Exploring Data Interaction and Nucleotide Alignment in a Multiple Gene Analysis of *Ips* (Coleoptera: Scolytinae)

ANTHONY I. COGNATO[1] AND ALFRIED P. VOGLER

Department of Entomology, The Natural History Museum, London SW7 5BD, UK, and
Department of Biology, Imperial College at Silwood Park, Ascot, Berkshire, SL5 7PY, UK

Abstract.—The possibility of gene tree incongruence in a species-level phylogenetic analysis of the genus *Ips* (Coleoptera: Scolytidae) was investigated based on mitochondrial 16S rRNA (16S) and nuclear elongation factor-1α (EF-1α) sequences, and existing cytochrome oxidase I (COI) and non-molecular data sets. Separate cladistic analyses of the data partitions resulted in partially discordant most-parsimonious trees but revealed only low conflict of the phylogenetic signal. Interactions among data partitions, which differed in the extent of sequence divergence (COI > 16S > EF-1α), base composition, and homoplasy, revealed that much of the branch support emerges only in the simultaneous analysis, particularly for deeper nodes in the tree, which are almost entirely supported through " hidden support" (sensu Gatesy et al., Cladistics 15:271-313, 1999). Apparent incongruence between data partitions is in part due to suboptimal alignments and bias of character transformations, but little evidence supports invoking incongruent phylogenetic histories of genetic loci. There is also no justification for eliminating or downweighting gene partitions on the basis of their apparent homoplasy or incongruence with other partitions, because the signal emerges only in the interaction of all data. In comparison with traditional taxonomy, the *pini*, *plastographus*, and *perturbatus* groups are polyphyletic, whereas the *grandicollis* group is monophyletic except for inclusion of the (monophyletic) *calligraphus* group. The *latidens* group and some European species are distantly related and closer to other genera within Ipini. Our robust cladogram was used to revise the classification of *Ips*. We provide new diagnoses for *Ips* and four subgeneric taxa. [Bark beetle systematics; DNA sequence alignment; elongation factor; mitochondrial 16S rRNA; phylogenetic congruence; phylogenetic information content; saturation.]

Molecular markers are particularly useful for investigating relationships between closely related species but they also may result in misleading phylogenetic conclusions due to historical phenomena prevalent at this taxonomic level. Deep coalescence, gene duplication, and horizontal gene transfer may result in different gene histories in the various portions of the genome (e.g., Pamilo and Nei, 1988; Doyle, 1992, 1997; Sota and Vogler, 2001), and thus the use of multiple loci is deemed necessary to provide a reliable estimate of the species tree (Doyle, 1997; Maddison, 1997). Some (e.g., Miyamoto and Fitch, 1995) have advocated keeping data compartmentalized as a prerequisite to recognizing the biological causes of incongruences, given that corroboration in separate analysis provides strongest evidence. Others (e.g., Bull et al., 1993) propose combining data on the condition that there is no significant conflict between partitions. Recently developed procedures such as gene tree parsimony accept incongruence as the manifestation of biological processes and account for these events in species tree reconstruction (Maddison, 1997; Slowinski and Page, 1999).

However, none of these approaches may be appropriate in a given situation because they ignore the role of data interaction and the reciprocal effects of character transformations in phylogenetic reconstruction from multiple gene partitions. Establishing gene tree-species tree incongruence, therefore, requires a proper test of conflict in phylogenetic signal between gene partitions at the level of character variation. Apparent character incongruence may have many causes, however, ranging from simple lack of sufficient character differences (Brower and DeSalle, 1994; Mardulyn and Whitfield, 1999), to convergent bias in sequence evolution and excessive homoplasy of character change (Felsenstein, 1978; Brower et al., 1996; Baker and DeSalle, 1997), to the inappropriate assignment of putative homologies in length-variable regions (Giribet and Wheeler, 1999). Convergent bias in sequence evolution as well as misalignment may produce an incorrect phylogenetic signal that conceals weaker signal representing the true phylogeny. The literature proposes similar

---

[1]Address for correspondence: Department of Entomology, Texas A&M University, College Station, Texas 77843-2475 , USA; E-mail: a-cognato@tamu.edu

procedures for dealing with both problems; the removal or downweighting of the affected nucleotides. Generally, removal of alignment-variable regions from data matrices is recommended before tree reconstruction (Swofford et al., 1996). Similarly, removal or downweighting of "saturated" character variation is commonplace, for example, in the analysis of third-codon positions in protein-coding regions or the hypervariable regions in rRNA genes (the latter also frequently being alignment ambiguous), reasoning that saturated data merely represent noise to be separated from phylogenetically informative data (references in Wenzel and Siddall, 1999). Apart from the questionable notion that homoplastic data do not contain phylogenetic signal (see Gatesy et al., 1999; Wenzel and Siddall, 1999), downweighting and removal of data can rarely be done according to objective criteria, given the difficulty of quantifying the power of "noise" to mislead phylogeny estimation.

However, with the availability of larger data sets from diverse parts of the genome, we can now more rigorously assess the effects of potentially misleading data. Data partitions with their respective dynamics of sequence evolution can be analyzed in the context of other partitions and provide the possibility of detecting the common signal that may be hidden by biases in character transitions (e.g., Olmstead and Sweere, 1994). We selected three loci with presumably different dynamics of character variation for such analysis, including a protein-coding nuclear gene (elongation factor-1 alpha; EF-1α), a mitochondrial protein-coding gene (cytochrome oxidase subunit I; COI), and a rRNA gene (mitochondrial 16S rRNA; 16S) affected by length variation. These loci reflect the mitochondrial-nuclear division of phylogenetic markers, for which the different modes of inheritance are thought to increase the chance of gene tree incongruence in closely related taxa. The loci also differ in base composition because mitochondrial genes of insects have a much greater AT content than the nuclear coding regions (Caccone et al., 1988). Investigation of interactions between these markers in a combined analysis can be used to assess the extent of incongruence and its causes and to separate true differences in gene phylogeny from the spurious incongruence resulting from specific properties of each data partition and procedures of analysis.

We apply this analysis to the determination of species-level relationships in the bark beetle genus *Ips* (Coleoptera: Scolytinae). Some species in this group are among the most destructive pests of conifer forests worldwide (Furniss and Carolin, 1977; S. L. Wood, 1982; Werner, 1988; Knizek and Zahradnik, 1998) and therefore are of considerable economic importance. The species also represent an interesting model system for evolutionary biology; they exhibit complex conspecific and heterospecific semiochemical communication systems that lend themselves to studies of chemical ecology, community ecology, predator-prey interactions, and the role of mate recognition in speciation (e.g., D. L. Wood, 1982; Raffa and Klepzig, 1989; Seybold et al., 1995; Cognato et al., 1997). These studies require accurate delimitation of species and knowledge of species relationships. *Ips* classification has frequently been used to infer generalities in behavior or biology of related species (Vité et al., 1972; Lanier and Wood, 1975; Lewis and Cane, 1990; Seybold, 1992; Seybold et al., 1995). However, the classification of *Ips* based on morphological and cytological similarities as well as intra- and interspecific breeding experiments is largely phenetic (reviewed by Cognato and Sperling, 2000). Several recent attempts to establish a phylogeny of *Ips* have left many questions unanswered because of little phylogenetic signal (Cognato, 2000) or limited taxonomic scope (Cane et al., 1990; Cognato et al., 1995; Page et al., 1997; Stauffer et al., 1997). A phylogenetic reconstruction that included 39 of 43 known species of *Ips* (Cognato and Sperling, 2000) and was based on COI sequences found most species groups under S. L. Wood's (1982) classification are not monophyletic. In addition, several species of *Ips* grouped with closely related outgroups such as *Orthotomicus* and *Pityokteines*, a fact also supported by a morphological data set and requiring the establishment of *Pseudips* for the former *concinnus* species group (Cognato, 2000).

Although these recent findings suggest the need for major changes in the classification of *Ips*, the COI data constitute but a single gene tree as the basis for these revisions. Here we describe the sequencing of portions of mitochondrial 16S and nuclear EF-1α genes to augment the existing data.

We detected apparent incongruence among gene trees and tested whether this was, in fact, evidence for different historical signal contained in the various partitions. We tested other potential causes of incongruence, particularly the effect of inappropriate homology assignment in sequence alignment and bias of changes in the highly variable third-codon positions of protein-coding regions. The analysis revealed that the distribution of conflicting signal within and between partitions is complex, but appropriate alignment of length variable regions of rRNA and intron sequences demonstrated a widely supported phylogenetic signal. On the basis of these results, we provide a sufficiently robust phylogeny for a revised classification of *Ips*.

MATERIALS AND METHODS

DNA Sequences and Protocols

A total of 44 species of *Ips* and closely related outgroups *(Orthotomicus caelatus, O. erosus, Pityogenes carinulatus, P. hopkinsi, Pityokteines sparsus, Pseudips concinnus*, and *P. mexicanus)* were included in this study. Mitochondrial COI sequences (769 characters) and a nonmolecular data set (14 morphological and 2 behavioral characters) were from Cognato and Sperling (2000) and Cognato (2000), respectively. The newly generated 16S and EF-1α sequences were amplified from the same DNA extractions used in the previous study (Cognato and Sperling 2000). For two additional species included in the earlier study, *I. hauseri* and *O. laricis*, PCR amplification of these genes failed, and the species were excluded from the current analysis.

A newly designed beetle-specific primer pair (LR-J-12961, 5′-TTT AAT CCA ACA TCG AGG, and LR-N-13398, 5′-CGC CTG TTT AAC AAA AAC AT) (Simon et al., 1994) was used to amplify partial 16S. Substantial length variation in the 16S rRNA gene was unexpectedly observed, with indels accounting for approximately one-third of the positions in the resulting data matrices. This was in contrast to species-level studies of other Coleoptera in a range of families, including Chrysomelidae, Cicindelidae, and Dytiscidae, in which only a few or no indels were observed at this taxonomic level (Funk et al., 1995; Mardulyn et al., 1997; Vogler and Welsh, 1997; Funk, 1999; Ribera et al., 2001).

A region of approximately 950 base pairs (bp) of the EF-1α gene, corresponding to 2103–3149 bp of the *Drosophila melanogaster* F1 copy (Hovemann et al., 1988), was amplified by using a combination of primers. Primers M3, M44-1, M51.9, rcM44.9, rcM51-1, and rcM53-2 were reproduced from Cho et al. (1995) and subsequently customized for *Ips*. The sense primers are EF-1α-S-2255 (5′-CCT GGG TAT TGG ACA AAC T), EF-1α-S-2515 (5′-GGA GTG AAA CAA CTT ATC GTT GG), and EF-1α-S-2360 (5′-GGT ACT GGT GAA TTT GAA G); the anti-sense primers are EF-1α-A-2692 (5′-CCA ACA TGT TGT CTC CAT GCC A), EF-1α-A-3015 (5′-ACA TTG TCA CCA GGG ACA GC), and EF-1α-A-3061 (5′-GGG TGG TTG AGC ACA ATG). PCR amplification of EF-1α was unsuccessful for *I. longifolia, I. mannsfeldi, I. nobilis, I. schmutzenhoferi, I. stebbingi, O. laricis*, and *Pityogenes calcaratus*. The characters in this partition were scored as missing for those species in all phylogenetic analyses. Successful PCR yielded amplified fragments between 680 and 706 bp long. In all *Ips* and outgroup species, the sequence includes a putative intron as much as 81 bp long, at position 2,269 of the corresponding *D. melanogaster* sequence. This sequence is identified as an intron sequence because of the lack of similarity of the translated amino acid sequence with that of other insect species, the presence of multiple stop codons, and high similarity to 5′ and 3′ splice site consensus sequences. The intron does not correspond to the position of known introns found in other insects, but it does occur within bp of intron 3 in *Apis* F2 (Danforth and Ji, 1998). After position 2,517, another intron is suspected for *Pseudips concinnus* and *P. mexicanus* because 54 nucleotides could not be aligned with the other *Ips* or outgroup species and similarity to a 3′ splice site. The 5′ end of the target EF-1α sequencing was unsuccessful for *I. amitinus, P. concinnus*, and *P. mexicanus*, sequences that start at nucleotide positions 2,598, 2,570, and 2,570, respectively. Thus the actual size of the suspected intron for *P. concinnus* and *P. mexicanus* is unknown. For *Ips acuminatus*, nucleotide heterozygosity was apparent in DNA sequence electropherograms in a single case at position 2,496. Amplification resulted in only a single PCR product, suggesting that the customized primers annealed to a single ortholog of EF-1α, even though multiple loci of the gene have been described for some insect species (Hovemann et al., 1988; Danforth and Ji, 1998).

