ACADEMIC PRESS

Sources of character conflict in a clade of water striders (Heteroptera: Gerridae)

Jakob Damgaard[a,*] and Anthony I. Cognato[b]

[a] *Department of Evolutionary Biology, Zoological Institute and Department of Entomology, Zoological Museum, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen, Denmark*
[b] *Department of Entomology, Texas A&M University, College Station, TX 77843-2475, USA*

Abstract

Incongruence among trees reconstructed with different data may stem from historical (gene tree-species tree conflict) or process (character change biases) phenomena. Regardless of the source, incongruent data, as determined with "global" measures of homoplasy, have often been excluded from parsimony analysis of the combined data. Recent studies suggest that these homoplasy measures do not predict the contribution of each character to overall tree structure. Branch support measures identify, on a character to node basis, sources of support and conflict resulting from a simultaneous analysis of the data. We implement these branch support measures to identify sources of character conflict in a clade of water striders consisting of *Gerris* Fabricius, *Aquarius* Schellenberg, and *Limnoporus* Stål species. Separate analyses of morphology, mitochondrial cytochrome oxidase I (COI), large mitochondrial ribosomal subunit (16SrRNA), and elongation factor-1α (EF-1α) data resulted in cladograms that varied in resolution and topological concordance. Simultaneous analysis of the data resulted in two trees that were unresolved for one node in a strict consensus. The topology agreed with current classification except for the placements of *Aquarius chilensis* and the *Aquarius remigis* species group closer to *Gerris* than to congeneric species. Branch support measures indicated that support derived from each data set varied among nodes, but COI had an overall negative effect on branch support. However, Spearman rank correlation of partitioned branch support values indicated no negative associations of branch support between any data sets and a positive association between EF-1α and 16SrRNA. Thus incongruence among data sets was not drastic and the gene-tree versus species tree phenomenon was not implicated. Biases in character change were a more likely reason for incongruence, although saturation curves and incongruence length difference for COI indicated little potential for homoplasy. However, a posteriori inspection of COI nucleotide change with reference to the simultaneous analysis tree revealed AT and codon biases. These biases were not associated with branch support measures. Therefore, it is difficult to predict incongruence or identify its cause. Exclusion of data is ill advised because every character is potentially parsimony informative.

© 2003 The Willi Hennig Society. Published by Elsevier Inc. All rights reserved.

Keywords: Molecular systematics; Insects; *Aquarius*; *Gerris*; *Limnoporus*; COI; 16SrRNA; EF-1α; Combined analysis; Hidden branch support

Cladistic analysis that incorporates characters from multiple presumably independent sources often results in well resolved trees, increased branch support and shorter analysis time as compared to separate analyses of the data partitions (e.g., Baker et al., 2001; Cognato and Vogler, 2001; Soltis et al., 1998). Even though these benefits and the philosophical underpinnings for simultaneous analysis are well documented (e.g., Baker and DeSalle, 1997; Brower et al., 1996; Kluge, 1989; Nixon

and Carpenter, 1996; Wenzel and Siddall, 1999), incongruence occurs among data sets. Critics of combined analysis cite incongruence as reason for separate analysis (e.g., Bull et al., 1993; Miyamoto and Fitch, 1995; Swofford, 1991), although their criticism is commonly based on incongruence measures of total tree length (Templeton, 1983; incongruence length difference test (ILD), Farris et al., 1995; Partitioned homogeneity test, Swofford et al., 1996). These methods may help uncover the disparate effects of gene histories (Doyle, 1992, 1997; Pamilo and Nei, 1988; Sota and Vogler, 2001). However, the statistical significance of these measures has

been questioned (Baker et al., 1998; Dolphin et al., 2000; Yoder et al., 2001) and they may overestimate the incongruence caused by biases in DNA sequence evolution (Baker et al., 2001; Cognato and Vogler, 2001; Yoder et al., 2001). Incongruence does not occur for all clades within a tree nor with the same magnitude. Thus, summarization of conflict in a single statistic can mislead decisions of data analysis (Yoder et al., 2001).

