tetraacetic acid (EDTA) (5:1 ratio) and 2 μl was electrophoresed on an Applied Biosystems 377 DNA sequencer. Sequences were imported into the Sequencher software program and the chromatograms were investigated.

12S rRNA.—The 12S rRNA locus was amplified with primers SR-J-14233 (alias 12Sbi) and SR-N-14588 (alias 12Sai) designed by T. Kocher (Kocher et al., 1989) and listed by Simon et al. (1994). The sequences were initially aligned using ClustalW (Higgins and Sharp, 1988). The secondary structure model proposed in Hickson et al. (1996) was then employed to refine the alignment. Sixty-three positions with ambiguous alignments were excluded from all analyses. Gaps were coded as an additional character. The 12S rRNA locus is divided into four process partitions according to Hickson et al. (1996): domain II, domain III stems, domain III loops, and domain IV (Appendix 3).

Cytochrome b.—The cytochrome b locus was amplified with primers CB-J-10933 (CB1) and CB-N-11367 (CB2) designed by Y. C. Crozier (Crozier and Crozier, 1992) and listed by Simon et al. (1994). Sequences were aligned using ClustalW (Higgins and Sharp, 1988) without any insertions or deletions. All nucleotides were included in the analysis. Amino acid composition was determined using MacClade 3.05 (Maddison and Maddison, 1992). Five cytochrome b process partitions were considered: first, second, and third codon positions, third-position transversions, and amino acids. These partitions are not all independent; only independent partitions are included in the final phylogenetic analyses.

**Phylogenetic Analysis**

The procedures for TE, TC and CDC have been summarized (Kluge, 1989; Bull et al., 1993; Jones et al., 1993; Miyamoto and Fitch, 1995). CE can be broken down into six discrete steps (Fig. 1). To facilitate comparison between the four techniques we have made slight modifications to TE, CDC, and TC, as discussed later.
We employed parsimony analysis using PAUP* 4.0d52 (provided by David Swofford), with one exception: We used MEGA (Kumar et al. 1993) to conduct a Jukes–Cantor neighbor-joining bootstrap analysis of cytochrome \( b \) replacement substitutions. We note that it is not essential to implement only one algorithm when constructing the TC and CE trees. On the contrary, it may behoove researchers to implement alternative methods when the branch lengths are highly uneven or there is a particular character bias (Felsenstein, 1978; Steel et al., 1993; Huelsenbeck, 1995).

PTP testing (Archie, 1989; Faith and Cranston, 1991) is integrated into each method. We employed the PTP test to investigate whether the observed tree length could have been obtained “by chance alone” (Archie, 1989; Faith, 1990; Carpenter, 1992; Faith, 1992).

We investigated nucleotide biases using the chi-square test of homogeneity of base frequencies across all taxa as implemented in PAUP* 4.0d52. Steel et al. (1993) found that PTP and bootstrap could provide misleading results where loci have independently acquired similar G + C compositions.

We used bootstrapping (Efron, 1982; Felsenstein, 1985), decay or support indices (Bremer, 1988; Donoghue et al., 1992), and the topology-dependent permutation (T-PTP) test (Faith, 1991) to investigate the monophyly of clades. We constructed 95% (Felsenstein and Kishino, 1993), 70% (Hillis and Bull, 1993), and 50% (majority rule) bootstrap consensus trees. For the T-PTP tests we employed 99 randomizations unless otherwise stated.

For CDC and CE we employed the incongruence length difference test of Farris et al., (1995) to investigate whether partitions are evolving under distinct biological processes. This random-partitioning test is an extension of a measure originally reported by Mickevich and Farris (1981) and is based on the null hypothesis of congruence.

For TC and CE we summarized the topological features from separate analyses by combinable-component consensus (Bremer, 1990). Among others, Lanyon (1993) and de Queiroz (1993) previously noted that only strongly supported nodes should be included in a final phylogenetic framework.

We investigated conflicts between partitions by employing decay indices (Bremer, 1988; Donoghue et al., 1992), T-PTP tests of nonmonophyly (Faith, 1991; Faith and Trueman, 1996; Swofford et al., 1996b), and compare-2 tests (Faith, 1991; Ballard, 1994; Swofford et al., 1996b; Faith and Trueman, 1996).

A Posteriori Phylogenetic Analysis

To further compare the four phylogenetic methods, we selected the lowest bootstrap percentage generating a topology consistent with the initial hypothesis. To be conservative, each bootstrap consensus value was rounded up to the nearest 5%. This (1) enables comparison of the approaches under the assumption that the initial hypothesis is correct and (2) facilitates prediction of bootstrap percentages that may be appropriate or future studies.

We were interested in comparing the tree lengths of the TE trees, the most parsimonious tree for each partition, and each maximally resolved a posteriori tree. We present tree lengths and statistically investigate length differences using the compare-2 PTP test (Faith, 1991; Ballard, 1994; Swofford et al., 1996b; Faith and Trueman, 1996) and the Templeton (1983) test. The latter utilizes a Wilcoxon signed-rank test of the relative number of steps required by each character on each of the respective trees.

Molecular Analysis

Saturation plots.—To further investigate the informativeness of each locus we plotted the number of substitutions in a specific partition (y-axis) against an estimate of time since divergence (x-axis). Distance estimates, using the Hasegawa et al. (1985) (HKY) model, were obtained from all alignable 12S rRNA and cytochrome \( b \) characters. We estimated gamma-distributed rates using PAUP* 4d52-3.

To investigate saturation between the outgroup and the ingroup, the pairwise divergences between \textit{Necrophilus} and all ingroup taxa and the pairwise divergences between all ingroup taxa were plotted with different symbols.

Substitution biases.—We employed the Wilcoxon test to compare the proportion of
nucleotides that occurred in the variable positions to the proportion in all positions. The substitution bias in the 12S rRNA stem partition may be expected to maintain stem stability. Ballard and Kreitman (1994) hypothesized that changes to C or G are slightly deleterious in insect mtDNA.

Results

Phylogenetic Analysis

Total evidence analysis.—The 871 characters in the three independently acquired data sets were pooled. These data are presented in Appendices 2, 3, and 4. The 12S rRNA and cytochrome b sequences have also been deposited in GenBank (accession numbers AFO21046-89). The 434 informative characters have significant structure as determined by PTP (Table 1). The resultant topology Fig. 3a substantially disagrees with our initial hypothesis (Fig. 2). In particular, there is conflict over the monophyly of the Omalinae (Amphichroum, Dropephylla, Eusphalerum). To investigate this apparent conflict we conducted a T-PTP test of Omalinae nonmonophyly. This test cannot reject monophyly of the Omalinae (T-PTP = 0.94, 1 step). Data not included in this analysis support monophyly of the Omalinae, specifically, larvae of Omalinae have a unique combination of derived characters including single-segmented urogomphi, five pairs of stigmata, and two pairs of laterosclerites on the first two or more abdominal segments (Newton and Thayer, 1995; Thayer, 1985).