For both genes, PCR cocktail contents and temperature profiles for thermocycling were as described in Cognato and Sperling (2000). An annealing temperature of 55°C was used for the EF-1α customized primers. PCR products were separated from primers and unincorporated nucleotides with a QIAquick PCR Purification Kit (QIAGEN Inc., Santa Clara, CA), following the manufacturer's instructions, and were directly sequenced with an ABI 377 sequencer. DNA sequences were determined in both directions. All 16S and EF-1α sequences were deposited in GenBank (16S: AF397470-AF397512; EF-1α: AF397612-AF397649).

Sequence Alignment

Sequence alignments were performed by using the optimization alignment procedure of Wheeler (1996), as implemented in POY Software (Gladstein and Wheeler, 1997). This procedure minimizes the number of insertions/deletions and nucleotide changes simultaneously on a tree by using parsimony. The analysis was performed on the combined data set including the length-invariable COI and the morphological data set, keeping each data partition separate in the POY analysis. To prevent spurious alignments, the EF-1α and the 16S partitions were further subdivided to separate the conserved and variable regions before the searches. The EF-1α data were split into two exon regions (at 1–545 and 628–766 bp) and the intron region (at 546–627 bp); the 16S was separated into three conserved regions (1–52, 135–170, and 307–481 bp) and two variable regions (53–136 and 171–306 bp). Heuristic searches for the most-parsimonious tree alignment were conducted with 50 random addition replicates (*-random 50* command) using POY version 2.6 on a Dell personal computer.

The influence of different alignment parameters on phylogenetic inferences was tested by changing the ratio of gap cost to change cost in POY. We assessed the ratio of 2:1 which has been shown to produce optimal alignments in other analyses (Giribet and Wheeler, 1999; Phillips et al., 2000), and the effect of variation in alignment parameters for ratios of 1:1 and 3:1. An output of the aligned matrix was produced by POY with the *-impliedalignment* option, and was converted to NEXUS format for subsequent analysis of tree length and support in PAUP*. The cost of the resulting character matrix was determined in PAUP* under parsimony optimization on the tree(s) generated by POY. The cost for a given tree alignment estimated by POY was usually found to be the same or very similar to that calculated with PAUP* (using the *Tree Scores* option and employing a step matrix specifying the same gap cost as in the POY searches and *"gaps as a fifth character state"*). However, using the heuristic search and TBR branch swapping in PAUP* revealed slightly shorter trees than those from the POY alignments in several cases; if a shorter tree could be found with PAUP* based on these alignments, we preferred that tree because the aligned matrices could also be used to calculate standard support values and branch lengths. By using the matrix obtained with *impliedalignment* in POY, we accepted that the aligned matrix reflects the correspondences of bases during alignment of the tree. Although it remains to be proven that these matrices are equivalent to character optimization during tree alignment, the potential error appears to be small, at least for the data sets considered here, with tree length and topology differing only slightly between both optimization procedures; accordingly, we used these matrices interchangeably with the tree alignment. Visual inspection of the implied alignment revealed rather "gappy" matrices, with homology assignments that appeared biologically unrealistic in some cases. However, because the true phylogeny of indels is not known, we accept that the alignment that produces the overall shortest tree alignment (from simultaneous analysis [SA] of all data) is the best reflection of homologies of bases. This alignment reflects the correspondences of nucleotide positions in a way that best accommodates the principle of congruence. In our alignments, "misplaced" indels often represent autapomorphies and thus would have a negligible effect on tree topology.

Multiple sequence alignments were also conducted with CLUSTAL W version 1.7 (Thompson et al., 1994) to compare alignments based on phenetic procedures. We used the default settings (slow/accurate, gap opening penalty: 15.00, gap extension penalty: 6.66, delay divergent sequences: 40%, DNA transitions weight: 0.50, DNA weight matrix: IUB, negative matrix: OFF). EF-1α coding sequences were unambiguously aligned because of

the conservation in amino acid sequence. The above alignments are available at the Society of Systematic Biologists website (www.systbiol.org).

Phylogenetic Analysis

Phylogenetic analyses were performed with PAUP* 4.0b3a (Swofford, 1998) in separate and SA of nonmolecular, COI, 16S, and EF-1α data sets. All most-parsimonious reconstructions were obtained by a heuristic search with 100 random stepwise addition replicates using PAUP* default settings and keeping a maximum of 200 trees for each replicate. Successive approximation was performed on multiple equally parsimonious trees, with data reweighted based on the rescaled consistency index, and the same tree searching conditions as used for the unweighted data. Bootstrap proportions were determined by performing 500 replicates. Patterns of sequence evolution on the SA tree were inferred by parsimoniously optimizing character change using MacClade 4 (Maddison and Maddison, 1992). The number of steps (changes) per character were graphed and compared with a Poisson distribution by using a goodness-of-fit test (Sokal and Rohlf, 1995). This test can be applied to determine whether nucleotide changes are stochastic and equiprobable, consistent with the Jukes–Cantor (Jukes and Cantor, 1969) model of change (for discussion, see Olmstead et al., 1998). To test for saturation of character change, branch lengths were calculated by assuming equal change (Jukes–Cantor) and plotted against a corrected nucleotide substitution model (Tamura and Nei, 1993), which accounts for unequal nucleotide frequencies, variation in substitution rates, and transition bias. An asymptotic relationship in plots of sequence divergence under either model was taken to indicate saturation of character change. The values plotted were calculated directly from the branches of the SA tree rather than from uncorrected pairwise or patristic divergence. Thus, the "saturation curve" reflects the inferred amount of character change on the branches.

To test for the dependency of branch support (BS) and node level (see below), molecular branch lengths were fitted by maximum likelihood to a molecular clock, which scaled branch lengths throughout the tree. This was done separately for each of the three molecular partitions on the SA tree. Maximum likelihood estimations were performed under the following settings in PAUP*: estimated ti/tv ratio, base frequencies determined empirically, among-site rate variation approximated to a gamma distribution with 4 rate categories and shape parameter = 0.5.

Assessment of Character Congruence and Nodal Support

The quality of different alignments was assessed based on congruence with the phylogenetic signal in other data partitions. As reasoned by Patterson (1988) and DePinna (1991), the assumption that a trait is homologous (of common descent) is supported when the trait is a synapomorphy for a monophyletic group of taxa; that is, the distribution of the trait is congruent with other data that support this group as monophyletic. For DNA sequence data, this assessment is made summarily for all alignment-sensitive positions: Those alignment parameters that support a tree topology closest to a topology obtained from other data (morphological or non-alignment-sensitive DNA data) are considered the optimal parameters (Wheeler, 1995). The extent of congruence was assessed based on (1) topological congruence, by inspecting the number of nodes identical across data partitions, and (2) character congruence, as measured by the incongruence length difference (ILD) (Mickevich and Farris, 1981). The latter was determined by subtracting the tree lengths from the separate analyses of data partitions from the SA tree length and dividing this value by the SA tree length, to normalize values for the total number of character changes (Phillips et al., 2000).

Support for individual nodes was assessed according to Bremer (1994) by comparing tree length of the globally shortest tree to that of the shortest trees constrained to not contain the focal node. Constraint trees for this analysis were generated with TreeRot (Sorensen, 1996) and then used to calculate partitioned Bremer (branch) support (PBS) (Baker and DeSalle, 1997) and hidden PBS (HPBS) (Gatesy et al., 1999). PBS, the contribution of each partition to the branch support for a particular node in SA, is determined by calculating the tree length for each partition constrained to the SA tree to assess the performance of each partition on the SA tree, following precisely the steps described

in Baker et al. (1998). Raw PBS values may not be directly comparable because large data sets may exhibit more support simply because they have more characters. Thus, we standardized PBS by dividing this value by the minimum unconstrained tree length for each partition. HPBS is the amount of conflicting (negative values) or supporting (positive values) characters in a partition, revealed through SA instead of a separate analysis of partitions constrained to the SA tree. For each node, HPBS was calculated as the difference between PBS and BS values (Gatesy et al., 1999). For this analysis, the calculation of BS was modified in cases where the focal node of the SA tree is either unresolved or not observed in the separate analyses of the data partitions. In these cases, all trees compatible with the focal node of the SA tree were analyzed by comparing tree lengths of the shortest SA trees constrained to the focal nodes. The sum of HPBS values for all data sets at each node equals the total hidden BS (HBS).

Finally, the PBS values were also used to assess congruence between partitions in comparisons of data matrices obtained under different alignment parameters, performing a Spearman's rank correlation on the PBS values obtained for each partition on the SA tree (Sota and Vogler, 2001). This metric permits the analysis of support from each data partition under separate analysis and SA on a node-by-node basis and takes into account both the extent of agreement in tree topology and the magnitude of the signal. Better rank correlation between alignment variable and invariable partitions in the Spearman's test is taken to indicate that the homology assignment of bases is improved.

RESULTS

Phylogenetic Analysis of Gene Partitions

Direct optimization-alignment (Wheeler, 1996) was used to produce alignments under different ratios of gap cost to change cost. Under the default cost ratio of 2:1, the SA of all data resulted in a total sequence length of 525 characters for 16S and 782 characters for EF-1α and a single most-parsimonious tree of length 4,828 (Fig. 1). This tree shows the deep separation of *Ips* from a group of related genera *Pityokteines*, *Pityogenes*, and *Orthotomicus* that also includes *I. mannsfeldi*, *I. nobilis*, and the *concinnus* (now *Pseudips*) and the

latidens groups. Within *Ips*, the tree suggests three major clades (γ, ϕ, ε), discussed below, the composition of which is fairly stable to the variation of parameters in the analysis. Compared with the classification of S. L. Wood (1982), the *plastographus* species group is polyphyletic, and the *pini* and *perturbatus* species groups are paraphyletic with respect to the *plastographus* group and several of the other species groups, including the *grandicollis* group (itself paraphyletic with the inclusion of the *calligraphus* group except *I. sexdentatus*) and the (monophyletic) *tridens* group. The *emarginatus* group is also paraphyletic at the base of clades γ, ϕ, ε. The trees are fairly stable to successive approximations; one iteration resulted in a single shortest tree that differs from the unweighted trees only in the relative position of clades *I. cembrae–I. typographus* and *I. amitinus–I. perroti*.

Trees resulting from separate analyses of the data partitions differed in topologies and their degree of resolution (Fig. 2) but were not grossly discordant with the SA tree. Trees from the separate analysis share approximately half the nodes (except for the nonmolecular data, which shares only four nodes) with the SA tree. Each of the three molecular partitions supports the placement of some *Ips* species among *Orthotomicus*, and the distant position of the *concinnus* group (*Pseudips*). Although the 16S and EF-1α trees were congruent for peripheral nodes, they differed with regard to the arrangement of major clades (γ, ϕ, ε). Compared with the other molecular data sets, trees obtained from the COI partition differed more strongly from the SA trees in the peripheral clades. Trees based on the nonmolecular partition showed little concordance, mostly because of their low resolution. Incongruence between individual trees occurs mostly at nodes with low BS (Fig. 2). However, despite the differences of trees, in no cases were species exhibiting clear morphological disparity grouped together.