Effects of a combined data analysis may be more accurately reflected in nodal (branch) support measures. Branch support (BS) (Bremer, 1994) and its elaborations, partitioned branch support (PBS), and hidden partitioned branch support (HPBS) (Baker and DeSalle, 1997; Gatesy et al., 1999) allow for the scrutiny of support or conflict for each character partition for each clade. These measures are calculated from comparisons of tree lengths between the globally shortest tree and the shortest trees constrained not to contain the focal node (BS). The tree lengths of separate data partitions (PBS) determine the contribution of each data set. Moreover, comparison of PBS values to the tree lengths of the separate analysis of each partition reveals the support or conflict that results from the simultaneous analysis (HPBS). Branch support is widely used in conjunction with, or as an alternative to, statistical measures of node support. The extent of its use has been questioned because raw BS values are not necessarily comparable among data sets or nodes of the same cladogram (Debry, 2001). This problem is remedied by standardization of the raw values with their division by tree length (Baker et al., 1998). However, BS values are mostly compared among data sets within a study to identify sources of character support. Potentially, HPBS can be used as a criterion for the choice of analysis variables and for the phylogenetic utility of characters. The first has been addressed for the choice of sequence alignment variables (Cognato and Vogler, 2001). The highest total HPBS was considered a robust assessment of nucleotide alignment and an indication of overall positive data interaction. The latter was explored for third codon positions of the mitochondrial cytochrome oxidase subunit I (COI) (Cognato and Vogler, 2001). In combined analysis, these characters contributed support for nodes regardless of phylogenetic rank. The generality of these findings is unknown because studies that apply HPBS values to the above questions are limited.

This study uses HPBS to identify sources of data conflict among morphological and molecular data sets for a clade of water striders (Heteroptera; Gerridae). The clade comprises three genera of mainly northern temperate species: *Gerris* Fabricius, *Aquarius* Schellenberg, and *Limnoporus* Stål. All three genera were treated as subgenera of *Gerris* (Hungerford and Matsuda, 1960; Matsuda, 1960) until recently, when their generic ranks were reinstated and species were assigned to presumed monophyletic species groups (Andersen, 1975, 1990,

1993; Andersen and Spence, 1992). Recent studies, including the use of DNA sequence data, have largely confirmed these species groups, but have also questioned their relationships to one another (Damgaard et al., 2000; Damgaard and Sperling, 2001; Sperling et al., 1997). In addition, the relationships among the three genera remain controversial (Andersen, 1993, 1995; Gallant and Fairbairn, 1996).

Materials and methods

We sampled 49 of 65 species of *Limnoporus*, *Aquarius*, and *Gerris*, including all species and subspecies of *Limnoporus* and *Aquarius* (Sperling et al., 1997; Damgaard et al., 2000), and representatives of all subgenera and monophyletic species groups of *Gerris* according to Andersen (1993) and Damgaard and Sperling (2001). For outgroup comparison we included *Gigantometra gigas* (China) which is the closest living relative to the three genera (Andersen, 1995; Matsuda, 1960). Table 1 lists species data and collection information.

Morphological characters

We used the 63 morphological characters from Andersen (1993) and Damgaard and Sperling (2001), and added characters 5, 30, and 32 from Andersen (1990) as our Nos. 64–66 following Damgaard et al. (2000) (see Andersen, 1990, 1993 for morphological characters and their states). Some characters (20, 25, 30, 32, 42, and 63) were defined conditionally and could only be scored if certain structures were present. These characters were scored as inapplicable (denoted by a question mark "?"). All multistate characters were treated as non-additive (states unordered). The morphological data matrix and character state definitions are available online as an electronic appendix (http://hisl.tamu.edu; http://www.cladistics.org; http://www.treebase.org).

DNA sequences and protocols

Most mitochondrial COI and elongation factor-1α (EF-1α) sequences were obtained from earlier studies (Damgaard et al., 2000; Damgaard and Sperling, 2001; Sperling et al., 1997). We added 426 bp of the large mitochondrial ribosomal subunit (16SrRNA) gene for all included taxa. This portion of the gene was amplified and sequenced using PCR-primers LR-J-13417 (5′-CGCCTG TTTAACAAAAACAT-3′) (Simon et al., 1994) and LR-J-12961 (5′-TTTAATCCAACATCGAGG-3′) (Cognato and Vogler, 2001). Conditions for PCR followed Cognato and Vogler (2001), and cycle sequencing protocols followed Damgaard et al. (2000).