Four nodes supported by bootstrap percentages of > 70 (TE70) are supported by the initial hypothesis (Figs. 2, 3b, 4). Also, two of the three nodes supported by 50% bootstrap consensus (TE50) are compatible with the initial hypothesis. Monophyly of (Habrocerus, Cyparium) supported by 50% bootstrap (Figs. 3b, 4) conflicts with the initial hypothesis (Fig. 2). T-PTP tests of nonmonophyly suggest these data can reject the hypothesis that (Tachinus, Habrocerus) are monophyletic (T-PTP = 0.03, 25 steps) but not that (Cyparium, Oxytelus) are monophyletic (T-PTP = 0.31, 14 steps). Further, data not included in this analysis support monophyly of (Oxytelus, Cyparium) and of (Tachinus, Habrocerus). Monophyly of (Oxytelus, Cyparium) is supported by loss of a vein in the anal area of the hind wing and specialized saprophagous/mycophagous mouthparts of adults and larvae. Derived morphological characters of larvae including deciduous urogomphi, epipharynx with asymmetrical setose fields (Newton, unpubl.; Ashe and Newton, 1993) and the traditionally accepted semilimuloid body shape unite (Tachinus, Habrocerus).

Taxonomic congruence analysis.—Each partition was analyzed independently and then a combinable-component consensus was constructed. For convenience we discuss the three independently acquired data sets separately.

Morphology.—The morphological data set has PTP structure (Table 1). This data set was divided into two partitions: head and nonhead characters. These partitions have significant PTP structure (Table 1). Bootstrapping the head partition supports two nodes at 95% level that are consistent with the initial
hypothesis. Bootstrapping the nonhead morphological characters supports the same two nodes as the head partition in addition to four other nodes at > 70% that do not conflict with the initial hypothesis (Figs. 2, 4). The nonhead and head partitions conflict over the sister taxon to (Conoplectus, Palimbolus) (Fig. 4). The nonhead partition supports monophyly of (Dasycerus (Conoplectus, Palimbolus)). The head morphological partition supports (Edaphus (Conoplectus, Palimbolus)). The initial hypothesis, decay indices, and compare-2 PTP testing support the former. In the non-head partition eight additional steps are required to decay the branch uniting Dasycerus with (Conoplectus, Palimbolus). Only two additional steps are required to decay the branch uniting Edaphus with (Conoplectus, Palimbolus) in the head partition. Two constraints were constructed for the compare-2 analysis, the first corresponding to (Dasycerus (Conoplectus, Palimbolus)) and the second to (Edaphus(Conoplectus, Palimbolus)). First, only the nonhead characters were included. Parsimony finds a shortest tree of 145 steps under the constraint of (Edaphus (Conoplectus, Palimbolus)) and 150 under the constraint of (Dasycerus(Conoplectus, Palimbolus)). In this case, the compare-2 PTP test fails to reject the null hypothesis that the result is due to chance (PTP = 0.06, 5 steps). Again, data not included in this analysis have some bearing on interpretation of the 50% bootstrap results. As discussed by Newton and Thayer (1995), derived larval characters support monophyly of both the Omaline Group (here Omaliiinae, Proteininae, Dasycriinae, and Pselaphinae) (tentorial bridge structure) and (Dasycriinae, Pselaphinae, Micropeplinae) (unarticulated urogomphi), to the exclusion of Eu aesthetinae, including Edaphus. Unfortunately, we did not have appropriate material to look at Dasycerus ovariole structure, which Newton and Thayer (1995) predicted would be of a type thus far unique to Pselaphinae + Proteininae + Micropeplinae (Welch, 1993).

In conflict with the initial hypothesis, the morphological nonhead partition groups Oxytelus and ((Opresus, Euconnus) (Edaphus, Dianous)) with a > 50% bootstrap (Fig. 4). T-PTP tests of nonmonophyly suggest these data cannot reject the hypothesis that (Cyparium, Oxytelus) and (Edaphus, Dianous, Opresus, Euconnus, Oxyporus, Achenomorphus,
Ontholestes, Oiceoptoma are each monophyletic (T-PTP = 0.06, 11 steps; T-PTP = 0.77, 11 steps respectively). Adult, larval, and pupal derived characters not included in this analysis (extraoral digestion with correlated mouthpart structures of adults and larvae; four pairs of functional pupal spiracles) support monophyly of the Staphylinine Group (here Edaphus through Oiceoptoma in Fig. 2). Monophyly of the Oxyteline Group (Oxytels and Cyparium in this study) is also supported by data not included in this analysis including loss of a vein in the anal area of the hind wing and specialized saprophagous/mycophagous mouthparts of adults and larvae (Lawrence and Newton, 1982; Newton and Thayer, 1995; Newton, unpublished observations).

Ontholestes, Oiceoptoma) are each monophyletic (T-PTP = 0.06, 11 steps; T-PTP = 0.77, 11 steps respectively). Adult, larval, and pupal derived characters not included in this analysis (extraoral digestion with correlated mouthpart structures of adults and larvae; four pairs of functional pupal spiracles) support monophyly of the Staphylinine Group (here Edaphus through Oiceoptoma in Fig. 2). Monophyly of the Oxyteline Group (Oxytels and Cyparium in this study) is also supported by data not included in this analysis including loss of a vein in the anal area of the hind wing and specialized saprophagous/mycophagous mouthparts of adults and larvae (Lawrence and Newton, 1982; Newton and Thayer, 1995; Newton, unpublished observations).

12S rRNA. — There is PTP structure in the 12S rRNA data set (Table 1). These characters were divided into four partitions: domain II, domain III stems, domain III loops, and domain IV characters. The former three partitions have PTP structure but the domain IV characters do not (Table 1). There are no nucleotide biases in the three partitions with PTP structure (domain II \( \chi^2_{63} = 25.92 \) and 37.57; domain III stems \( \chi^2_{63} = 25.37 \) and 38.62; domain III loops \( \chi^2_{63} = 22.62 \) and 28.28, \( P > 0.05 \) for all and variable positions,
respectively). One node resolved by each of the three partitions at the 70% bootstrap consensus level does not conflict with the initial hypothesis (Fig. 4).

Support for (Opresus, Edaphus) at the 50% bootstrap consensus level in the domain III stems partition conflicts with the initial hypothesis (Fig. 4). This support is at least partially due to compensatory changes that maintain stem stability (Fig. 5). Six synapomorphies unite (Opresus, Edaphus) in the three most parsimonious trees (Table 1) and 3 steps are required to decay the node. Two of these involve compensatory changes and, as a result, it may be argued that these positions should be down-weighted. At position 103 in stem 36 there is a substitution from C to U, while at position 218 there is a substitution from G to A (Fig. 5, Appendix 3). There are also compensatory changes in stem 48 (A → U at position 311 and U → A at position 324) (Appendix 3). T-PTP tests of nonmonophyly suggest these data cannot reject the hypothesis that (Opresus, Euconnus) and (Edaphus, Dianous) are monophyletic (T-PTP = 0.55, 4 steps; T-PTP = 0.55, 5 steps). Several derived characters (not included in this study) support the nodes (Edaphus, Dianous) and (Opresus, Euconnus) in the initial hypothesis (Newton, unpublished observations). Adults of the former have lost wing-folding patches from the abdominal terga; larvae have the mala highly reduced and finger-like and the apical maxillary palp segment with a membranous flexation point near the middle. Like other Scydmaenidae, Opresus and Euconnus have the adult elytra long, nearly or completely covering the abdomen (a secondary development), and all known scydmaenid larvae have club-shaped antennae, no ligula, three or fewer sternmata, and urogomphi fixed or absent (Newton, 1991; Opresus larvae are unknown). Several additional features that differ between these pairs of taxa are of uncertain polarity, so it is not clear which of the two pairs they support.