The four partitions contribute in different ways to the total support of the SA tree of Figure 1. The PBS contributed by 16S was greatest, 331.9, whereas COI and EF-1α contributed 216 and 238.4, respectively, and the contribution of the nonmolecular data set was negligible (Table 1, bottom). On a per character basis, that is, standardizing PBS for the number of steps each partition contributes to the SA, the signal in 16S was again
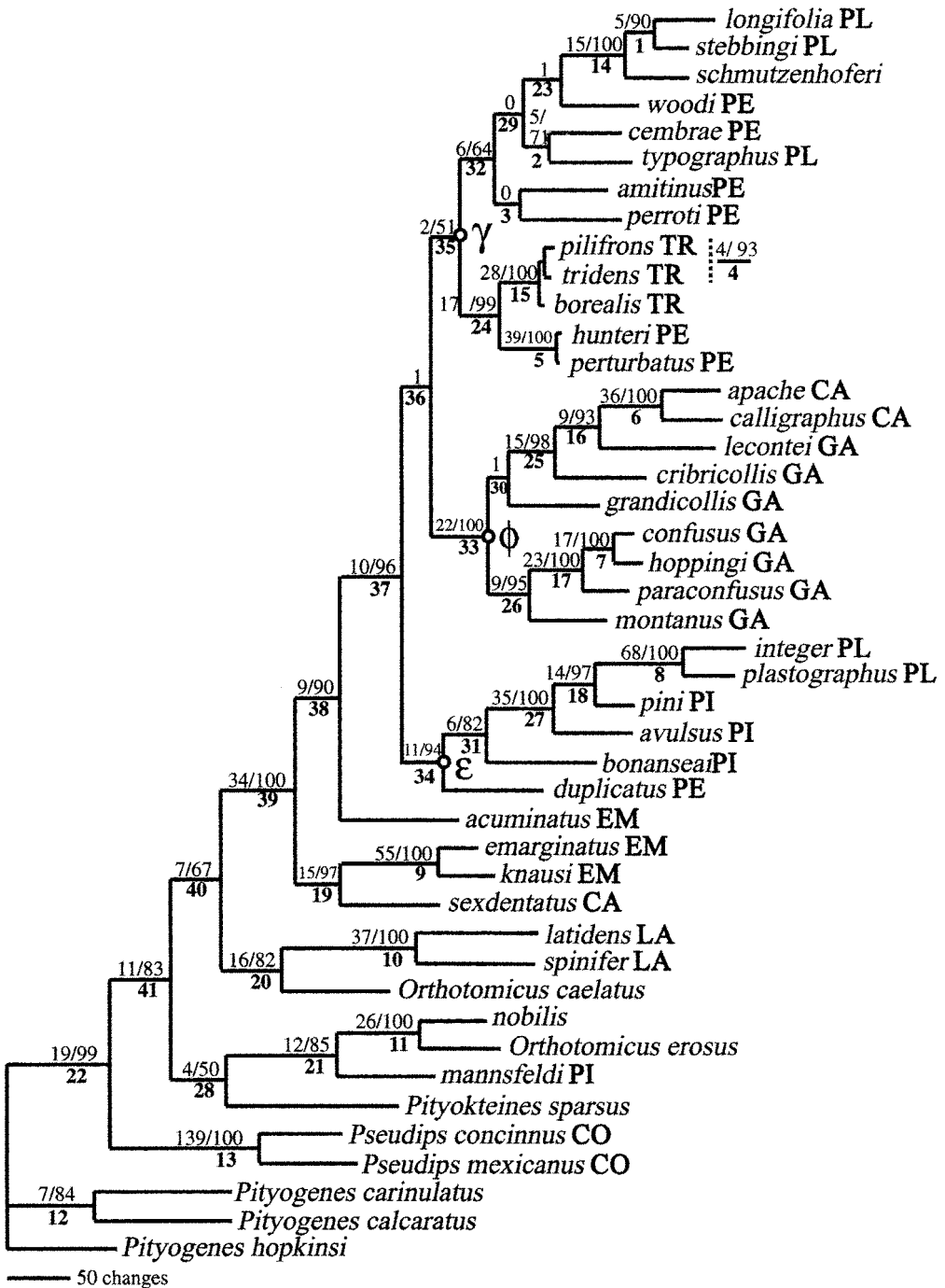
FIGURE 1. One of two most-parsimonious trees obtained from SA of all data partitions by using sequence alignments created by POY with gap cost: change cost = 2. CI = 0.38, RI = 0.53, length = 4,828 steps. BS and bootstrap values >50% are given above branches. Bold numbers below branches indicate nodes for the calculation of PBS values in Table 1. Greek letters refer to clades discussed in the text. Capital letters after the taxon names refer to species groups of S. L. Wood (1982): CA, *calligraphus*; CO, *concinnus*; EM, *emarginatus*; GA, *grandicollis*; LA, *latidens*; PE, *perturbatus*; PI, *pini*; PL, *plastographus*; TR, *tridens*. Zero BS represents nodes that are unresolved in the strict consensus of the shortest trees.

TABLE 1. Partitioned branch support (PBS), branch support (BS), hidden PBS (HPBS), and hidden BS (HBS) for nodes of combined data tree reconstructed with POY (Fig. 1) and partitioned data sets for gap cost = 2. For alternative alignment parameters, including the CLUSTAL alignment, only the total of the individual nodal values are given at the bottom of the Table. NA = not applicable because of missing taxa.

Node	% bootstrap	COI			16S rRNA			EF-1α			Nonmolecular			Total PBS	Total BS	HBS
		PBS	BS	HPBS	PBS	BS	HPBS	PBS	BS	HPBS	PBS	BS	HPBS			
1	90	8	9	−1	−1.5	−3	1.5	1	NA	NA	−1.5	0	−1.5	5	6	−1
2	71	−3	−3	0	2.5	1	1.5	4	1	3	1.5	−1	2.5	5	−2	7
3	>50	1	−3	4	−1.5	−1	−0.5	−1	−1	0	1.5	0	1.5	0	−5	5
4	93	−1	−1	0	0	0	0	5	5	0	0	0	0	4	4	0
5	100	23	23	0	7.5	10	−2.5	7.4	3	4.4	1.1	0	1.1	39	36	3
6	100	15	10	5	16	18	−2	3	3	0	2	0	2	36	31	5
7	100	9.7	9	0.7	6.5	8	−1.5	1.3	0	1.3	−0.5	0	−0.5	17	17	0
8	100	15	11	4	46	46	0	5	4	1	2	0	2	68	61	7
9	100	27.3	15	12.3	12.5	12	0.5	13.3	13	0.3	1.8	1	0.8	54.9	41	13.9
10	100	6	8	−2	28.5	14	14.5	3	25	−22	−0.5	0	−0.5	37	47	−10
11	100	11.3	10	1.3	14.2	15	−0.8	1	NA	NA	0.5	0	0.5	26	25	1
12	84	3	−1	4	4	4	0	0	NA	NA	0	0	0	7	3	4
13	100	23	20	3	21.5	21	0.5	90	−9	99	4.5	3	1.5	139	35	104
14	100	9.3	8	1.3	3.8	4	−0.2	−0.7	NA	NA	1.5	0	1.5	14.6	12	2.6
15	100	14	13	1	10.5	8	2.5	4	2	2	−0.5	1	−1.5	28	24	4
16	93	4.7	1	3.7	−2.2	−4	1.8	7.3	5	2.3	−0.8	−1	0.2	9	1	8
17	100	11	11	0	8.8	6	2.8	4	2	2	−0.8	0	−0.8	23	19	4
18	97	−5	−1	−4	20.5	18	2.5	0	−1	1	−1.5	0	−1.5	14	16	−2
19	97	−4	−3	−1	10.8	10	0.8	7.7	5	2.7	0.5	−2	2.5	15	10	5
20	82	0.5	−5	5.5	7.5	7	0.5	11	6	5	−3	−2	−1	16	6	10
21	85	6	−1	7	5	6	−1	0	NA	NA	1	−1	2	12	4	8
22	99	4	5	−1	9.5	7	2.5	5	15	−10	0.5	0	0.5	19	27	−8

TABLE 1. (*Continued*) Partitioned branch support (PBS), branch support (BS), hidden PBS (HPBS), and hidden BS (HBS) for nodes of combined data tree reconstructed with POY (Fig. 1) and partitioned data sets for gap cost = 2. For alternative alignment parameters, including the CLUSTAL alignment, only the total of the individual nodal values are given at the bottom of the Table. NA = not applicable because of missing taxa.

Node	% bootstrap	COI			16S rRNA			EF-1α			Nonmolecular			Total PBS	Total BS	HBS
		PBS	BS	HPBS	PBS	BS	HPBS	PBS	BS	HPBS	PBS	BS	HPBS			
23	>50	1	−4	5	−1.5	−2	0.5	−1	NA	NA	1.5	0	1.5	1	−6	7
24	99	10	7	3	5.5	4	1.5	1.8	0	1.8	−0.3	0	−0.3	17	11	6
25	98	10	1	9	3.5	4	−0.5	4	3	1	−2.5	−1	−1.5	15	7	8
26	95	−4.3	−1	−3.3	6.2	5	1.2	4.7	1	3.7	2.5	1	1.5	9.1	6	3.1
27	100	4.5	0	4.5	32.5	30	2.5	−1	−2	1	−1	0	−1	35	28	7
28	50	3	−6	9	0.5	1	−0.5	0	1	−1	0.5	−2	2.5	4	−6	10
29	>50	1	0	1	−1.5	−5	3.5	−1	−4	3	1.5	−2	3.5	0	−11	11
30	>50	3	1	2	−1.5	3	−4.5	−1	0	1	−1.5	−1	−0.5	1	3	−2
31	82	−2	−1	−1	5.3	3	2.3	2.6	1	1.6	0.1	0	0.1	6	3	3
32	64	3	1	2	3.5	3	0.5	1	−3	4	−1.5	−1	−0.5	6	0	6
33	100	1	−1	2	9.5	4	5.5	10.5	3	7.5	1	0	1	22	6	16
34	94	4	−1	5	3	3	0	4	4	0	0	−2	2	11	4	7
35	51	−5	0	−5	5.5	2	3.5	2	−2	4	−0.5	−1	0.5	2	−1	3
36	>50	−2	−1	−1	1.5	−2	3.5	3	1	2	−1.5	−1	−0.5	1	−3	4
37	96	5	−1	6	1	−1	2	4.5	2	2.5	−0.5	0	−0.5	10	0	10
38	90	2	1	1	5	0	5	4	2	2	−2	−2	0	9	1	8
39	100	2	−1	3	14	13	1	16	12	4	2	0	2	34	24	10
40	67	2	−8	10	0	−1	1	5	5	0	0	−2	2	7	−6	13
41	83	−1	−1	0	9.5	8	1.5	6	3	3	−3.5	0	−3.5	11	10	1
Total		**216**	**120**	**96**	**331.9**	**279**	**52.9**	**238.4**	**105**	**133.1**	**3.6**	**−16**	**19.6**	**789.6**	**488**	**301.6**
Gap cost = 1																
Total		**306.65**	**242**	**64.65**	**285.75**	**204**	**81.75**	**162.2**	**90**	**72.2**	**16.8**	**36**	**−19.2**	**771.4**	**572**	**199.4**
Gap cost = 3																
Total		**90.3**	**343**	**−252.7**	**498.3**	**365**	**133.3**	**236.2**	**161**	**66.1**	**−11.2**	**49**	**−60.2**	**813.6**	**918**	**−113.5**
CLUSTAL alignment																
Total		**163.7**	**113**	**50.7**	**89.4**	**8**	**81.4**	**88.15**	**97**	**−8.85**	**−0.25**	**−24**	**23.75**	**324.75**	**193**	**80.75**
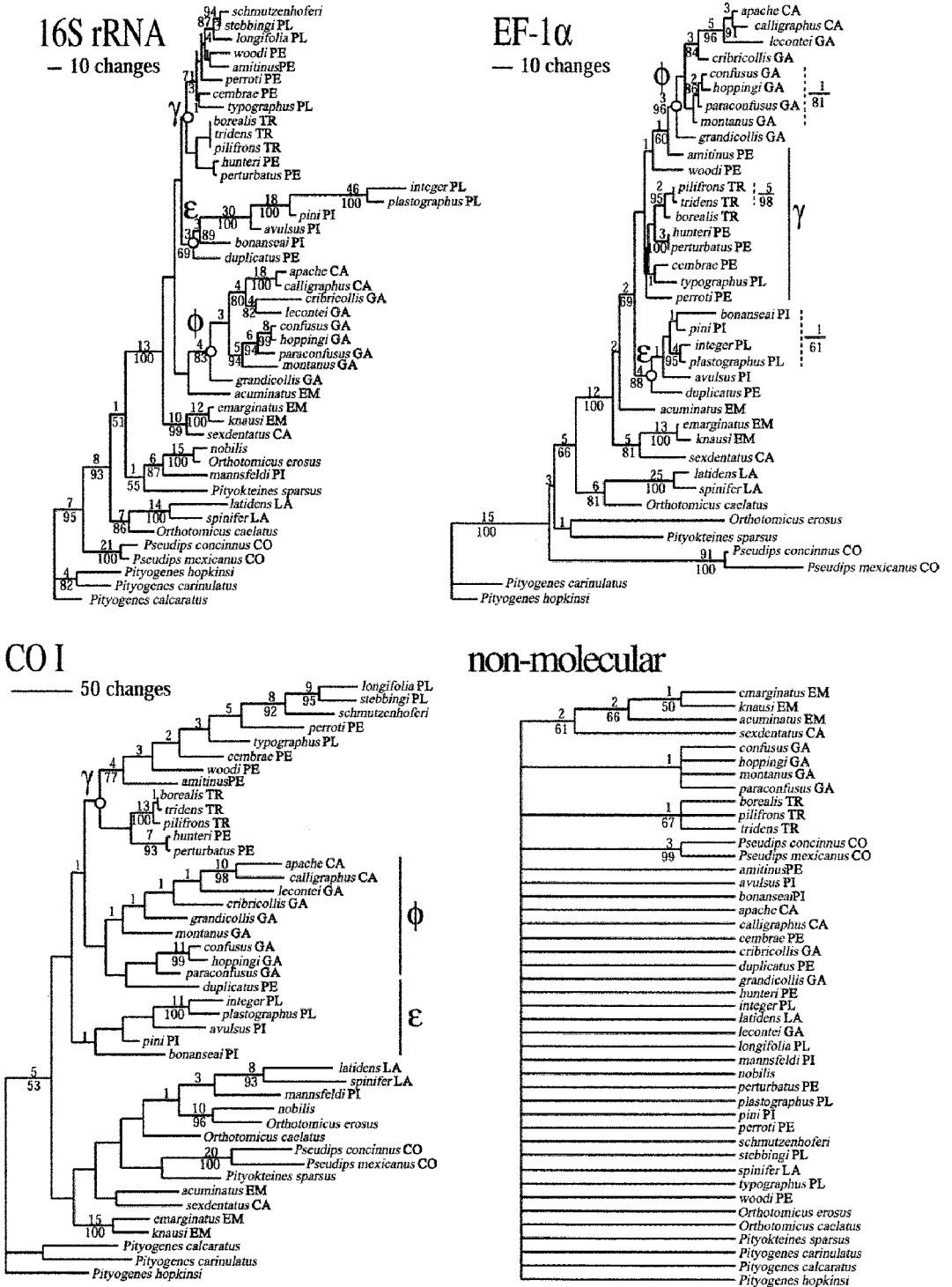
FIGURE 2.   Phylograms obtained from parsimony analyses of individual data partitions. BS values are given above the branches and bootstrap values >50% are given below. Nodes without BS values are unresolved in the strict consensus of the shortest trees. For all trees, symbols are discussed in the text and legend to Figure 1. 16S: 1 of 6 shortest trees of cost 1,171, CI = 0.56, RI = 0.70; EF-1α: 1 of 16 shortest trees of cost 860, CI = 0.65, RI = 0.72; CO I: 1 of 2 shortest trees of cost 2672, CI = 0.23, RI = 0.39; nonmolecular data (consensus of 65,000 trees): cost 52, CI = 0.58, RI = 0.84.