Due to access to new specimens, we generated new COI and EF-1α sequences for *A. adelaidis*, *A. amplus*,

Table 1
Species, geographic locality, and GenBank Accession Nos

Taxon	Locality	Legit	EF1-α	COI	16SRNA
Gigantometra Hungerford and Matsuda					
G. gigas (China, 1925)	Viet Nam, Che Linh NE Hanoi	L. Rowe	AF200280	AY425235	AY425183
Limnoporus Stål					
L. canaliculatus (Say, 1832)	Canada/Ontario, Chaffey Locks	J.R. Spence	AY425262	AY425251	AY425208
L. esakii (Miyamoto, 1958)	Japan, Honshu	J.R. Spence	AY425266	U83341	AY425216
rufoscutellatus group (Andersen and Spence, 1992)					
L. dissortis (Drake & Harris, 1930)	Canada/Ottawa	J.R. Spence	AY425260	U83336	AY425228
L. genitalis (Miyamoto, 1958)	Japan, Hokkaido	J.R. Spence	AY425261	U83339	AY425217
L. notabilis (Drake & Hottes, 1925)	Canada/Vancouver	J.R. Spence	AY425263	U83334	AY425210
L. rufoscutellatus (Latreille, 1807)	Finland, lab culture	J.R. Spence	AF200268	U83337	AY425212
Aquarius Schellenberg					
najas group (Andersen, 1990; comb. nov.)					
A. cinereus (Puton, 1869)	Morocco, Azrou/Ifrane Area	ZMUC Exp.	AF200279	AF200248	AY425234
A. najas (De Geer, 1773)	Denmark, Zealand, Lellinge Å	N.M. Andersen	AY425254	AF200736	AY425199
A. ventralis (Fieber, 1861)	Bulgaria, Strandzha Mts.	N. Simov	AY425255	AY425249	AY425196
conformis group (Andersen, 1990)					
A. conformis (Uhler, 1878)	USA/KY, Red River	J.J. Krupa	AF200272	AF200255	AY425186
A. elongatus (Uhler, 1896)	Japan, Kyushu, Fukuoka	K. Hayashi	AF200276	AY425250	AY425200
A. nebularis (Drake & Hottes, 1925)	USA/KY, Elkhorn River	J.J. Krupa	AY425259	AF200254	AY425185
paludum group (Andersen, 1990)					
A. adelaidis (Dohrn, 1860)	Singapore, Lower Pierce Reservoir	C.M. Yang	AY425252	AY425248	AY425202
A. antigone (Kirkaldy, 1899)	Australia, A.C.T., Murrumbidges	R. Hauser	AF200271	AF200260	AY425205
A. distanti (Horvath, 1899)	Botswana, Okawango-Delta	L. Grau	AF200265	AF200253	AY425198
A. fabricii (Andersen, 1990)	Australia, W.A., Kimberley	R. Hauser	AF200274	AF200261	AY425204
A. lili (Polhemus and Polhemus, 1994)	Indonesia, Timor, Lili River	J.T. & D.A. Polhemus	AF200270	AF200258	AY425232
A. paludum paludum (Fabricius, 1794)	Denmark, Zealand, Agersø	F.A.H. Sperling	AY425253	AF200737	AY425188
A. p. amamiensis (Miyamoto, 1958)	Japan, Amami-Oshima	K. Hayashi	AF200273	AF200256	AY425197
Genus incerta sedis I					
remigis group (Andersen 1990; comb. nov.)					
Gerris amplus (Drake & Harris, 1938)	Mexico, La Ciudad	M. Caterino	AY425258	AY425242	AY425182
G. nyctalis (Drake & Hottes, 1925)	USA/Colorado, Estes Park	J.T. Polhemus	AF200267	AY425243	AY425233
G. remigis (Say, 1832)	USA/Calif. Del Puerte Canyon	J. Damgaard	AF200277	AF200251	AY425190
Aquarius remigoides (Gallant and Fairbairn, 1996)	USA/NC, Doughton Park	D.J. Fairbairn	AF200269	AF200249	AY425230
Genus incerta sedis II					
G. chilensis (Berg, 1881)	Chile, V region, Laguna Verde	Chr. Villagra	AY425265	AY425246	AY425189