Cytochrome b.—There is no PTP structure when all cytochrome b characters are included or when the locus is divided into first, second, and third codon position (Table 1). The amino acid and third-position transversion partitions have PTP structure (Table 1). These partitions are not independent and the amino acid partition was selected because it includes only replacement changes, whereas the third-position transversion partition is a heterogeneous assemblage of silent and replacement changes. Consistent with the initial hypothesis, the amino acid characters support monophyly of (Conoplectus, Palimbolus) at the 70% level. The third-position transversion partition supports no nodes that are consistent with the initial hypothesis.

In conflict with the initial hypothesis, the amino acid characters support monophyly of (Cyparium, Habrocerus) at the 50% bootstrap consensus level. Nine synapomorphies unite (Cyparium, Habrocerus) and two steps are required to decay the node. Four of these changes are from isoleucine and three are changes to isoleucine. In contrast, only 44% of all changes throughout the amino acid tree include isoleucine. There are 14 isoleucines in the 48 parsimony-informative amino acid positions sequenced in Habrocerus and 7 in Cyparium. The mean number of isoleucines in the parsimony-informative positions for all taxa is 8.64 ± 0.55. These data suggest that there is an increase in the proportion of isoleucines in Habrocerus, possibly causing a long branch effect. Additional sequencing of taxa within the Habrocerinae is required. T-PTP tests of nonmonophyly suggest these data cannot reject the hypothesis that (Tachinus, Habrocerus) and (Cyparium, Oxytelus) are monophyletic (T-PTP = 0.17, 6 steps; T-PTP = 0.33, 5 steps). A neighbor-joining analysis with gamma-distributed rates (0.5, 1, 2) does not unite (Cyparium, Habrocerus) but in conflict with the initial hypothesis supports (Oxytelus, Oxyporus) at the 50% bootstrap. As discussed in the section on Total Evidence Analysis, additional characters not included in this analysis also support (Oxytelus, Cyparium) and (Tachinus, Habrocerus).
FIGURE 5. The 12S rRNA secondary structure model following Hickson et al. (1996). Stems are shaded, while equivocally alignable bases are denoted by hatched circles. Upper case characters denote positions that are invariant. Lower case characters denote the majority rule base at that position. If two bases are equally present in the majority the IUPAC code is employed. Insertions are shown with an open triangle. The six changes on the lineage to (Opresus, Edaphus) are shown with arrows and an open diamond is placed beside compensatory base changes in the (Opresus, Edaphus) lineage.
**Conditional Data Combination Analysis.**—For CDC we partitioned the data into process partitions, tested each partition for homogeneity, pooled independent combinable partitions, and constructed a tree. The head and nonhead morphological partitions, the domain II and domain III 12S rRNA partitions, and the cytochrome b amino acid partition have PTP structure (Table 1) and are not statistically incongruent \( (P < 0.01, 1,159\) steps). As described earlier, we did not include the cytochrome b third-position transversion partition.

The CDC procedure resolves more nodes than either TE or TC at the 95%, 70%, and 50% bootstrap consensus levels (Figs. 4, 7). However, only at the 95% bootstrap consensus (CDC95) are all nodes supported by the initial hypothesis.

Eight of the nine nodes supported in the CDC50 tree, including monophyly of Staph-

---

**Table 2.** Summary of nodes supported by each method that are consistent with the initial hypothesis shown in Fig. 2.

<table>
<thead>
<tr>
<th>Method</th>
<th>Phylogenetic analyses</th>
<th>No. nodes</th>
<th>A posteriori phylogenetic analyses</th>
<th>No. nodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total evidence</td>
<td>TE70</td>
<td>4</td>
<td>TE55</td>
<td>5</td>
</tr>
<tr>
<td>Taxonomic congruence</td>
<td>TC70</td>
<td>4</td>
<td>TC65</td>
<td>5</td>
</tr>
<tr>
<td>Conditional data combination</td>
<td>CDC95</td>
<td>5</td>
<td>CDC75</td>
<td>5</td>
</tr>
<tr>
<td>Compatible evidence</td>
<td>CE95-70</td>
<td>5</td>
<td>CE75-75</td>
<td>7</td>
</tr>
</tbody>
</table>

---

**Figure 6.** Taxonomic congruence trees constructed by combinable-component consensus of six partitions. The roman numerals in the circles represent the number of partitions supporting that node. (a) Constructed from the 95% bootstrap consensus. (b) Constructed from the 70% bootstrap consensus. (c) Constructed from the 50% bootstrap consensus.
ylinidae + Scydmaenidae + Silphidae, were present in, or consistent with, the original morphology based hypothesis (Figs. 2, 4). The ninth node, \((Oiceoptoma, Tachinus)\), was not included in the initial hypothesis. T-PTP tests of nonmonophyly suggest that we cannot reject monophyly of the initially hypothesized nodes of \((Oiceoptoma, Ontholestes, Achenomorphus, Oxyopus, Opresus, Euconnus, Dianous, Edaphus)\) (T-PTP = 0.98, 13 steps) or \((Tachinus, Habrocerus)\) (T-PTP = 0.17, 14 steps). Adult, larval, and pupal derived characters not included in this analysis (extraoral digestion by adults and larvae; four pairs of functional pupal spiracles) unite \(Oiceoptoma\) with the rest of the Staphylinine Group (here, \(Edaphus\) through \(Oiceoptoma\) in Fig. 2) and derived morphological characters of larvae including deciduous urogomphi, epipharynx with asymmetrical setose fields (Ashe and Newton, 1993; Newton, unpublished observations) and semilimuloid adult body shape unite \(Tachinus\) with the Tachyporine Group (here, \(Tachinus\) and \(Habrocerus\)).

**Compatible Evidence Analysis.**—For CE we divided the data into process partitions, tested each discrete partition for hierarchical structure, tested independent partitions having hierarchical structure for homogeneity, pooled combinable partitions, and constructed a tree. The CE constraint was based upon the 95\% CDC topology (CE95) (Fig. 7). The 70\% and 50\% CDC topologies contain a node that conflicts with the initial hypothesis. We then analyzed the characters in each partition separately under the constraint generated from the combined analysis, and constructed a combinable component consensus tree. No additional nodes were supported by a 95\% bootstrap of each partition (CE95-95), and the combinable-component consensus trees that summarize the topological features shared among the CE95-70 and CE95-50 bootstrap consensus trees are presented (Fig. 8).

CE95-70 includes three constrained and two unconstrained nodes that are congruent or consistent with the initial hypothesis (Table 2). The nonhead morphological and the 12S rRNA domain III loop partitions support monophyly of the ingroup excluding \(Neopelatops\), while the head partition supports monophyly of \((Amphichroum, Eusphalerum)\) (Figs. 4, 8a).