TABLE 2. Pairwise analysis of congruence among data sets based on POY analyses. Note that significant congruence of 16S rRNA and EF-1α is achieved with a gap cost of 2.

	EF-1α			COI			16S rRNA			Nonmolecular		
	1[a]	2	3	1	2	3	1	2	3	1	2	3
EF-1α	—	—	—	0.006	0.007	0.014	0.007	0.006	0.01	0.018	0.018	0.013
CO I	0.26	0.09	0.18	—	—	—	0.007	0.01	0.013	0.007	0.008	0.012
16S rRNA	−0.042	0.34*	0.08	0.27	0.29	0.3	—	—	—	0.018	0.015	0.007
Nonmolecular	0.42	0.29	0.09	0.31	0.31	0.48 *	0.22	0.15	0.25	—	—	—

[a] 1, 2, and 3 refer to gap cost values.
* $P < 0.01$.
Above diagonal, ILD/steps of SA tree. Below diagonal, Spearman's rank correlation of the PBS.

the greatest, with a value of 0.28, closely followed by EF-1α (0.27), whereas the contributions of COI and the nonmolecular data set were considerably less, 0.08 and 0.05, respectively. This attests to the fact that despite the large number of characters, the COI partition does not overwhelm the smaller data partitions, as might be expected. When the contribution of each partition for each node in the SA tree was assessed, we found a large number of nodes supported by some partitions and contradicted by others, in some cases differing by high numerical values (for example, node 19, Table 1). Only 8 of the 41 nodes are universally supported with positive PBS values in all partitions, although this value increases to 19 nodes if only the molecular data sets are considered. No one data set provides appreciably more support or conflict for individual nodes, although the nonmolecular data set provides little support for any node. Ranks of nodes according to the extent of node support (PBS values) were poorly correlated in Spearman's rank test, except for a significant rank correlation of EF-1α and 16S (Table 2, middle column for the 2:1 gap cost ratio). Despite this generally low correlation in node support, character conflict between most partitions based on ILD values was weak, the values being greatest for the pairwise comparisons of nonmolecular data with the two genes with length-variable 16S and EF-1α (Table 2). Data interactions between different partitions were further analyzed based on HPBS values; generally high levels of hidden support (positive HPBS) were revealed by combining data in the SA (Table 1). Similar to the distribution of signal in the PBS analysis, hidden support is widely scattered between nodes and data partitions throughout the SA tree (Table 1).

Variation of Alignment Parameters

We explored the effects of different alignment parameters on tree topology and data interactions under lower and higher gap cost ratios. When gap cost was 1, the number of characters increased for 16S and EF-1α (to 558 and 791, respectively). Also, two trees were obtained whose consensus differed by recovering only two of the three major clades (ε and φ), the relationship of these clades among the species of clade γ being unresolved (not shown). Under a gap cost of 3, the number of characters decreased for 16S and EF-1α (to 502 and 766, respectively). Clades γ, ε, and φ were recovered but with minor rearrangements within these clades and with clade φ as sister to the other two clades (not shown). However, the inclusion of *I. acuminatus* within clade γ was surprising, given the morphological disparity of this species. Finally, a phenetic alignment produced by CLUSTAL recovered the three major clades, γ, φ, and ε. Under all of these alternative alignments, the strict consensus trees obtained from each analysis were less resolved than those obtained with a gap cost of 2. Yet, despite several differences in the details of tree topologies, the SA under these alignment parameters generally also supported the placement of some *Ips* species among *Orthotomicus*; recovered the *grandicollis* group as paraphyletic with respect to *calligraphus* group; found *pini, plastographus*, and *perturbatus* groups to be broadly polyphyletic; and recovered the monophyletic *tridens* group.

Different alignment parameters had a clear effect on the extent of character congruence between data partitions. Although the total PBS for all data partitions combined was similar under all three gap cost to change cost parameters, the contribution of the four

partitions to the total PBS differed substantially (Table 1, bottom rows). The PBS was more or less evenly distributed among the molecular data sets for gap costs of 1 and 2, but under a gap cost of 3 nearly half of the total PBS was contributed by the 16S, and total PBS for COI dropped to very low levels, indicating that the SA tree topology under these conditions is mostly determined by the 16S partition. In addition, the analysis with a gap cost of 3 is the only one in which the nonmolecular data showed clear conflict with the SA. Total PBS determined for the phenetic (CLUSTAL) alignment was less than half the total PBS determined by the direct-optimization alignments (Table 1, bottom). In the latter alignment, the PBS for COI was the greatest and contributed about half of the total data support of the combined analysis. Comparisons of HBS in the gap costs of 1 and 2 alignments revealed substantial hidden support (sensu Gatesy et al., 1999) (199.4 and 301.4, respectively) and lower positive values under the phenetic alignment (80.75). In contrast, hidden conflict is revealed for a gap cost of 3 (−113.5), in particular in the COI partition (Table 1). The relatively greater conflict in the gap cost 3 alignment was also apparent in the increase in ILD (0.02 per change of the SA tree) versus 0.013 and 0.014 under gap costs of 1 and 2, respectively. The largest ILD was found after CLUSTAL alignment (0.022 per change on the SA tree, 0.019 if gapped positions are excluded), an increase in ILD of nearly 60% over the best alignment.

Pairwise comparisons of partitions resulted in similarly low (Farris et al., 1994) character incongruence, but again, the values were generally greatest under a gap cost of 3 (Table 2) and the CLUSTAL alignment. Spearman's rank correlations of PBS values in pairwise comparisons were mostly nonsignificant for gap cost alignment parameters of 1 and 3, except for a significant correlation between nonmolecular and COI data for a gap cost of 3. The phenetic alignment also performed poorly in this test (not shown), but removal of the gapped characters generally improved the correlation of PBS values (to 0.23, not significant). Thus, the extents of congruence in pairwise comparisons were similar to the values of ILD from the SA of all partitions (above). They were also consistent with the results from the analysis of total HPBS values (Table 1), which suggested the most hidden conflict

under high gap costs and under the phenetic alignment. Thus these alignment parameters result in generally less satisfactory homology assignment, given the congruence criterion. The fact that removal of gapped characters in the CLUSTAL alignment reduced the apparent character conflict also supports the conclusion that a suboptimal alignment compounds the effects of any conflicting phylogenetic signal that might affect these data. Thus, the analysis under gap costs of 1 and 2 exhibited the most concordance among data sets; only the gap cost of 2 resulted in a significant correlation of PBS in the two alignment sensitive partitions 16S and EF-1α, however, and therefore represents the best homology assignment among the alignment parameters tested.

Gene Properties and Implications for Branch Support

Although character concordance is greatest with a gap cost of 2, the concordance is not overwhelming and BS values remain negative for several nodes (Table 1). Differences in the rates of divergence and amounts of saturation among and within the genes may in part explain this conflict. The extent of sequence divergence, calculated along the branch lengths of the SA tree, is approximately twice as much in COI as in either 16S or EF-1α (Fig. 3A; Table 3). COI also exhibits a greater proportion of phylogenetically informative nucleotide positions (43%) than does 16S (38%) or EF-1α (26%) (Table 3). For each gene, both divergence and phylogenetically informative characters are aggregated in the length-variable regions and at third-codon positions. In 16S the variable regions represent 966 of the 1,187 steps contributed by that partition, and the intron sequence in EF-1α contributes 461 of 872 steps, yet they constitute only about one-half or one-fifth, respectively, of the total number of characters in either partition (Table 3). Similarly, the third-codon positions contribute the vast majority of the remaining phylogenetic information (295 steps for EF-1α and 2,284 for COI; Table 3). Yet, the extent of homoplasy for each of the classes of characters is not closely correlated with the proportion of variable sites (Table 3). For example, the consistency index (CI) of the EF-1α intron sequences (CI = 0.777), the 178 characters of which contribute a similar amount of steps to the SA tree as do the remaining 611 characters of
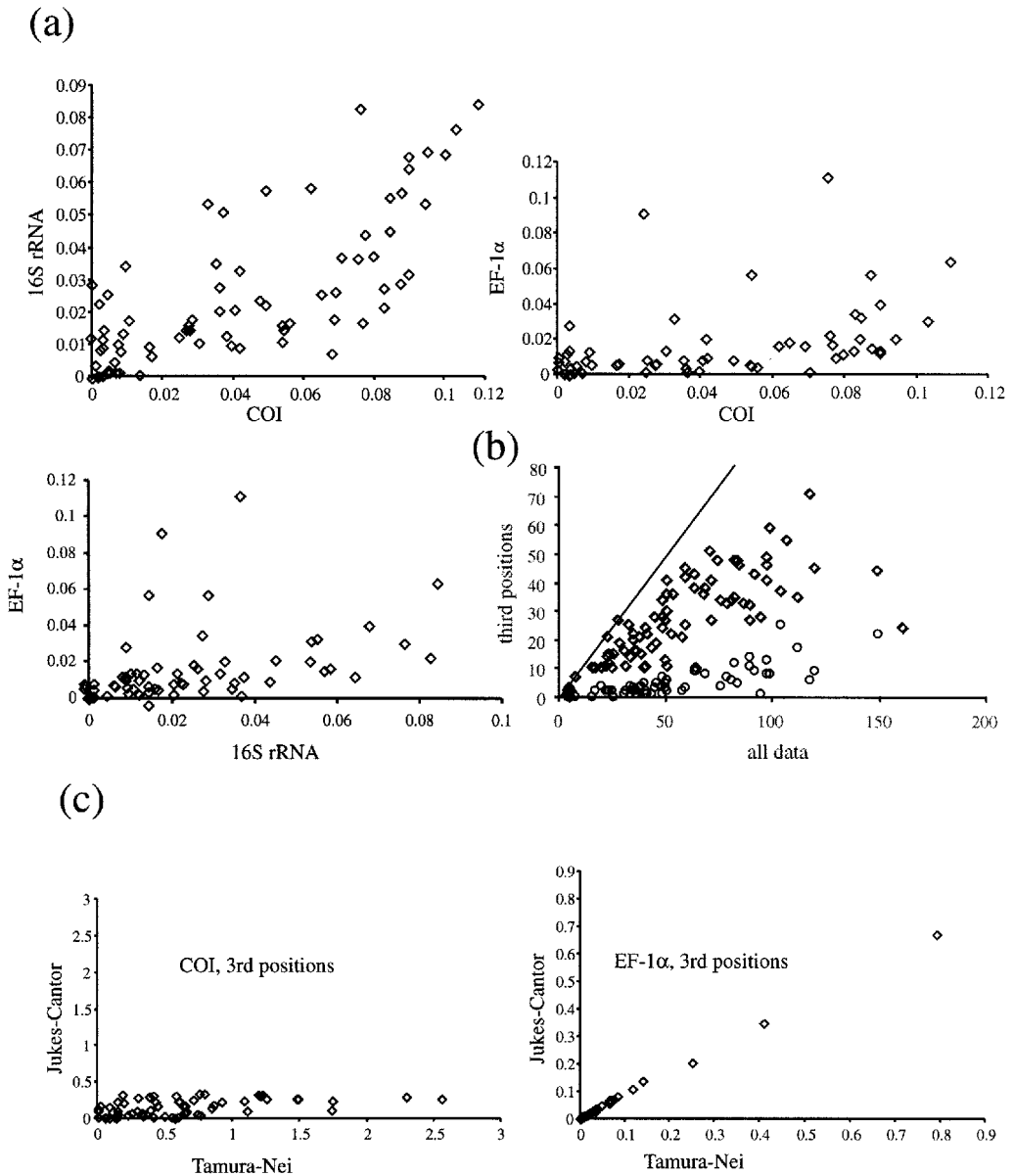
FIGURE 3.   (a) Scatter plots for COI, 16S, and EF-1α branch lengths constrained to Figure 1 and determined under the Jukes–Cantor (1969) model of nucleotide change. (b) Character change of COI (diamonds) and EF-1α (circles) third-codon positions versus character change of all data determined from branch lengths of Figure 1. (c) Relationship between divergence according to the Jukes–Cantor nucleotide model versus the Tamura and Nei (1993) model for third-codon positions of COI and EF- 1α.