Gerris Fabricius

subgenus *Gerriselloides* (Hungerford & Matsuda, 1958)

Taxon	Location	Collector			
G. asper (Fieber, 1861)	Switzerland, Berne	J.R. Spence	AY425183	AF251101	AF251081
G. lateralis (Schummel, 1832)	Denmark, Zealand, Grib Skov	J. Damgaard	AY425231	AF251102	AF251082
subgenus *Macrogerris* (Andersen, 1993)					
G. gracilicornis (Horvath, 1879)	Japan, Kochi University Campus	T. Harada	AY425201	AF251244	AF251080
G. yezoensis (Miyamoto, 1958)	Japan, Honshu, Mutsu	M. Muraji	AY425203	AY425245	AY425257
subgenus *Gerris* (Fabricius, 1794)					
gillettei group (Andersen, 1993)					
G. gillettei (Lethierry & Severin, 1896)	USA/NM, Jemez Mts.	G. Arnqvist	AY425218	AF251105	AF251085
G. incognitus (Drake & Hottes, 1925)	USA/NM, Bosque del Apache	G. Arnqvist	AY425207	AY425247	AY425256
G. pingreensis (Drake & Hottes, 1925)	Canada, Alberta, Dunstable	J.R. Spence	AY425214	U83345	AF200275
G. sphagnetorum (Gaunitz, 1947)	Sweden, Bjädiesjö	J. Damgaard	AY425206	AF251103	AF251083
nepalensis group (Andersen, 1993)					
G. nepalensis (Distant, 1910)	Japan, Kochi	T. Harada	AY425192	AF251106	AF251086
odontogaster group (Andersen, 1993)					
G. argentatus (Schummel, 1832)	Denmark, Zealand, Holte	J.R. Spence	AY425194	AF251107	AF251087
G. babai (Miyamoto, 1958)	China, Tianjin	J.R. Spence	AY425225	AF251108	AF251090
G. buenoi (Kirkaldy, 1911)	Canada/Ottawa	J.R. Spence	AY425191	U383343	AF251089
G. odontogaster (Zetterstedt, 1828)	Denmark, Zealand, Holte	J.R. Spence	AY425219	AF251109	AY425270
argenticollis group (Andersen, 1993)					
G. argenticollis (Parshley, 1916)	USA/VA, Pr.William F.P.	J.T. Polhemus	AY425193	AF251111	AY425269
latiabdominis group neo. comb.					
G. latiabdominis (Miyamoto, 1958)	Japan, Sakurai	T. Harada	AY425215	AF251110	AF251291
thoracicus group (Andersen, 1993)					
G. costae costae (Herrich-Schäffer, 1850)	Austria, Scheibbs, Dürrenstein	Seyfert	AY425226	AY425236	AY425267
G. c. fieberi (Stichel, 1938)	Greece, Peloponnissos	V. Michelsen	AY425209	AY425237	AY425268
G. c. poissoni (Wagner & Zimmermann, 1955)	Spain, Pyrenees	V. Michelsen	AY425213	AF251113	AF251094
G. thoracicus (Schummel, 1832)	Denmark, Zealand, Tåstrup	J. Damgaard	AY425187	AF251114	AF251093
marginatus group (Andersen, 1993; neo. comb.					
G. alacris (Hussey, 1921)	USA/PA, Henry's Mill	J.R. Spence	AY425221	AF251115	AF251095
G. comatus (Drake & Hottes, 1925)	Canada/Ottawa	J.R. Spence	AY425222	U383344	AF251096
G. incurvatus (Drake & Hottes, 1925)	Canada/BC, Abbotsford	L. Rowe	AY425220	AY425240	AY425271
G. marginatus (Say, 1832)	USA/NC, Weaverville, L.Louise	G. Arnqvist	AY425211	AY425241	AY425272
swakopensis group (Andersen, 1993)					
G. swakopensis (Stål, 1858)	Kenya, Masai Mara	J.R. Spence	AY425223	AF251112	AY425264
lacustris group (Andersen, 1993)					
G. brasili (Poisson, 1941)	Spain, Murcia, Sierra Segura	G. Arnqvist	AY425229	AY425238	AY425273
G. gibbifer (Schummel, 1832)	Denmark, Bornholm, Bolshavn	J. Damgaard	AY425195	AF251117	AF251098
G. lacustris (Linnaeus, 1758)	Denmark, Zealand, Holte	ZMUC exp.	AY425227	AF200278	AF200278
G. maculatus (Tamanini, 1946)	Tunisia, Tamara		AY425224	AF251116	AY425274