In CE95-50 the nonhead, domain II, domain III loop, and amino acid partitions support monophyly of the ingroup excluding \(Neopelatops\) (Figs. 4, 8b). The head partition supports monophyly of \((Amphichroum, Eusphalerum)\). In conflict with the initial hypothesis (Fig. 2), the amino acid partition supports monophyly of \((Cyparium, Habrocerus)\) (Figs. 4, 8b). As discussed earlier, one explanation for this result is that \(Habrocerus\) has an elevated number of isoleucines in the amino acid partition.

**A Posteriori Phylogenetic Analysis**

Maximally resolved trees consistent with the initial hypothesis.—Following Miyamoto et al. (1994) and Wheeler (1995), we prefer the method that supports the highest number of
nodes congruent with the initial hypothesis. CE75-75 supports the complete set of six nodes supported by the three other methods and \((\text{Dasycerus}, \text{Conoplectus}, \text{Palimbolus})\) (Fig. 9, Table 2). A posteriori analysis demonstrated that the result of the constrained bootstrap analysis is linked with the initial hypothesis. CE75-75 analysis produces five constrained and two unconstrained nodes, but the \((\text{Amphichroum, Eusphalerium})\) node supported in CE95-70 was not among these. We suggest that CE75-75 may be appropriate for future studies.

**Tree lengths.**—The total evidence data set contained all the morphological data and all the alignable 12S rRNA and cytochrome \(b\) sequence data. This consensus tree was 2,143 steps in length, one step shorter than the TE55/CDC75, TC65, and CE75-75 trees (Table 3). The one-step difference occurs because \((\text{Opresus, Euconnus})\) does not occur in any of the 2,143-step trees; it does occur in four of the eleven 2,144-step trees and is supported by a 59% bootstrap consensus of the combined data (Fig. 3b).

We tested whether the TE55/CDC75, TC65, and CE75-75 trees were significantly different in length. The compare-2 PTP test suggests that there is no statistical difference between the TE55/CDC75 topology and the TC65 and CE75-75 topologies (head morphological PTP = 0.30 and 0.36, nonhead morphological PTP = 0.82 and 0.91, domain II PTP = 0.82 and 0.91, domain III stems PTP = 0.80 and 0.75, domain III loops PTP = 0.30 and 0.24, cytochrome \(b\) amino acids PTP = 0.92 and 0.98). The Templeton test (1983) supports the conclusion that there are no significant differences among the three topologies in any partition. An alternative approach would be to calculate the tree length difference between the TE tree topology and each a posteriori tree topology for the total data set and for each of the homogenous partitions. We present the tree length differences for these tests (Table 3), but we suggest that these data should be treated with caution. The most parsimonious tree for each partition is not constant and the TE topology has less potential for character optimization within each partition (Table 3).

**Molecular Analysis**

**Saturation plots.**—To investigate further the informativeness of each locus, we plotted the number of substitutions against an HKY (Hasegawa et al., 1985) estimate of relative divergence. These plots demonstrate three main points.

First, these data from 12S rRNA and cytochrome \(b\) support the hypothesis that the ingroup had an explosive radiation. The pairwise divergence values between \textit{Necrophilus} and each ingroup taxon fall within the scatter of ingroup divergences (Figs. 10, 11). If all branches were short and of equal length it
Table 3. Increase in the number of steps from the minimum tree length

<table>
<thead>
<tr>
<th>Data set</th>
<th>Minimum tree (steps)</th>
<th>TE^a (additional steps)</th>
<th>TE55/CDC75^b (additional steps)</th>
<th>TC65 (additional steps)</th>
<th>CE75-75 (additional steps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total evidence</td>
<td>2,143</td>
<td>0/4</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Morphology</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Head</td>
<td>145</td>
<td>28/25</td>
<td>4</td>
<td>7</td>
<td>12</td>
</tr>
<tr>
<td>Nonhead</td>
<td>200</td>
<td>22/18</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12S rRNA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Domain II</td>
<td>72</td>
<td>17/16</td>
<td>5</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Domain III stems</td>
<td>283</td>
<td>33/29</td>
<td>14</td>
<td>13</td>
<td>19</td>
</tr>
<tr>
<td>Domain III loops</td>
<td>175</td>
<td>18/19</td>
<td>5</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>Cytochrome b</td>
<td>284</td>
<td>30/31</td>
<td>13</td>
<td>12</td>
<td>15</td>
</tr>
</tbody>
</table>

^a To facilitate comparison we calculate TE using all data (before the slash) and the six congruent partitions with hierarchical structure (after the slash).

^b These trees have the same topology.

Figure 9. Comparison of the a posteriori support for specific nodes. The tree representing the initial hypothesis is shown in Fig. 2. The bootstrap support or the number of independent partitions supporting a node is denoted by shading. Here the lowest bootstrap percentage, rounded up to the nearest 5%, that generated a topology consistent with the initial hypothesis is presented. Neopelatops is not included. H_o = initial hypothesis; TE55 = total evidence derived from a 55% bootstrap consensus; TC65 = taxonomic congruence, derived from 65% bootstrap support and constructed by combinable-component consensus; CDC75 = conditional data combination derived from a 75% bootstrap consensus; CE75-75 = compatible evidence derived from a 75% bootstrap consensus tree followed by a combinable-component 75% bootstrap consensus of each partition. Note that (Conoplectus, Palimbolus) is nested within (Dasycerus, Conoplectus, Palimbolus) in CE75-75.
Figure 10. Divergence between taxa plotted against an estimate of time since divergence for the 12S rRNA data. The 14 domain IV characters are not plotted because the three substitutions were not parsimony-informative. (a) Domain II. (b) Domain III stem transitions. (c) Domain III stem transversions. (d) Domain III loop transitions. (e) Domain III loop transversions.
Figure 11. Divergence between taxa plotted against an estimate of divergence time for the cytochrome \( b \) data. (a) First codon positions. (b) Second codon positions. (c) Third codon positions. (d) Third codon transitions. (e) Third codon transversions.
may be expected that the pairwise divergence values between the outgroup *Necrophilus* and all ingroup taxa would cluster in the top right of each plot.

Second, divergences in all partitions show some leveling. This suggests that each partition is saturated at least to some degree (Figs. 10, 11).

Third, proportionately more transversions than transitions appear to occur in the 12S rRNA domain III stems (Fig. 10c vs. 10b), loops (Fig. 10e vs. 10d) and cytochrome *b* third codon positions (Fig. 11e vs. 11d). In contrast to these data, Springer and Douzery (1996) noted more 12S rRNA transitions than transversions in mammals. This suggests that different evolutionary forces may be acting in insect and mammal mtDNA. A less likely explanation is that the difference may be due to our difficulty in accurately estimating divergence time. We employed the HKY model (Hasegawa et al., 1985) with gamma-distributed rates. In contrast, Springer and Douzery (1996) estimated divergence times based on the more complete mammal fossil record and on single-copy DNA hybridization molecular clocks.

Substitution biases.—This substitution bias in the 12S rRNA stems retains stem stability. Variable characters in 12S rRNA stems have significantly more C's and fewer U's than all positions combined (Table 4). There is also the trend to have more G's and fewer A's in variable positions than all positions combined.