the exon, is more than for any of the three classes of coding regions (CI = 0.491, 0.524, and 0.532 for first-, second-, and third-codon positions); accordingly, the greater average variation per site in a partition does not necessarily result in greater homoplasy.

In fact, the number of steps per site varies widely, with a large proportion of invariable sites present in each of the three gene

partitions and a large number of invariable sites present in all partitions except the third-codon positions of COI (Fig. 4). The distributions of changes per site in no case fit a Poisson distribution (even when the large number of invariant sites are removed from the test), indicating substantial site rate heterogeneity between sites and inconsistency with the Jukes–Cantor model of nucleotide

TABLE 3.   Nucleotide divergence patterns observed for COI, 16s, and EF-1α for *Ips* and outgroup species. Branch lengths were determined from the SA tree in Figure 1. Mean ti/tv ratios were calculated from average pairwise species comparisons.

	Size matrix	Variable sites	Inform sites	Cost on SA tree	CI	Mean branch length	ti/tv
16S rRNA							
Total	525	309	234	1,187	0.554	0.023	0.54
Length Invariable reg.	265	67	60	221	0.398	0.009	1.2
Length variable reg.	260	242	174	966	0.590[a]	0.058	0.39
EF-1α							
Total	782	234	187	872	0.655	0.015	1.61
Total exon	611	162	115	411	0.518	0.016	2.8
1st codon pos.	204	21	14	53	0.491	0.007	1.9
2nd codon pos.	203	23	15	63	0.524	0.008	1.2
3rd codon pos.	203	118	86	295	0.532	0.041	3.46
Intron	178	158	99	461	0.777[a]	0.038	0.68
COI							
Total	769	340	311	2,699	0.224	0.04	1.22
1st codon pos.	256	68	58	363	0.262	0.014	4.47
2nd codon pos.	256	23	16	52	0.615	0.002	0.89
3rd codon pos.	257	249	237	2,284	0.209	0.118	1.23

[a] Note that many informative sites are contributed by gaps (weight = 2).

change (Goldman, 1990). The number of character changes per site, however, are by far greatest in third-codon positions of COI, which has only a small number of invariable sites (Figs. 3B, 4). This may be an indication that character changes are largely saturated (i.e., additional character changes do not lead to further divergence of sequences). This is supported by plots of uncorrected versus corrected branch lengths, which revealed
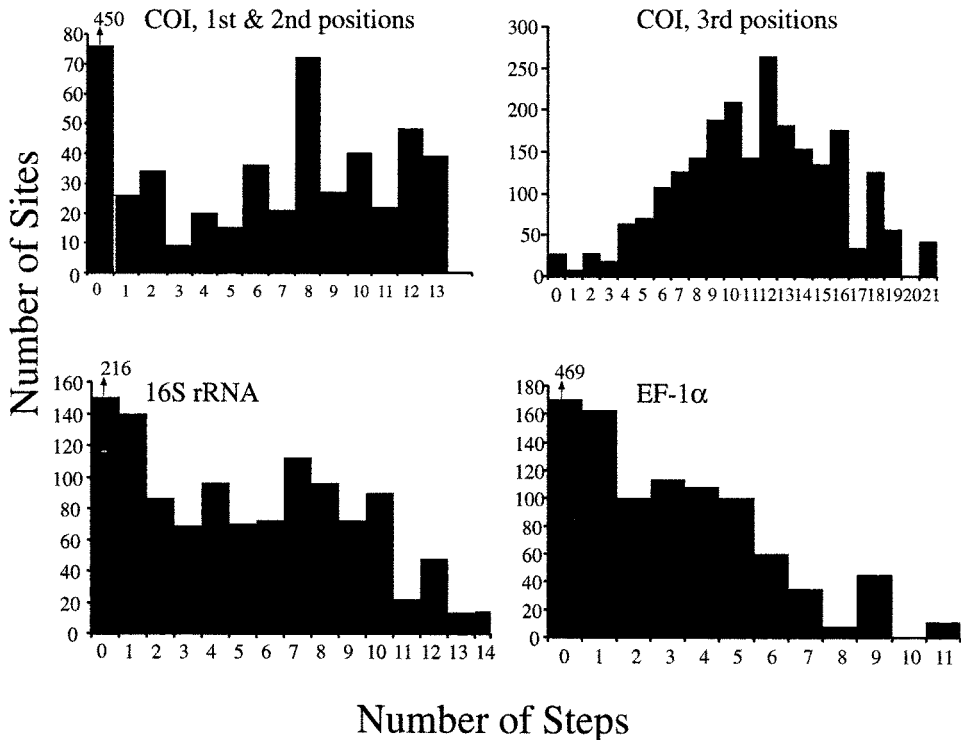

FIGURE 4.   Distribution of nucleotides sites exhibiting parsimony-inferred change for COI first- and second-codon positions, COI third-codon positions, 16S all positions, and EF-1α all positions. All distributions are significantly different ($\alpha = 0.001$) from a Poisson distribution.

minimal increase of sequence divergence with increased phylogenetic divergence under the Jukes–Cantor model but continued to increase under a more complex Tamura–Nei model in the third-codon positions of COI (Fig. 3C). In contrast, the Jukes–Cantor versus Tamura-Nei plot deviated little from a linear correlation in the third-codon positions of the EF-1α gene (Fig. 3C), indicating no such saturation effect in this partition.

Differences in the type and rate of character changes between partitions are also evident from studies of ti/tv bias. Between close relatives, ti/tv ratios are high; with increased divergence, the ti/tv ratio drops off in all loci but tends towards an asymptote at different values in each case. Whereas 16S exhibits a value of 0.5, as is expected under the Jukes–Cantor model of equal change, both EF-1α (ti/tv $\approx$ 2) and COI (ti/tv $\approx$ 1; Cognato and Sperling, 2000) reach asymptotes at values greater than that for neutral expectations. Similar conclusions are also apparent from calculations of mean ti/tv ratios from pairwise sequence comparisons (Table 3). The differences in ti/tv observed between genetic loci could be due to the different bias in the type of character changes in these loci. Nucleotide changes in protein-coding regions are predominantly at third-codon positions (Table 3; Fig. 3) and thus may be biased as a consequence of the preponderance of synonymous changes and the pressure to maintain codon usage. However, the high AT compositional bias in mtDNA would result in a relatively higher proportion of transversions (mostly those involving A and T) and thus keep the skew in ti/tv ratio relatively lower. In 16S, the skew towards A/T transversions is an even stronger possibility because that gene is not constrained by codon usage; selection from secondary structure interactions may also be involved.

Given the widely differing amounts of sequence divergence and homoplasy, the analysis of nodal support was extended to test whether particular partitions were more relevant to resolving certain portions of the tree, for example, if support for deep nodes was mostly derived from the more conserved data partitions. To obtain comparable measures of sequence divergence throughout the tree (Fig. 1), we constrained the molecular partitions to conform to a molecular clock under a maximum likelihood model.

The resulting branch lengths provide ultrametric estimates of divergence for every node and hence should provide comparable divergence data. We observed that BS values generally decreased towards deeper nodes (Fig. 5) for each of the three molecular data sets, particularly if only the saturated COI third-codon positions were considered ($r = -0.74$) (this correlation was also observed for 16S [not shown], with $r = -0.28$). Because this finding applies to all three data sets, including the more conserved EF-1α and 16S, it argues against the notion that conservative data are more useful at recovering deeper nodes.

The negative correlation of BS values with distance from the tips of the tree produces largely negative BS values at the deep nodes, indicating that alternative trees would better reflect the relationships among the deeper branches according to these data. This effect is seen in COI, particularly in third-codon positions (Fig. 5A), but also partly in EF-1α (Fig. 5B), in which the statistics are based on fewer character changes. Yet, several of these nodes have substantial net support, as reflected in the values for HPBS, which increase towards deeper nodes. Thus, total support for these deeper nodes arises almost entirely from "hidden support," which is revealed only through the SA of all characters in the partition. The discrepancy between total BS values for the full COI data and the full EF-1α data, and the values for the respective third-codon positions, indicates the phylogenetic importance of the first- and second-codon positions and the EF-1α introns for extracting the phylogenetic signal.

DISCUSSION

Differences in topologies (Fig. 2) and node support (Table 1) for each partition may suggest that the genes studied reflect different histories, possibly confirming the expectation of incongruence of gene trees at the species level. However, measuring the incongruence of trees obtained from individual partitions is complex. We investigated two factors that can conceal the phylogenetic signal contained in the data: errors with homology assignment in length-variable regions and differences in types and rates of character transformations. Some degree of conflict is ubiquitous in phylogenetic data, and its implication for molecular systematics
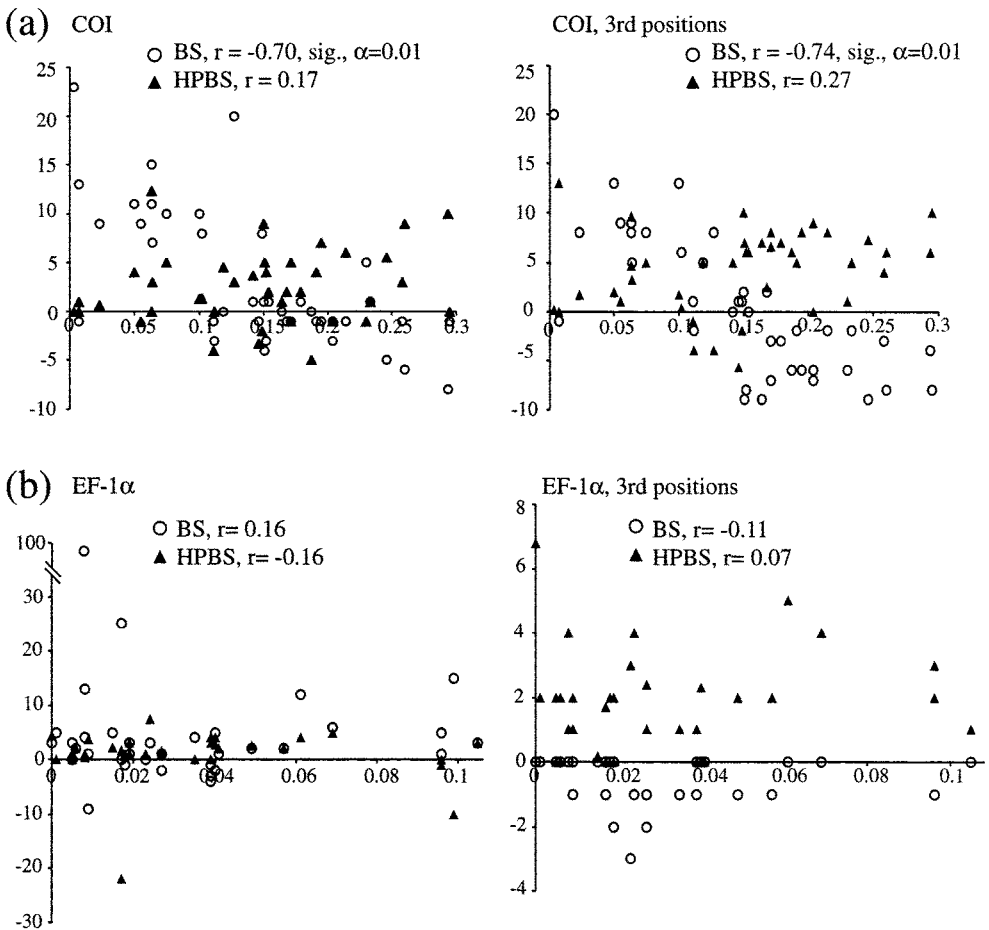
FIGURE 5.   Relationship between PBS values and nodal distance from tips of tree. (a) All COI data and COI third-codon positions; (b) all EF-1α data and EF-1α third-codon positions. Nodal distance is the maximum likelihood branch length from a given node to the tip of the tree (see text for details).

has become more relevant with the growing data from multiple genetic loci. Clearly, we saw strong interactions between and within data partitions, and found it difficult to pin-point "conflict" as a discrete phenomenon assignable to particular partitions.