A. chilensis, *A. ventralis*, *Gerris brasili*, *G. costae fieberi*, *G. incurvatus*, *G. marginatus*, and *G. yezoensis* and EF-1α sequences for *Limnoporus canaliculatus*, *dissortis*, and *notabilis* based on published primers and PCR-conditions (Damgaard et al., 2000; Damgaard and Sperling, 2001; Sperling et al., 1997). All DNA sequences were assembled in Sequencher (Gene Codes Corporation, Ann Arbor, Michigan). Length variation of the COI and EF-1α sequences was not observed and they presented no alignment ambiguity. Sequences generated for 16sRNA only varied in length by a maximum of six nucleotides, thus alignment gaps were inserted using ClustalX (Thompson et al., 1997). Suboptimal alignments have been produced by this program using the default parameters (Cognato and Vogler, 2001). However, better alignments are produced when a user defined guide tree is used (Caterino and Vogler, 2002) and an optimality criterion such as the ILD (Phillips et al., 2000) is implemented for the choice of alignment parameters. We aligned the 16SrRNA data with the gap/cost alignment value of, 2, 5, 10, 15 (default), 35, and 55. All other alignment values were left as default. Alignments were guided with a parsimoniously reconstructed tree based on combined COI, EF-1α, and morphology data. Tree reconstruction followed the cladistic analysis methodology described below. The first of 16 most parsimonious trees (tree length: 1827 steps) was chosen as the guide tree. The ILD values for each alignment were calculated following Phillips et al., 2000. Gap/cost values of 5 and 10 produced the same alignment and the lowest ILD value (0.03565). This alignment was used in all subsequent tree reconstructions. The guide tree, alignment, and character matrix are available online at http://hisl.tamu.edu; http://www.cladistics.org; http://www.treebase.org. All newly generated sequences were deposited in GenBank (Table 1).

Heterozygosity in EF-1α

As observed for water striders and other insects (Cognato and Vogler, 2001; Damgaard et al., 2000; Damgaard and Sperling, 2001; Moran et al., 1999; Reed and Sperling, 1999) we recorded base ambiguities (simultaneous presence of different nucleotides at certain positions) in EF-1α DNA electropherograms. For water striders, the ambiguous sites in EF-1α electropherograms show only two alternatives at each position. The number of such sites per specimen varies between 1 and 11 (Damgaard et al., 2000). Most base ambiguities occurred in synonymous third positions, but a few polymorphisms were found in all codon positions (Damgaard et al., 2000).

We consistently amplified only a single amplicon, and found a maximum of 2% ambiguous sites, which suggested the possibility of multiple copies of EF-1α in water striders is low. There is 18–25% nucleotide difference between EF-1α copies in *Drosophila* and *Apis* (Danforth and Ji, 1998). We therefore interpreted the base ambiguities as heterozygosity due to different alleles of the same gene copy (Reed and Sperling, 1999), and coded the sites as polymorphic for the phylogenetic analyses using the nucleotide ambiguity code implemented in Sequencher.

Cladistic analyses

Equally weighted parsimony analyses were performed for all tree searches using the program PAUP* 4b2 (Swofford, 1998) heuristic search and 100 random-taxon-addition iterations. For simultaneous analysis (SA), the morphological and molecular data sets were converted into standard character states (0–3) in MacClade (Maddison and Maddison, 1992) and analyzed using parsimony under the same conditions as above. In addition, a ratchet tree search performed with NONA (Goloboff, 1993) and the following settings, Fitch parsimony, gap = fifth character state, # iterations = 200, # trees/iteration, characters sampled = 178, recovered the same trees as the simultaneous analysis performed with PAUP*. Clade reliability was estimated using bootstrap values and branch support. Bootstrap values were generated in PAUP* from 500×10 random addition replicates. Branch support (BS), partition branch support (PBS), hidden partition branch support (HPBS), and hidden branch support (HBS) measures were calculated following published instructions and guidelines (Baker et al., 1998; Cognato and Vogler, 2001; Gatesy et al., 1999). Constraint trees for these analyses were generated with TreeRot (Sorenson, 1999).