Data presented here are consistent with the hypothesis of Ballard and Kreitman (1994) that changes to C or G are slightly deleterious in insect mtDNA. In the less constrained regions there are significantly more A's and U's or T's than C's and G's in variable positions than in all positions combined. Variable positions in 12S rRNA domain III loops have significantly more A's and U's and significantly fewer C's and G's than in all positions (Table 4). Further, there are significantly more A's and T's and fewer C's and G's in variable first- and third-codon positions than in all the positions.

Substitution heterogeneity also exists in the domain II 12S rRNA and cytochrome *b* second-position partitions. Additional data are required to test the generality of this result. Domain II variable positions show significantly more A's and C's and fewer U's. There is no significant difference in the numbers of G's. In the variable cytochrome *b* second positions there are significantly more A's and fewer G's, while the numbers of C's and T's do not differ significantly.

**DISCUSSION**

CE is a robust method for analyzing multiple data sets. CE sets a base of nodes obtained by CDC analysis and then investigates what further agreement may arise in a universe where these nodes are given. Thus, the CDC tree defines the minimum resolution of the CE topology. Congruence analyses are then employed to investigate the potential for additional phylogenetic resolution. Estabishing

### Table 4. Investigation of substitutional biases

<table>
<thead>
<tr>
<th>Data set</th>
<th>A</th>
<th>C</th>
<th>G</th>
<th>T or U</th>
</tr>
</thead>
<tbody>
<tr>
<td>12S rRNA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Domain II</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Domain III stems</td>
<td>−1.41</td>
<td></td>
<td>4.11*</td>
<td>−3.07*</td>
</tr>
<tr>
<td>Domain III loops</td>
<td>3.62*</td>
<td>−4.11*</td>
<td>−3.91*</td>
<td>4.04*</td>
</tr>
<tr>
<td>Cytochrome <em>b</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First position</td>
<td>4.07*</td>
<td>−4.04*</td>
<td>−4.02*</td>
<td>2.04*</td>
</tr>
<tr>
<td>Second position</td>
<td>4.11*</td>
<td>−1.38</td>
<td>−4.11*</td>
<td>−0.6</td>
</tr>
<tr>
<td>Third position</td>
<td>4.12*</td>
<td>−4.11*</td>
<td>−4.11*</td>
<td>4.12*</td>
</tr>
</tbody>
</table>

*Substitution biases were investigated by comparing the proportion of nucleotides that occurred in the variable positions compared to all positions using the Wilcoxon signed ranks test. The direction of change is represented by a positive (higher number in variable positions) or a negative value (higher number in all positions combined).

*Denotes significance at 0.05.
such constraints should preclude the problem raised by Barrett et al. (1991) that a consensus tree can contradict the tree inferred from pooled data. In this study CE75-75 supported more nodes that are consistent with the initial hypothesis than any other procedure. We suggest that these bootstrap proportions may be appropriate for future studies that include multiple data sets.

CDC, TC, and CE partition the data into evolutionary process partitions (Bull et al., 1993; Miyamoto and Fitch, 1995). In this study 11 process partitions were defined. We defined two morphological partitions because we expected that selection on feeding mode may influence the evolution of head characters. We defined four 12S RNA partitions: domain II, domain III stems, domain III loops, and domain IV. It is likely that the evolutionary forces acting on stems and loops differ. However, this simple partitioning may be inadequate. Hickson et al. (1996) suggested that patterns of conservation of and variability in the paired and unpaired regions make differential weighting in terms of stems and loops unsatisfactory. We investigated a total of five partitions in the cytochrome b locus. These partitions are not all independent. It is methodologically tractable under the criterion of parsimony to partition the cytochrome b locus into first, second and third codon positions, third-position transversions, and amino acids. However, all but the latter are heterogeneous mixtures of synonymous and nonsynonymous changes. In this study we employed the cytochrome b amino acid partition in the TC, CDC, and CE analyses.

There are at least three reasons why partitions may appear to evolve differently. The signal may be swamped by homoplasy, methods for investigating whether process partitions should be combined may be inadequate, or distinct and conflicting processes may be operating. Testing for hierarchical structure should test the first element while investigating the homogeneity of data partitions should test the second. We tested hierarchical structure using the PTP test and for incongruence length difference using the test of Farris et al. (1995). Cunningham (1997) found that the partition homogeneity test distinguished between cases where combining data generally improved phylogenetic accuracy and cases where accuracy of the combined data suffered relative to the individual partitions. However, further investigation of the robustness of alternative tests is required (Hillis, 1991; Huelsenbeck, 1991; Källersjö et al., 1992; Huelsenbeck and Bull, 1996; Lyons-Weiler et al., 1996).

Partitioning data may also be advantageous because it (1) assists the investigation of evolutionary forces, (2) facilitates critical exploration of the data, and (3) permits critical examination of methods for presenting data. However, not all process partitions are both self-defining and methodologically tractable. As a result, the validity of some partitions may be the subject of debate. In the extreme case, defining each character as a partition is not recommended (Miyamoto and Fitch, 1995).

We investigated substitution biases in each molecular partition by comparing the proportion of nucleotides that occurred in the variable positions to the proportion in all positions. Our results are consistent with the notion that there is a substitution bias to maintain secondary structure. Support for (Opresus, Edaphus) in the domain III stems partition that conflicts with the initial hypothesis is at least partially due to compensatory changes that maintain stem stability. In the less constrained regions of the variable 12S rRNA loops and the variable cytochrome b first and third codon positions there are significantly fewer C’s and G’s and more A’s and T’s. These data are consistent with the hypothesis of Ballard and Kreitman (1994) that changes to C or G are slightly deleterious at less constrained sites in the mitochondrial genome of Drosophila.

Plots of divergence against time are employed as a heuristic guideline for the onset of saturation. These plots clearly show that the outgroup-ingroup divergences cluster within the ingroup-ingroup comparisons (Figs. 10, 11). One explanation for this result is a massive radiation of staphylinid beetles, as has been hypothesized in previous morphological studies (Newton and Thayer, 1995). An alternative explanation is that an inappropriate outgroup was defined. Our attempts to address this latter question by sequencing three other outgroup beetle taxa (two species of Tribolium [Tenebrionidae] and one of Cicindela [Carabidae]) for both 12S rRNA and cytochrome b
(data not presented) failed to resolve the problem. The pairwise HKY distance estimates fell within the cluster of ingroup-ingroup divergences. Further, the outgroup rooted on a long ingroup branch, suggesting that the sequences were effectively randomized (Swoford et al. 1996a).

A partition is not considered to be saturated if the number of substitutions is linearly related to time since divergence. However, it is not clear how this should be calculated or when this is likely to be observed. Calculations of linear regressions and hyperbolic curves between transversions and transitions assume that the pairwise distance estimates are independent and thus researchers are left to infer saturation. In this study transitions appeared to be saturated by 8% divergence. In contrast to these data, Hackett (1996) showed that for tanagers in the genus Ramphocelus (Aves) third-position transitions at the ND2 locus exhibit signs of saturation by about 12% sequence divergence. Friedlander et al. (1994) suggested that below about 20–30% divergence the pairwise divergence should relate to the number of informative characters.