Alignment has long been recognized as a source of data conflict. Using POY, data interaction between alignment-variable and -invariable portions can now be assessed and brought to bear on the tree topology simultaneously. This procedure strongly reduced conflict in comparison with procedures that do not permit data interaction during alignment, such as the CLUSTAL method. The latter is entirely static, and under the parameters used here, the amount of congruence and total tree support is much lower

than in any of the POY alignments. Our POY analysis favors the default alignment parameter (gap cost:change cost = 2:1) over higher and lower gap costs, although a wider exploration of the parameter space would be needed before drawing conclusions about the optimal values. The observations from the current analysis indicate that increasing gap costs to assign relatively more weight to the alignment-sensitive regions, in particular the 16S region, results in increased conflict with the COI partition (Table 1). An additional increase in gap cost can be expected to enhance this effect because indels receive even more weight, resulting in length variation to become increasingly more important for determining the topology of the SA tree and probably increasing total conflict across

the partitions. The study of data interactions indicates a clear upper limit to the gap cost.

An intriguing finding from the study of multiple data sets is the widespread observation of "hidden support"—the fact that signal contained in individual data partitions is immaterial in the combined analysis if it is not also shared by other partitions. In that case, the hidden signal common to all partitions (presumably because of common history) is greater than noise or misleading information inherent only to a single partition (Gatesy et al., 1999). The fact that a large amount of concealed signal is present in these data suggests that the specifics of character evolution in different genetic loci, rather than different gene history, should be invoked as the major explanation for the apparent incongruence between these data sets.

Currently, little is known about the quantity of hidden support in molecular data in general, and about the performance of HBS values to measure it. Species-level phylogenetic studies, as used here, are particularly useful to address these issues because they represent the most complete set of taxa available to test character changes in a phylogeny. Phylogenetically misleading signal (i.e., homoplasy) and a departure of character change from random expectations (as revealed by the deviation from a Poisson distribution in the types of character transformations; Olmstead et al., 1998) in molecular data has many causes and may lead to different patterns of homoplastic character variation and possibly saturation. The analysis of ti/tv ratio (see Results) and character variation and saturation (Figs. 3 and 4) clearly demonstrates the differences in evolutionary dynamics of the loci analyzed here. How these various biases could affect the recovery of the phylogenetic signal will be difficult to quantify, and the application of different weighting schemes for "correcting" such differences in a parsimony framework (e.g., Cunningham, 1997) would have to be extremely complex.

Biases in nucleotide change should affect the more divergent taxa and the recovery of deeper nodes more strongly than the tips of the tree; it is commonly held, therefore, that an optimal level of sequence divergence exists for resolving relationships at a particular hierarchical level. In fact, several stud-ies suggest that mitochondrial, ribosomal, and nuclear genes support different taxonomic levels (Friedlander et al., 1992; Brower and DeSalle, 1994, 1998; Simon et al., 1994; Funk et al., 1995; Miyamoto and Fitch, 1995; Whitfield and Cameron, 1998; Reed and Sperling, 1999). The basal nodes in our plots of nodal BS versus distance from the tips (in the most apparently saturated data, the third-codon position in the COI; Fig. 5B) were found to be almost exclusively supported by hidden support, that is, support extracted from the data only in the interaction with other (types of) nucleotide partitions. Presumably, the combining with other data, including the first- and second-codon positions of the COI gene, which are affected by different sets of homoplasies, recovers phylogenetic signal difficult to detect beneath the homoplastic character variation in the third-codon positions.

We find it striking that different loci exhibit very different patterns of node support (PBS and HPBS) at any given node (Table 1). Besides a clear general trend from higher to lower nodal support values towards the base of the tree, the PBS values vary substantially between data partitions at given nodes, and conflict and support for particular nodes emerge equally from the combining of data sets (although in sum the extra support prevails over conflict). Because few data sets to date have been analyzed this way, it is difficult to speculate on the generalities of these observations. Further subdivision of the data (according to synonymous/nonsynonymous positions, or intron vs. exon sequences) further attest to the diverse nature of the data conflict. For example, the EF-1α coding positions are almost entirely in conflict with the tree (negative PBS) but their hidden support is high (positive HPBS) (Fig. 5). These effects need further exploration through comparisons with other data sets and at different hierarchical levels. However, despite the highly diverse patterns of character change and the persistent differences between sets of characters, the basic topologies of resolved clades obtained from individual partitions do not differ strongly. Saturation and various types of bias in the data, including errors (or randomness) in homology assignment because of difficulties with aligning, may in this case have little effect on tree topology.

Even a weak signal may be fairly resilient to "noise" in the data (Wenzel and Siddall, 1999).

Lastly, the findings of our study leave the question about the combinability of these data. SA allows for the recognition and quantification of hidden support and conflict for each node among data sets. For example, a large amount of hidden support within the apparently saturated third-codon positions would have gone undiscovered if these data were analyzed separately. Our data thus add to the evergrowing list of studies discovering support for phylogenetic signal from the interactions of different data types. In a framework that argues against combining conflicting data partitions (Bull et al., 1993), some would exclude from the analysis the alignment-variable regions in 16S and EF-1α (Swofford et al., 1996) or other types of strongly conflicting data (Huelsenbeck et al., 1996). However, conflicting data in *Ips* contain information relevant to the discovery of the phylogenetic signal, which in part is concealed by saturation, uncertainty of alignment, selection for particular amino acid replacement, and possibly other factors. Conflict, therefore, is difficult to define once the data have been more closely dissected and consistent differences between classes of characters (presumed process partitions) have been detected. Given the complexity of conflicting signal in every partition, excluding a given data partition on the basis of a summary statistic of conflict, such as the partition homogeneity test, can hardly be justified.

Unless there is evidence for introgression of particular genotypes (Harrison, 1993), paralogy, or recombination between unrelated genomes (Doyle, 1997), inferring these processes from the observation of phylogenetic conflict alone is problematic, because of the difficulty in identifying true incongruence of gene trees and species trees (Sota and Vogler, 2001). Differences in phylogenetic signal resulting from stochastic processes and various mutational and selectional biases produce apparent incongruence even if the phylogenetic history of these loci is congruent. Removing presumed incongruent data partitions is likely to change the interactions of the remaining partitions but does not constitute a solution to the problem of conflict. Sound estimates of phylogeny at the species level will probably result only from the inclusion—and simultaneous analysis—of a large number of characters.

Taxonomic Revision of Ips

A stable revision of *Ips* necessitates a robust phylogenetic hypothesis, which required the above thorough data analysis. SA using POY and a gap cost:change cost of 2 (Fig. 1) produced the best-supported *Ips* phylogenetic hypothesis to date. The revised classification of species within *Ips* is based on this tree (Appendix). Synapomorphies define monophyletic groups (Fig. 1) in this revised classification. Thus molecular and nonmolecular characters contribute to the definition of groups through parsimony-based phylogenetic analysis. Morphological characters are used to diagnose taxa, which are named in accordance with established nomenclatural code (I.C.Z.N., 1999). Four subgenera are proposed to replace species groups (S. L. Wood, 1982) and are named in association with clades of *Ips*. A comparison of the previous and revised classifications is given in Table 4. Six species are considered *incertae sedis*. Specimens of *I. chinensis*, *I. hauseri*, *I. nitidus*, and *I. ussuriensis* were unavailable for complete comparison of characters. The placement of *I. acuminatus* is not inclusive of a clade defining a subgenus, and inclusion of *I. duplicatus* in the proposed classification would cause taxonomic ambiguity (discussed in the Appendix). In addition, the following nomenclatural changes for species were made. *Ips cribricollis* is removed from synonymy with *I. grandicollis* (Wood, 1977; Wood and Bright, 1992) and *I. schmutzenhoferi* is removed from synonymy with *I. stebbingi* (Wood, 1992; Wood and Bright, 1992). Lanier (1987) provided detailed evidence for morphological and chromosomal differences between *I. cribricollis* and *I. grandicollis*. Furthermore, *I. cribricollis* and *I. grandicollis* are reciprocally monophyletic, according to COI data (Cognato and Sperling, 2000), and this study suggests that *I. cribricollis* is sister to *I. grandicollis* and four additional species (Fig. 1). Morphological differences, such as the median lobe of the male genitalia, and reciprocal monophyly distinguish *I. schmutzenhoferi* and *I. stebbingi*. Both *I. cribricollis* and *I. schmutzenhoferi* represent diagnosable lineages separate from other closely related species.

TABLE 4. Comparison of *Ips* species groups (defined by S. L. Wood, 1982) to the proposed revision based on the current analysis (*Eurasian species).

Species groups	Monophyletic groups
concinnus group—*I. concinnus, I. mexicanus*	*concinnus* group—*Pseudips*
emarginatus group—*I. emarginatus, I. knausi, I. acuminatus**	*Ips (Emarips)—I.emarginatus, I. knausi, I. sexdentatus**
pini group—*I. pini, I avulsus, I. bonanseai, I. mannsfeldi**	
plastographus group—*I. plastographus, I. integer, I. typographus* I. longifolia*, I. stebbingi*, I.schmutzenhoferi** was not addressed	*Ips (Bonips)—I. bonanseai, I. plastographus, I. integer, I. pini, I. avulsus (I. mannsfeldi* is moved to Orthotomicus)*
perturbatus group—*I. perturbatus, I. hunteri, I. woodi, I. perroti, I. cembrae*, I. amitinus*, I. duplicatus**	*Ips (Ips)—I. typographus*, I. cembrae*, I. woodi, I. longifolia*, I. stebbingi*, I. schmutzenhoferi*, I. perroti, I. amitinus*, I. tridens, I. pilifrons, I. borealis, I. perturbatus, I. hunteri*
tridens group—*I. tridens, I. pilifrons, I. borealis*	
grandicollis group—*I. paraconfusus, I. confusus, I. hoppingi, I. lecontei, I. montanus, I. grandicollis*	*Ips (Granips)—I. grandicollis, I. lecontei, I. montanus, I paraconfusus, I. confusus, I. hopping, I. cribricollis, I. calligraphus, I. apache*
calligraphus group—*I. calligraphus, I. sexdentatus**	
latidens group—*I. latidens, I. spinifer, I. erosus**	*(latidens* group—*moved to Orthotomicus)*

Ips latidens (LeConte), *I. spinifer* (Eichhoff), *I. mannsfeldi* (Wachtl), and *I. nobilis* (Wollaston) are removed from *Ips*. Characters of the antenna, elytra, male genitalia, and phylogenetic placement suggest that these species are more closely related to *Orthotomicus* Ferrari. For nomenclatural purposes they are tentatively placed within *Orthotomicus*; however, a phylogenetic revision including additional Ipini is needed to elucidate their classification.