Phylogenetic congruence was addressed globally and locally. Globally, ILD for pair-wise comparisons of the data (Phillips et al., 2000) were used. Local (at each node) congruence between each data set was measured with Spearman's rank correlations of PBS values obtained for each partition on the simultaneous analysis tree (SA) (Sota and Vogler, 2001). This metric permits the analysis of support from each data partition under separate and simultaneous analysis on a node-by-node basis and factors in both the level of agreement in tree topology and the magnitude of the signal. A positive correlation indicates congruence of support, a negative correlation indicates opposing support, and no correlation indicates that support is not associated with any pair-wise combination of data.

Potential saturation was compared with branch lengths estimated under different models of nucleotide evolution. For each data set, branch lengths were constrained to the SA tree and calculated under a model of equiprobable character change (Jukes and Cantor, 1969; see Olmstead et al., 1998 for discussion) and a complex model (Tamura and Nei, 1993) that accounts for unequal nucleotide frequencies, variation in substitution

rates, and transition bias. These values were plotted and an asymptotic curve was interpreted as an indication of saturated character change.

We also associated branch support (BS and HBPS) with node level in order to test the phylogenetic utility of each gene. A positive correlation indicates support occurs mostly at deep nodes, negative correlation indicates support occurs mostly at peripheral nodes, and no correlation indicates support is scattered among the nodes. The distance from taxa to nodes was calculated by scaling branch lengths of the SA tree with a molecular clock using maximum likelihood optimization. Thus, the branch lengths calculated for each molecular data set are comparable when graphed. Maximum likelihood estimations were performed under the following settings in PAUP*: ti/tv ratio equal, base frequencies equal, among-site rate variation equal (Cognato and Vogler, 2001).

Results

A simultaneous analysis of all data resulted in two most parsimonious trees (MPTs), one of them shown in Fig. 1. The two trees differ in the arrangement of three closely related *Gerris* species (node 9), so the zero value of all branch support measures for this node indicates that it is unresolved in a consensus of the two MPTs (Table 2). The SA tree generally agrees with current generic taxonomy and previous phylogenetic analyses; *Limnoporus* (node 50) and *Gerris* (Node 27) are monophyletic. However, *Aquarius* is paraphyletic; one clade (node 31) is the sister group to *Gerris* while another clade (node 44) is sister group to the clade that includes *Gerris* and node 31. Relationships of subgenera and species groups for the three genera are further discussed below. Separate analyses of the molecular data sets produced trees that differed in resolution and topology, but exhibited a similar pattern to the SA tree, except for generally less resolution. Only the separate analysis of morphological characters revealed all three genera as monophyletic. *Limnoporus* was monophyletic in all separate analyses. *Gerris* was monophyletic in the analysis of EF-1α alone, but paraphyletic in the analysis of COI, and polyphyletic in the analysis of 16SrRNA. *Aquarius* was paraphyletic in analyses of 16SrRNA and EF-1α and polyphyletic in the analysis of COI. However, these relationships were either poorly supported or unresolved in a strict consensus of the MPTs resulting from the separate analyses.