CE analysis of the constrained homogeneous partitions with hierarchical structure gives researchers flexibility to choose methods and/or models appropriate for each partition. In this study Habrocerus exhibited an elevated number of isoleucines. This may bias parsimony in the amino acid partition. Naylor and Brown (1997) noted that isoleucine and valine produced the poorest fit to a mitochondrial genealogy of 19 taxa whose interrelationships are widely accepted. A bias is introduced when the evolutionary process does not satisfy the assumptions crucial to a particular method. If the bias becomes sufficiently great, it may override the support for the true phylogeny and lead the researcher to an incorrect conclusion. In these cases the method is said to be positively misleading or inconsistent (Felsenstein, 1978). Future studies of staphylinoid systematics that include cytochrome b data should include additional taxa within the Tachyporine Group (Ashe and Newton, 1993) to break the long branch to Habrocerus.

CE facilitates the identification of conflicting biological signals between partitions. In this study, partitioning the morphological data into head and nonhead characters resulted in conflict over the phylogenetic affinities of (Conoplectus, Palimbolus) (Fig. 4). It is likely that this conflict is related to selection on preaceous mouthpart characters in the head partition. Decay indices, compare-2 tests, and constrained analyses support (Dasycerus (Conoplectus, Palimbolus)) and (Dianous (Edaphus (Opresus, Euconnus))). To the extent that they have been specified, characters used previously to justify a relationship between Pselaphinae (represented here by Conoplectus and Palimbolus) and a subset (here Edaphus, Dianous, Opresus, and Euconnus) of the Staphyline Group have been a few mouthpart structures and vague aspects of adult body form (especially coxal structure). We deem these likely to be subject to convergent selection in response to similarities in microhabitat and feeding habits (predation on microarthropods in forest litter), particularly in light of the findings of Newton and Thayer (1995) that morphological features of diverse body parts support monophyly of both the Omaline Group (here Dropephylla, Eusphalerum, Amphichroum, Conoplectus, and Palimbolus) and the Staphyline Group (here Oxyporus, Ontholestes, Achenmorpha, Edaphus, Dianous, Opresus, Euconnus, and perhaps Oiceoptoma). Future analyses should include additional independent morphological partitions such as external characters of other life stages or internal characters (e.g., larvae, Thayer, 1985; ovariole structure, Welch, 1993).

Despite the advantages of CE there are at least two potential problems that must be addressed before the technique can be routinely employed. Specifically, CE analysis may accept a tree that is not the most parsimonious, and it may have a biased estimator of nodal support.

CE may influence character compatibility if the constrained nodes are not present in the most parsimonious reconstruction of the homogeneous data partitions with hierarchical structure. This means the constrained patterns may not be congruent with those of global parsimony (Maddison et al., 1984). Under the procedure employed here, CDC generates the most parsimonious tree for combinable process partitions with hierarchical structure. CE then performs a second analysis in a universe where
the well supported nodes are constrained. The principle of this analysis is similar to the two-step procedure of Maddison et al. (1984) and is consistent with the tenets of meta-analysis. When all the data were included in this study, the TE topology was one step shorter than the TE55, TC65, and CE75-75 topologies. When the congruent partitions with hierarchical structure were analyzed independently, the TE55 and CDC75 topology was shorter than the CE75-75 topology in five of the six partitions. However, the compare-2 PTP and Templeton tests suggest that the increase in tree length that results from constraining nodes in CE75-75 is not greater than that attributable to chance. Constraining nodes for subsequent phylogenetic analysis is not a new idea. Moritz et al. (1992) used a step matrix to discount transitions at first and third codon positions, then used the shortest topology to constrain the basal branching order in an analysis that heavily weighted transversions over transitions.

CE may have a biased estimator of nodal support. In this study we compared the initial hypothesis to each of the four methods using the bootstrap proportions of 95%, 70% and 50%. We then calculated the minimum bootstrap percentage that remained compatible with the initial hypothesis. This procedure demonstrated that the results of the constrained CE bootstrap analysis were intimately linked with the unconstrained analysis. Three constrained and two unconstrained nodes supported by a CE95-70 bootstrap consensus support our initial hypothesis. To prevent conflict with this hypothesis, the a posteriori analysis indicated that decreasing the unconstrained bootstrap consensus by 20% to CE-75 necessitated a 5% increase in the constrained bootstrap analysis to CE75-75.

Further research, possibly through simulation, is required to determine the utility of CE75-75. However, it may be predicted that CE may perform well when there are correlated characters in one part of the tree but not in another or when one or more partitions have missing data. In this study we have shown that CE constrains a set of well-supported CDC nodes. If we are willing to accept this constrained topology and the partitions employed to generate this topology resolve further nodes, we suggest that this represents strong evidence supporting these additional nodes. Methodologically, CE appears to have advantages over the other procedures discussed. First, it sets a base of nodes obtained by CDC analysis and then investigates what further agreement may arise in a universe where these nodes are given. In this study CE75-75 resolves more nodes than any other procedure. Second, CE investigates disagreements between partitions in a universe where these nodes are given. In this study, this facilitated investigation of conflicts between partitions. Thus, CE enables researchers to both generate a well-corroborated tree and investigate character conflicts.

ACKNOWLEDGMENTS

We thank the Field Museum Evolution Discussion Group, Kirrie Ballard, Dan Faith, Anna Graybeal, Larry Hedges, David Maddison, Janet Voight, Peter Wagner, and three anonymous reviewers for their comments and discussion. We thank David Swoford for permission to use PAUP*. This work is based upon molecular data collected in the Field Museum’s Biochemistry Laboratories, renovated with support from the National Science Foundation (grant STI-921446) and operated with support by an endowment from the Pritzker Foundation. DNA sequence data were collected on an Automated DNA sequencer purchased with funds granted to the Field Museum by the National Science Foundation (Grant BIR9419732). Funds for the biochemical work were provided in part by the Field Museum as part of J.W.O.B’s start-up funds. Fieldwork by A.F.N. and M.K.T. was supported by the Marshall Field Fund of the Field Museum.

REFERENCES


Faith, D. P., and J. W. H. Trueeman. 1996. When the topology-dependent permutation test (T-PTP) for monophyly returns significant support for monophyly, should that be equated with (a) rejecting a null hypothesis of nonmonophyly, (b) rejecting a null hypothesis of “no structure,” (c) failing to falsify a hypothesis of monophyly, or (d) none of the above? Syst. Biol. 45:580–586.