This revised classification (Table 4) is a conservative and perhaps coarse assessment of species associations. Indeed, more data will help resolve poorly supported nodes and streamline the taxonomy. However, this revision of *Ips* represents species groups as evolutionary lineages. The subgenera are considered monophyletic and, in conjunction with the tree (Fig. 1), can be used to make evolutionary inferences of species' biology. For example, America has been suggested as the geographic origin of *Ips* because of the greater diversity of Nearctic *Ips* species (S. L. Wood, 1982). However, the results of this study do not support this hypothesis (Fig. 1). Eurasian species are phylogenetically distributed among the North American species, and the Eurasian *I. sexdentatus* and *I. acuminatus* are basal to the majority of *Ips* species, indicating that *Ips* diver-

sification involved exchange between North American and Eurasian lineages. One subgenus (including the *grandicollis* and *calligraphus* species groups; S. L. Wood, 1982) and the *tridens* species group (S. L. Wood, 1982) unambiguously represent a major North American *Ips* species diversification exclusive of Eurasian species.

REFERENCES

BAKER, R. H., AND R. DESALLE. 1997. Multiple sources of character information and the phylogeny of Hawaiian Drosophilids. Syst. Biol. 46:654–673.

BAKER, R. H., X. B. YU, AND R. DESALLE. 1998. Assessing the relative contribution of molecular and morphological characters in simultaneous analysis trees. Mol. Phylogenet. Evol. 9:427–436.

BREMER, K. 1994. Branch support and tree stability. Cladistics 10:295–304.

BROWER, A. V. Z., AND R. DESALLE. 1994. Practical and theoretical considerations for choice of a

DNA sequence region in insect molecular systematics, with a short review of published studies using nuclear gene regions. Ann. Entomol. Soc. Am. 87:702–716.

BROWER, A. V. Z., AND R. DESALLE. 1998. Patterns of mitochondrial versus nuclear DNA sequence divergence among nymphalid butterflies: The utility of wingless as a source of characters of phylogenetic inference. Insect Mol. Biol. 7:73–82.

BROWER, A. V. Z., R. DESALLE, AND A. VOGLER. 1996. Gene trees, species trees, and systematics: A cladistic perspective. Annu. Rev. Ecol. Syst. 27:423–450.

BULL, J. J., J. P. HUELSENBECK, C. W. CUNNINGHAM, D. L. SWOFFORD, AND P. J. WADDELL. 1993. Partitioning and combining data in phylogenetic analysis. Syst. Biol. 42:384–397.

CACCONE, A., G. D. AMATO, AND J. R. POWELL. 1988. Rates and patterns of scnDNA and mtDNA divergence within the *Drosophila melanogaster* subgroup. Genetics 118:671–683

CANE, J. H., M. W. STOCK, D. L. WOOD, AND S. J. GAST. 1990. Phylogenetic relationships of *Ips* bark beetles (Coleoptera: Scolytidae): Electrophoretic and morphometric analysis of the *grandicollis* group. Biochem. Syst. Ecol. 19:359–368.

CHO, S., A. MITCHELL, J. C. REGIER, C. MITTER, R. W. POOLE, T. P. FRIEDLANDER, AND S. ZHAO. 1995. A highly conserved nuclear gene for low-level phylogenetics: *Elongation factor-1α* recovers morphology-based tree for Heliothine moths. Mol. Biol. Evol. 12:650–656.

COGNATO, A. I. 2000. Phylogenetic analysis reveals new genus of Ipini bark beetle (Scolytidae). Ann. Entomol. Soc. Am. 93:362–366.

COGNATO, A. I., S. O. ROGERS, AND S. A. TEALE. 1995. Species diagnosis and phylogeny of the *Ips grandicollis* group (Coleoptera: Scolytidae) using Random Amplified Polymorphic DNA. Ann. Entomol. Soc. Am. 88:397–405.

COGNATO, A. I., S. SEYBOLD, D. L. WOOD, AND S. A. TEALE. 1997. A cladistic analysis of pheromone evolution in *Ips* bark beetles (Coleoptera: Scolytidae). Evolution 51:313–318.

COGNATO, A. I., AND F. A. H. SPERLING. 2000. Phylogeny of *Ips* DeGeer species inferred from mitochondrial cytochrome oxidase I sequence. Mol. Phylogenet. Evol. 14:445–460.

CUNNINGHAM, C. W. 1997. Is congruence between data partitions a reliable predictor of phylogenetic accuracy? Empirically testing an iterative procedure for choosing among phylogenetic methods. Syst. Biol. 46:464–478

DANFORTH, B. N., AND S. JI. 1998. Elongation factor-1-alpha occurs as two copies in bees: Implications for phylogenetic analysis of EF-1-alpha sequences in insects. Mol. Biol. Evol. 15:225–235.

DE PINNA, M. C. C. 1991. Concepts and tests of homology in the cladistic paradigm. Cladistics 7:367–394.

DOYLE, J. J. 1992. Gene trees and species trees: Molecular systematics as one character taxonomy. Syst. Bot. 17:144–163.

DOYLE, J. J. 1997. Trees within trees: Genes and species, molecules and morphology. Syst. Biol. 46:537–553.

FARRIS, J. S., M. KÄLLERSJÖ, A. G. KLUGE, AND C. BUTLIN. 1994. Testing significance of congruence. Cladistics 10:315–320.

FELSENSTEIN, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst. Zool. 27:401–410.

FRIEDLANDER, T. P., J. C. REGIER, AND C. MITTER. 1992. Nuclear gene sequences for higher level phylogenetic analysis: 14 promising candidates. Syst. Biol. 41:483–490.

FUNK, D. J. 1999. Molecular systematics of cytochrome oxidase I and 16S rRNA from *Neochlamisus* leaf beetles and the importance of sampling. Mol. Biol. Evol. 16:67–82.

FUNK, D. J., D. J. FUTUYMA, G. ORTI, AND A. MEYER. 1995. Mitochrondial DNA sequences and multiple data sets: A phylogenetic study of phytophagous beetles (Chrysomelidae: *Ophraella*). Mol. Biol. Evol. 12:627–640.

FURNISS, R. L., AND V. M. CAROLIN. 1977. Western Forest Insects. Misc. Publ. No. 1339. USDA Forest Service, Washington, DC.

GATESY, J., P. O'GRADY, AND R. H. BAKER. 1999. Corroboration among data sets in simultaneous analysis: Hidden support for phylogenetic relationships among higher level Artiodactyl taxa. Cladistics 15:271–313.

GIRIBET, G., AND W. C. WHEELER. 1999. On gaps. Mol. Phylogenet. Evol. 13:132–143.

GLADSTEIN, D. S., AND W. C. WHEELER. 1997. POY: The optimization of alignment characters. Program and documentation. New York, NY. Available from the authors at ftp.amnh.org/pub/molecular.

GOLDMAN, N. 1990. Maximum likelihood inference of phylogenetic trees, with special reference to a Poisson process model of DNA substitution and to parsimony analyses. Syst. Zool. 39:345–361.

HARRISON, R. G., ed. 1993. Hybrid zones and the evolutionary process. Oxford Univ. Press, New York.

HOPPING, G. R. 1963. The North American species in groups II and III of *Ips* DeGeer (Coleoptera: Scolytidae). Can. Entomol. 95:1202–1210.

HOPPING, G. R. 1964. The North American species in group IV and V of *Ips* DeGeer (Coleoptera: Scolytidae). Can. Entomol. 96:970–978.

HOVEMANN, B., S. RICHTER, U. WALLDORF, AND C. CZIEPLUCH. 1988. Two genes related to cytoplasmic elongation factors 1α (EF-1α) in *Drosophila melanogaster* with continuous and stage specific expression. Nucleic Acids Res. 16:3175–3194.

HUELSENBECK, J. P., J. J. BULL, AND C. W. CUNNINGHAM. 1996. Combining data in phylogenetic analysis. TREE 11:152–158.

INTERNATIONAL COMMISSION ON ZOOLOGICAL NOMENCLATURE (ICZN). 1999. International code of zoological nomenclature. International Trust for Zoological Nomenclature, London.

JUKES, T. H., AND C. R. CANTOR. 1969. Evolution of protein molecules. Pages 21–132. *in* Mammalian protein metabolism (H. N. Munro, ed.). Academic Press, New York.

KNIZEK, M., AND P. ZAHRADNIK. 1998. To the distribution of *Ips duplicatus* (Sahlberg) (Coleoptera: Scolytidae) in the Czech Republic. Pages 332–333 *in* Book of abstracts, VIth European Congress of Entomology, Ceske Budejovice, August 23–29 (V. Brunnhofer and T. Soldan, eds.). University of South Bohemia, Ceske Budejovice, Czech Republic.

LANIER, G. N. 1972. Biosystematics of the genus *Ips* (Coleoptera: Scolytidae) in North American: Hopping's groups IV and X. Can. Entomol. 104:361–388.

LANIER, G. N. 1987. The validity of *Ips cribricollis* (Eichhoff) as distinct from *I. grandicollis* (Eichhoff) and occurrence of both species in Central America. Can. Entomol. 119:179–187.

LANIER, G. N., AND D. L. WOOD. 1975. Specificity of response to pheromones in the genus *Ips* (Coleoptera: Scolytidae). J. Chem. Ecol. 1:9–23.

LEWIS, E. E., AND J. H. CANE. 1990. Pheromonal specificity of southeastern *Ips* pine bark beetles reflects phylogenetic divergence (Coleoptera: Scolytidae). Can. Entomol. 122:1235–1238.

MADDISON, W. P. AND D. R. MADDISON. 1992. *MacClade: Analysis of Phylogeny and Character Evolution, Version 3.0.* Sinaver Associates, Sunderland, Massachusetts.

MADDISON, W. P. 1997. Gene trees in species trees. Syst. Biol. 46:523–536.

MARDULYN, P., M. C. MILINKOVITCH, AND J. M. PASTEELS. 1997. Phylogenetic analyses of DNA and allozyme data suggest that *Gonioctena* leaf beetles (Coleoptera: Chrysomelidae) experienced convergent evolution in their history of host-plant family shifts. Syst. Biol. 46:699–725.

MARDULYN, P. AND J. B. WHITFIELD. 1999. Phylogenetic signal in the COI, 165, and 285 genes for inferring relationships among genera of microgastrinae (Hymenoptera; Braconidae): Evidence of a high diversification rate in this group of parasitoids Mol. Phylogenet. Evol. 12:282–294.

MICKEVICH, M. F., AND J. S. FARRIS. 1981. The implications of establishing congruence in Menidia. Syst. Zool. 30:351–370.

MIYAMOTO, M. M., AND W. M. FITCH. 1995. Testing species phylogenies and phylogenetic methods with congruence. Syst. Biol. 44:64–76.

OLMSTEAD, R., P. A. REEVES, AND A. C. YEN. 1998. Patterns of sequence evolution and implications for parsimony analysis of chloroplast DNA. Pages 164–187 *in* Molecular systematics of plants II. DNA sequencing (D. E. Soltis, P. S. Soltis, and J. J. Doyle, eds.). Kluwer, Boston.

OLMSTEAD, R., AND J. SWEERE. 1994. Combining data in phylogenetic systematics: An empirical approach using three molecular data sets in Solenaceae. Syst. Biol. 43:467–481.

PAGE, M., L. J. NELSON, G. J. BLOMQUIST, AND S. J. SEYBOLD. 1997. Cuticular hydrocarbons as chemotaxonomic characters of pine engraver beetles (*Ips* spp.) in the *grandicollis* subgeneric group. J. Chem. Ecol. 23:1053–1099.

PAMILO, P., AND M. NEI. 1988. Relationships between gene trees and species trees. Mol. Biol. Evol. 5:568–583.

PATTERSON, C. 1988. Homology in classical and molecular biology. Mol. Biol. Evol. 5:603–625.

PHILLIPS, A., D. JANIES, AND W. C. WHEELER. 2000. Multiple sequence alignment in phylogenetic analysis. Mol. Phylogenet. Evol. 16:317–330.

RAFFA, K. F., AND K. D. KLEPZIG. 1989. Chiral escape of bark beetles from predators responding to bark beetle pheromone. Oecologia 80:566–569.

REED, R. D., AND F. A. H. SPERLING. 1999. Interaction of process partitions in phylogenetic analysis: An example from the swallowtail butterfly genus *Papilio*. Mol. Biol. Evol. 16:286–297.

RIBERA, I., T. G. BARRACLOUGH, AND A. P. VOGLER. 2001. The effect of range size on clade evolution in lotic and lentic lineages of diving beetles (genus *Ilybius* and *Hydroporus*). Mol. Ecol. 10:721–735.

SEYBOLD, S. J. 1992. The role in the olfactory-directed aggregation behavior of pine engraver beetles in the genus *Ips* (Coleoptera: Scolytidae). Ph.D. Dissertation Univ. of California, Berkeley.