Received 18 February 1997; accepted 9 September 1997. Associate Editor: D. Cannatella

**APPENDIX I**

**SPECIMENS EXAMINED**

All specimens from which morphological and molecular data (listed next) were gathered had been collected and identified by A.F.N. and M.K.T., except as noted. The notation (70%) indicates specimens collected into acid alcohol (70% ethanol with 5% by volume of glacial acetic acid added), then stored in 70% ethanol before DNA extraction; all others were collected and stored in 95% ethanol. A pinned or pointed specimen from each series is marked as a voucher and deposited in the Field Museum Nercrophillus hydrophiloides Guérin-Méneville-USA: CALIFORNIA: Monterey Co., Los Padres NF, Bottchers Gap, 625 m, Quercus, etc., woodland; 19.iii.1995, in roiling and muddy gilled mushrooms; Neopelatops sp.1-AUSTRALIA: TASMANIA: Liffey Forest Reserve, near picnic area, 560 m, site 918, Eucalyptus obliqua forest with rainforest understorey; 31.i.1993, pyrethrin-fogging large old Eucalyptus log chunk w/tan slime mold about to fruit; Drocephylly cacti (Schwarz) (70%)—USA: ARIZONA: Cochise Co., Chiricahua Mts, East Turkey Creek, 1.1 km SE of on US FS road 42, 1,950 m, Quercus-Pinus edulis-Juniperus woodland; 9.iv. 1995, on white flowers Canthos fenderi 0930hr; Eusphaerum sp. —USA: CALIFORNIA: Monterey Co., Los Padres N.F., Skinner Ridge Trail nr. Bottchers Gap, 650 m, Quercus-Arbutus menziesii woodland; 19.iii.1995, on light blue Canthos flowers; Amphichroum sp. —USA: CALIFORNIA: same as preceding; Megarthrus americanus Sachse—USA: ILLINOIS: Union Co., Shawnee N.F., Little Grand Canyon, 120–200 m site 967, 15.x.1995, mixed hardwood forest; in rotting gilled mushrooms, A. Newton, M. Thayer, O. Helmy; Proteinus sp.—USA: ILLINOIS, same as preceding; Dasycerus angulicollis Horn (70%)—USA: CALIFORNIA: Monterey Co., Los Padres N.F, Nacimiento-Fergusson Rd. at pass (11.2 km from California Hwy. 1), 840 m, site 953, Quercus-Arbutus menziesii-Pinus coulteri woodland; 18.iii.1995, FMHD# 95–47, berlese, forest leaf and log litter; Conoplectus canaliculatus (LeConte)—USA: ILLINOIS: Union Co., Shawnee N.F., Little Grand Canyon, 120–200 m, site 967, 15.x.1995, mixed hardwood forest; berlese, log and leaf litter FMHD# 95–82, M. Thayer, Palimbolus victoriae (King) (70%)—AUSTRALIA: TASMANIA: Southwest N.P., Gordon R. Rd, 10.1 km SE Strathgordon, 325 m, site 903, open Eucalyptus forest with Acacia, Eucryphiad, etc. understorey; 24.1.1995, FMHD# 93–52, berlese, leaf and log litter; Edaphus sp. (americanus Puthz n titius Motschulsky)—USA: ILLINOIS: Union Co., Shawnee N.F., F.R. 236, SE of McGee Hill, 175 m, site 966, 14.x.1995, mixed hardwood forest in south facing ravine; FMHD# 95–80, berlese, old litter in dry stream bed, A. Newton, M. Thayer, O. Helmy; and USA: ILLINOIS: FMHD# 95–82, as already described; Dianous nitidius LeConte (70%)—USA: NEW MEXICO: Lincoln Co., Eagle Ck. at Carlton Canyon (W of Alto), 2,390 m, site 944, Pseudotsuga menziesii-Pinus ponderosa forest; 3.iii.1995, FMHD# 95–2, berlese, wet debris, forest stream; Opresus sp. (det. W. Suter)—USA: WISCONSIN: Kenosha Co., Somers, Holmes Forest; 26.ix.1995 FMHD# 95–84, berlese oak stump buttress tree hole, W. Suter; Euconnus sp.—USA: ILLINOIS: FMHD# 95–82, as already described; Oxytus stygius Say—USA: ILLINOIS: Union Co., Shawnee N.F., Little Grand Canyon, 120–200 m site 967, 15.x.1995, mixed hardwood forest; in yellow gilled mushrooms, A. Newton, M. Thayer, O. Helmy; Achenomorphus corticinus (Gravenhorst)—USA: ILLINOIS: FMHD# 95–82, as already described; Ontholestes cingulatus Erichson—USA: ILLINOIS: Cook Co., Western Springs, Bemis Woods North, site 943, oak/mixed hardwood forest; 2-16.vii.1994, FMHD# 94–6, carrion trap (squid); Oiceoptoma nveadorance (Forster)—USA: ILLINOIS same as preceding; Tachinus luridus (Fabricius)—USA: ILLINOIS: same as preceding; Habrocerus capillarcornis (Gravenhorst)—USA: ILLINOIS: Cook Co., Western Springs, Bemis Woods North, site 943, oak/mixed hardwood forest; 2-16.vii.1994, FMHD# 94–5, berl., rotting logs; Cyparium concolor (Fabricius)—USA: ILLINOIS: FMHD# 95–82, as described earlier; Oxytelus convergens LeConte—USA: ILLINOIS: FMHD# 94–6, as described earlier.
Appendix 2

Morphological characters included in this study (see Newton and Thayer [1995] for details). Characters 11–50 form the head partition; characters 1–10 and 51–112 the nonhead partition. “X” indicates characters excluded by Newton and Thayer (1995) and in this study and “A” denotes characters excluded in this study only. A “?” indicates unknown and a “-” denotes inapplicable.

<table>
<thead>
<tr>
<th>Character</th>
<th>X</th>
<th>A</th>
<th>X</th>
<th>A</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Necrophilus</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Neopelatesops</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Drupesphylia</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Buphalecula</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Amphicrinia</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Megarchirus</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Proteinus</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Dasyurus</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Congepactus</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pallidobulus</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Alcyopus</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Dianus</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Opesmus</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Bucconus</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Oppgurus</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Articulars</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Cinecephala</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Tschimius</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Nebrocerus</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Oxyphorus</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Oxytulus</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
APPENDIX 3

The secondary structure alignment of domains II-IV of staphylinoid 12S rRNA. The helix (stem) numbers are shown above the alignment and the bases involved in pairing are shaded for each taxon. Open boxes enclose positions that include gaps but were included in the analysis. Hatched boxes enclose positions that include gaps and have been removed from the analysis.
## Appendix 4

The alignment of the staphylinoid cytochrome \(b\) locus.

<table>
<thead>
<tr>
<th>Taxon</th>
<th>Alignment of the staphylinoid cytochrome (b) locus</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Macrophilus</em></td>
<td>TGGAGAACCTTTCATTGGAGGAGCTTATCTAGCTATCTCTTGCGACTTACAGGTGAATTCTCCTACGTGTTGAGG</td>
</tr>
<tr>
<td><em>Neopeltops</em></td>
<td>TGGAGAACCTTTCATTGGAGGAGCTTATCTAGCTATCTCTTGCGACTTACAGGTGAATTCTCCTACGTGTTGAGG</td>
</tr>
<tr>
<td><em>Dropephylla</em></td>
<td>TGGAGAACCTTTCATTGGAGGAGCTTATCTAGCTATCTCTTGCGACTTACAGGTGAATTCTCCTACGTGTTGAGG</td>
</tr>
<tr>
<td><em>Busphalerum</em></td>
<td>TGGAGAACCTTTCATTGGAGGAGCTTATCTAGCTATCTCTTGCGACTTACAGGTGAATTCTCCTACGTGTTGAGG</td>
</tr>
<tr>
<td><em>Amphichroa</em></td>
<td>TGGAGAACCTTTCATTGGAGGAGCTTATCTAGCTATCTCTTGCGACTTACAGGTGAATTCTCCTACGTGTTGAGG</td>
</tr>
<tr>
<td><em>Lanatosus</em></td>
<td>TGGAGAACCTTTCATTGGAGGAGCTTATCTAGCTATCTCTTGCGACTTACAGGTGAATTCTCCTACGTGTTGAGG</td>
</tr>
<tr>
<td><em>Proteinus</em></td>
<td>TGGAGAACCTTTCATTGGAGGAGCTTATCTAGCTATCTCTTGCGACTTACAGGTGAATTCTCCTACGTGTTGAGG</td>
</tr>
<tr>
<td><em>Dasyxenus</em></td>
<td>TGGAGAACCTTTCATTGGAGGAGCTTATCTAGCTATCTCTTGCGACTTACAGGTGAATTCTCCTACGTGTTGAGG</td>
</tr>
<tr>
<td><em>Conoplectus</em></td>
<td>TGGAGAACCTTTCATTGGAGGAGCTTATCTAGCTATCTCTTGCGACTTACAGGTGAATTCTCCTACGTGTTGAGG</td>
</tr>
<tr>
<td><em>Palimboelus</em></td>
<td>TGGAGAACCTTTCATTGGAGGAGCTTATCTAGCTATCTCTTGCGACTTACAGGTGAATTCTCCTACGTGTTGAGG</td>
</tr>
<tr>
<td><em>Staphylinus</em></td>
<td>TGGAGAACCTTTCATTGGAGGAGCTTATCTAGCTATCTCTTGCGACTTACAGGTGAATTCTCCTACGTGTTGAGG</td>
</tr>
<tr>
<td><em>Dianurus</em></td>
<td>TGGAGAACCTTTCATTGGAGGAGCTTATCTAGCTATCTCTTGCGACTTACAGGTGAATTCTCCTACGTGTTGAGG</td>
</tr>
<tr>
<td><em>Opresus</em></td>
<td>TGGAGAACCTTTCATTGGAGGAGCTTATCTAGCTATCTCTTGCGACTTACAGGTGAATTCTCCTACGTGTTGAGG</td>
</tr>
<tr>
<td><em>Eucoraxus</em></td>
<td>TGGAGAACCTTTCATTGGAGGAGCTTATCTAGCTATCTCTTGCGACTTACAGGTGAATTCTCCTACGTGTTGAGG</td>
</tr>
<tr>
<td><em>Oxyporus</em></td>
<td>TGGAGAACCTTTCATTGGAGGAGCTTATCTAGCTATCTCTTGCGACTTACAGGTGAATTCTCCTACGTGTTGAGG</td>
</tr>
<tr>
<td><em>Acheneramphus</em></td>
<td>TGGAGAACCTTTCATTGGAGGAGCTTATCTAGCTATCTCTTGCGACTTACAGGTGAATTCTCCTACGTGTTGAGG</td>
</tr>
<tr>
<td><em>Oxyporus</em></td>
<td>TGGAGAACCTTTCATTGGAGGAGCTTATCTAGCTATCTCTTGCGACTTACAGGTGAATTCTCCTACGTGTTGAGG</td>
</tr>
<tr>
<td><em>Oxyporus</em></td>
<td>TGGAGAACCTTTCATTGGAGGAGCTTATCTAGCTATCTCTTGCGACTTACAGGTGAATTCTCCTACGTGTTGAGG</td>
</tr>
</tbody>
</table>

---

**Note:** The table above shows the alignment of the staphylinoid cytochrome \(b\) locus for various species. Each row represents a different taxon, and the alignment is represented as a sequence of nucleotides.
null
Macrophilus
Neolepatoce
Dropephylla
Bupholerum
Amphichromus
Megarthrus
Proteinus
Dasycerus
Conoplectus
Pallimulus
Eulophus
Dianthus
Oryxus
Euxanthis
Oxytus
Acanthomorphus
Ortholestes
Cloeotoma
Tachinus
Habrocerus
Cyprilus
Oxytelus

ATAGAATCCATCAATCGAATTCAAGAAAGCTTTTATTGGAATACGATATCAATCCCAAAGATTGA
ATTAGAATCCATCAATCGAATTCAAGAAAGCTTTTATTGGAATACGATATCAATCCCAAAGATTGA
TTTAGAATCCATCAATCGAATTCAAGAAAGCTTTTATTGGAATACGATATCAATCCCAAAGATTGA
TTTAGAATCCATCAATCGAATTCAAGAAAGCTTTTATTGGAATACGATATCAATCCCAAAGATTGA
TTTAGAATCCATCAATCGAATTCAAGAAAGCTTTTATTGGAATACGATATCAATCCCAAAGATTGA
TTTAGAATCCATCAATCGAATTCAAGAAAGCTTTTATTGGAATACGATATCAATCCCAAAGATTGA
TTTAGAATCCATCAATCGAATTCAAGAAAGCTTTTATTGGAATACGATATCAATCCCAAAGATTGA
TTTAGAATCCATCAATCGAATTCAAGAAAGCTTTTATTGGAATACGATATCAATCCCAAAGATTGA
TTTAGAATCCATCAATCGAATTCAAGAAAGCTTTTATTGGAATACGATATCAATCCCAAAGATTGA
TTTAGAATCCATCAATCGAATTCAAGAAAGCTTTTATTGGAATACGATATCAATCCCAAAGATTGA
TTTAGAATCCATCAATCGAATTCAAGAAAGCTTTTATTGGAATACGATATCAATCCCAAAGATTGA
TTTAGAATCCATCAATCGAATTCAAGAAAGCTTTTATTGGAATACGATATCAATCCCAAAGATTGA
TTTAGAATCCATCAATCGAATTCAAGAAAGCTTTTATTGGAATACGATATCAATCCCAAAGATTGA
TTTAGAATCCATCAATCGAATTCAAGAAAGCTTTTATTGGAATACGATATCAATCCCAAAGATTGA
TTTAGAATCCATCAATCGAATTCAAGAAAGCTTTTATTGGAATACGATATCAATCCCAAAGATTGA
TTTAGAATCCATCAATCGAATTCAAGAAAGCTTTTATTGGAATACGATATCAATCCCAAAGATTGA
TTTAGAATCCATCAATCGAATTCAAGAAAGCTTTTATTGGAATACGATATCAATCCCAAAGATTGA
TTTAGAATCCATCAATCGAATTCAAGAAAGCTTTTATTGGAATACGATATCAATCCCAAAGATTGA
TTTAGAATCCATCAATCGAATTCAAGAAAGCTTTTATTGGAATACGATATCAATCCCAAAGATTGA
TTTAGAATCCATCAATCGAATTCAAGAAAGCTTTTATTGGAATACGATATCAATCCCAAAGATTGA
TTTAGAATCCATCAATCGAATTCAAGAAAGCTTTTATTGGAATACGATATCAATCCCAAAGATTGA
TTTAGAATCCATCAATCGAATTCAAGAAAGCTTTTATTGGAATACGATATCAATCCCAAAGATTGA
TTTAGAATCCATCAATCGAATTCAAGAAAGCTTTTATTGGAATACGATATCAATCCCAAAGATTGA
TTTAGAATCCATCAATCGAATTCAAGAAAGCTTTTATTGGAATACGATATCAATCCCAAAGATTGA

Downloaded from https://academic.oup.com/sysbio/article-abstract/47/3/367/1703689 by guest on 06 December 2018