SEYBOLD, S. J., T. OHTSUKA, D. L. WOOD, AND I. KUBO. 1995. The enantiomeric composition of ipsdienol: A chemotaxonomic character for North American populations of *Ips* spp. in the *pini* subgeneric group (Coleoptera: Scolytidae). J. Chem. Ecol. 21:995–1016.

SHULL, V. L., A. P. VOGLER, M. D. BAKER, D. R. MADDISON, AND P. M. HAMMOND. 2001. Basal relationships in adephagan beetles inferred from 18S ribosomal RNA sequences: Evidence for a monophyletic Hydradephaga. Syst. Biol. (in press).

SIMON, C., F. FRATI, A. BECKENBACH, B. CRESPI, H. LIU, AND P. FLOOK. 1994. Evolution, weighting, and phylogenetic utility of mitochrondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Ann. Entomol. Soc. Am. 87:651–701.

SLOWINSKI, J. B., AND R. PAGE. 1999. How should species phylogenies be inferred from sequence data? Syst. Biol. 48:814–825

SOKAL, R. R., AND F. J. ROHLF. 1995 . Biometry: The principles and practice of statistics in biological research, 3rd edition. W.H. Freeman & Co., New York.

SORENSEN, M. D. 1996. TreeRot. Univ. of Michigan, Ann Arbor.

SOTA, T., AND A. P. VOGLER. 2001. Incongruence of mitochondrial and nuclear gene trees in the carabid beetles *Ohomopterus*. Syst. Biol. 50:39–59.

STAUFFER, C., F. LAKATOS, AND G. M. HEWITT. 1997. The phylogenetic relationships of seven European *Ips* (Scolytidae, Ipinae) species. Insect Mol. Biol. 6:1–8.

SWOFFORD, D. L. 1998. PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods), 4.0b3a. Sinauer Associates, Sunderland, Massachusetts.

SWOFFORD, D. L., G. J. OLSEN, P. J. WADDELL, AND D. M. HILLIS. 1996. Phylogenetic inference. Pages 407–514 *in* Molecular systematics (D. M. Hillis, C. Moritz, and B. K. Mable, eds.). Sinauer Associates, Sunderland, MA.

TAMURA, K., AND M. NEI. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. 10:512–526.

THOMPSON, J. D., D. HIGGINS, AND T. J. GIBSON. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties, and weight matrix choice. Nucleic Acids Res. 22:4673–4680.

VITÉ, J. P., A. BAKKE, AND J. A. A. RENWICK. 1972. Pheromones in *Ips* (Coleoptera: Scolytidae): Occurrence and production. Can. Entomol. 104:1967–1975.

VOGLER, A. P., AND A. WELSH. 1997. Phylogeny of North American *Cicindela* tiger beetles inferred from multiple mitochondrial DNA sequences. Mol. Phylogenet. Evol. 8:225–235.

WENZEL, J. W., AND M. E. SIDDALL. 1999. Noise. Cladistics 15:51–64.

WERNER, R. A. 1988. Recommendations for suppression of an *Ips perturbatus* outbreak in interior Alaska using integrated control. Pages 189–195 *in* Integrated control of Scolytid bark beetles (T. L. Payne and H. Saarenmaa, eds.). Virginia Polytechnic Institute and State Univ., Blacksburg.

WHEELER, W. C. 1995. Sequence alignment, parameter sensitivity, and the phylogenetic analysis of molecular data. Syst. Biol. 44:321–331.

WHEELER, W. C. 1996. Optimization alignment: The end of multiple sequence alignment in phylogenetics? Cladistics 12:1–9.

WHITFIELD, J. B., AND S. A. CAMERON. 1998. Hierarchical analysis of variation in mitochondrial 16S gene among Hymenoptera. Mol. Biol. Evol. 15:1728–1743.

WOOD, D. L. 1982. The role of pheromones, kairomones, and allomones in the host selection and colonization behavior of bark beetles. Annu. Rev. Entomol. 27:411–446.

WOOD, S. L. 1977. New synonymy and new species of American bark beetles (Coleoptera: Scolytidae), part V. Great Basin Nat. 37:383–394.

WOOD, S. L. 1982. The bark and ambrosia beetles of North America (Coleoptera: Scolytidae), a taxonomic monograph. Great Basin Nat. Mem. 6.

WOOD, S. L. 1992. Nomenclatural changes and new species of Platypodidae and Scolytidae (Coleoptera), part II. Great Basin Nat. 52:78–88.

WOOD, S. L., AND D. E. BRIGHT. 1992. A catalog of Scolytidae and Platypodidae (Coleoptera), part 2: Taxonomic index, volumes A and B. Great Basin Nat. Mem. 13.

APPENDIX—REVISION OF *Ips* CLASSIFICATION

Ips DeGeer, 1775:190. Type species: *Dermestes typographus* Linnaeus, subsequent designation by Bergroth 1884:230.

Cumatotomicus Ferrari, 1867:44. Type species: *Bostrichus stenographus* Duftschmidt. Synonymy: Hagedorn 1910:47.

Cyrtotomicus Ferrari, 1897:44. Type species: *Bostrichus acuminatus* Gyllenhal, subsequent designation by Hopkins 1914:120. Synonymy: Eichhoff 1878:256.

Ips is defined by the synapomorphies of all data sets, which support node 39 of the tree (Fig. 1; Table 1, PBS values). This genus is distinguished from other Ipini genera by the following characters: antennal club flattened, oval, anterior face of club with either straight, weakly to strongly bisinuate, or acutely angulate sutures; apical margin of elytral declivity strongly explanate; second and third spines on declivital margin connected by a low tumescence; declivital face without spines; seminal trough of male genitalia consist of either two compressed rods crossing each other or two parallel rods held close together ending in a point or broadly.

Thirty-two species examined in this study are included within *Ips* (Fig. 1): *I. acuminatus* (Gyllenhal), *I. amitinus* (Eichhoff), *I. apache* Lanier, *I. avulsus* (Eichhoff), *I. bonanseai* (Hopkins), *I. borealis* Swaine, *I. calligraphus* (Germar), *I. cembrae* (Heer), *I. confusus* (LeConte), *I. cribricollis* (Eichhoff), *I. duplicatus* (Sahlberg), *I. emarginatus* (LeConte), *I. grandicollis* (Eichhoff), *I. hoppingi* Lanier, *I. hunteri* Swaine, *I. integer* (Eichhoff), *I. knausi* Swaine, *I. lecontei* Swaine, *I. longifolia* (Stebbing), *I. montanus* (Eichhoff), *I. paraconfusus* Lanier, *I. perroti* Swaine, *I. perturbatus* (Eichhoff), *I. pilifrons* Swaine, *I. pini* (Say), *I. plastographus* (LeConte), *I. schmutzenhoferi* Holzschuh, *I. sexdentatus* (Boerner), *I. stebbingi* Strohmeyer, *I. tridens* (Mannerheim), *I. typographus* (Linnaeus), and *I. woodi* Thatcher. The following species are included in *Ips* based on morphology: *I. chinensis* Kurenzov and Kononov,

I. hauseri Reitter (COI nucleotide sequence also confirms monophyly with *Ips*; Cognato and Sperling, 2000), *I. nitidus* Eggers, and *I. ussuriensis* Reitter.

Ips (Emarips) *Cognato; new subgenus*

Type species: *Tomicus emarginatus* LeConte, 1876.
Emarips is defined by the synapomorphies of non-molecular, EF-1α and 16S data, which support node 19 of the tree (Fig. 1; Table 1, PBS values). The species are distinguished from the other *Ips* species by large size (>5 mm), presence of spines on both first and second interstrial spaces on the margin of the elytral declivity, shape of seminal trough (two straight rods held close together, ending in a broad tip), and the large internal sac of the male genitalia.

This subgenus comprises *I. emarginatus, I. knausi,* and *I. sexdentatus*, and this group is basal to the other *Ips* species. The sister species *I. (E.) emarginatus* and *I. (E.) knausi* have been recognized as a group in past systematic studies (Hopping, 1963; S. L. Wood, 1982), and their monophyly is well supported by all data sets (node 9, Table 1). However, *I. (E.) sexdentatus* was not considered to be related to these species presumably because of the very different spine morphology of the elytral declivity.

Ips acuminatus shares pleisomorphic morphological characters with *Ips (Emarips)* but shares molecular synapomorphies with the remaining *Ips* species. All molecular data support node 38, which includes *I. acuminatus* and the remaining *Ips* species.

Ips (Bonips) *Cognato; new subgenus*

Type species: *Tomicus bonanseai* Hopkins, 1905.
Bonips is defined by the synapomorphies of non-molecular, EF-1α, and 16S data that support node 31 of the tree (Fig. 1; Table 1, PBS values). The species are distinguished from the other *Ips* species by the combination of the following characters: seminal trough of the male genitalia consisting of two parallel rods held close together ending in a point, presence of one spine after the third spine on the margin of the elytral declivity margin, and the absence of discal interstrial punctures on the elytra.

This subgenus contains *I. avulsus, I. bonanseai, I. integer, I. pini,* and *I. plastographus*. Sister species *I. plastographus* and *I. integer* (node 8) are distinguishable from other *Ips (Bonips)* species by their larger size (>4.0 mm), acutely angulate antennal sutures, and presence of a median epistomal tubercle. *Ips duplicatus* is sister to the remaining species; its relationship to the *Ips (Bonips)* species has been debated (Hopping, 1964; Lanier, 1972). Including this species with *Ips (Bonips)* would introduce morphological homoplasy for the following characters in *I. duplicatus*: discal interstrial punctures on the elytra, shape of seminal trough (two compressed rods crossing each other), and absence of a stridulatory apparatus. This homoplasy would compromise the diagnosis of *Ips (Bonips)* species. Because *Ips (Bonips)* is monophyletic in a parsimony analysis without *I. duplicatus, I. duplicatus* is not considered part of *Ips (Bonips)*.

In addition, Figure 1 shows *I. (B.) bonanseai* as basal to *I. (B.) pini* and *I. (B.) avulsus*, contrary to morphological similarity between *I. (B.) bonanseai* and *I. (B.) pini* (Lanier, 1972). Monophyly of the former group is only weakly supported (Table 1); hence additional phylogenetically informative characters will be required to test the ambiguous positions of *I. (B.) bonanseai* and *I. duplicatus*.

These species breed and feed on *Pinus* species and occur throughout North America and Mexico.

Ips (Granips) *Cognato; new subgenus*

Type species: *Tomicus grandicollis* Eichhoff, 1868.

Granips is defined by the synapomorphies of all data sets that support node 33 of the tree (Fig. 1; Table 1, PBS values). The presence of two or three spines after the third spine on the margin of the elytral declivity and the seminal trough of two compressed rods crossing each other, taken together, distinguish the species of this subgenus from other *Ips*.

This subgenus contains *I. apache, I. calligraphus, I. confusus, I. cribricollis, I. grandicollis, I. hoppingi, I. lecontei, I. montanus*, and *I. paraconfusus*. Sister species *I. (G.) apache* and *I. (G.) calligraphus* (node 6) are diagnosable from other *Ips (Granips)* species by the presence of three spines after the third spine on the elytral declivity margin and by punctures with one or more setae on the discal interstriae. *Ips (G.) montanus, I. (G.) paraconfusus, I. (G.) confusus*, and *I. (G.) hoppingi* (node 26) are differentiated from the other *Ips (Granips)* species by the placement of first spine on the second interstrial space on the elytral declivity margin.

Ips (Ips) *DeGeer; new subgenus*

Type species: *Dermestes typographus* Linnaeus, subsequent designation by Bergroth, 1884.

Ips (s. str.) is defined by the synapomorphies of EF-1α and 16S data that support node 35 of the tree (Fig. 1; Table 1, PBS values). Collectively, the punctures with one or more setae on the discal interstriae and presence of one spine after the third spine on the elytral declivity margin distinguish this subgenus from all *Ips* species except *I. duplicatus*.

This subgenus contains *I. amitinus, I. borealis, I. cembrae, I. hunteri, I. longifolia, I. perroti, I. perturbatus, I. pilifrons, I. schmutzenhoferi, I. stebbingi, I. tridens, I. typographus*, and *I. woodi*. These species tend to have a geographic distribution of high latitude, high altitudes, or both. The clade (node 24) that contains *I. (I.) perturbatus, I. (I.) hunteri, I. (I.) borealis, I. (I.) tridens*, and *I. (I.) pilifrons* is well supported by molecular data (Table 1). All species primarily breed and feed in North American *Picea*. The relationships of some of these North American spruce *Ips* are in question, given the observed polyphyly among populations for some species (Cognato and Sperling, 2000). One well-supported clade (node 14) includes species inhabiting the Himalayas.