An overwhelming pattern of PBS for the SA tree was not observed because the contribution of each data set to total support varied. For example, 16SrRNA and COI contributed the most support (199.3 and 148.9, respectively), followed by EF-1α (142.2) and morphology (33.5) (Table 2). However, when PBS was stan-
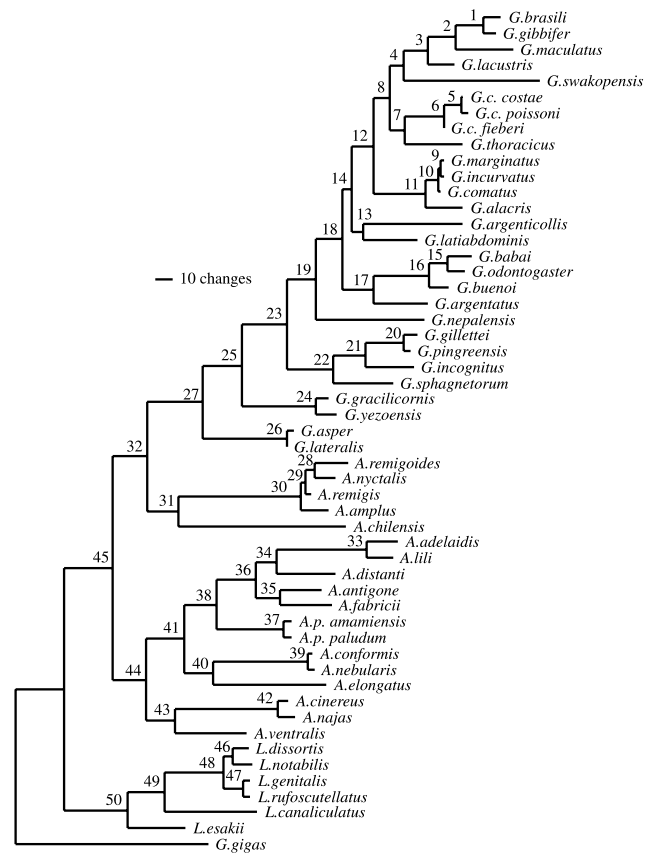

Fig. 1. A cladogram of *Limnoporus*, *Aquarius*, and *Gerris*, species rooted with *G. gigas*. One of two most parsimonious trees (length 2272) obtained from a simultaneous parsimony analysis of the 1787 characters from morphological and molecular data sets using 100 random addition replicates conducted in PAUP* (Swofford 1998). Numbers refer to clades in Table 3, which also reports branch support and bootstrap values for the different clades. Note that node 9 is unresolved in a strict consensus tree.

dardized by the number of steps in each data set that contributed to the overall signal, COI contributed the least information in the combined analysis tree (16SrRNA, 0.95; EF-1α, 0.33; morphology, 0.22; and COI, 0.10) despite the greater number of characters in this partition. In addition, half of the support was derived from the simultaneous analysis as measured by the HBS. Overall, total phylogenetic information was revealed for 16SrRNA, while generally more homoplasy was observed for COI (Table 2, HPBS). Branch support and contradiction also varied at individual nodes and only 6 nodes are supported by all data sets. No one data set provided appreciably more or less information except for morphology. Correlation of PBS values was only observed between the 16SrRNA and EF-1α data sets (Table 2). Added together, a few nodes (39, 30, and 26) exhibited values several times larger then the average PBS per node (10.5) (Table 4, total PBS). The negatively supported nodes indicated that the data interaction can also result in homoplasy (Table 2, node 26). However,

Table 2
Partitioned Branch support (PBS), branch support (BS), hidden PBS (HPBS), and hidden BS (HBS) for nodes of simultaneous analysis tree (Fig. 1) and partitioned data sets

	Bootstrap	Morphology			COI			EF-1α			16S			Total		
		PBS	BS	HPBS	PBS	BS	HPBS	PBS	BS	HPBS	PBS	BS	HPBS	PBS	BS	HPBS
Node 1	100.0	0.0	0.0	0.0	14.0	13.0	1.0	1.0	1.0	0.0	0.0	-7.0	7.0	15.0	7.0	8.0
Node 2	99.0	3.3	1.0	2.3	4.5	0.0	4.5	5.3	5.0	0.3	0.0	-7.0	7.0	13.0	-1.0	14.0
Node 3	91.0	2.0	2.0	0.0	2.0	1.0	1.0	0.0	0.0	0.0	0.0	-7.0	7.0	4.0	-4.0	8.0
Node 4	<50	1.5	-2.0	3.5	1.0	-1.0	2.0	-0.5	-1.0	0.5	0.0	-3.0	3.0	2.0	-7.0	9.0
Node 5	100.0	3.0	0.0	3.0	2.0	2.0	0.0	-1.0	0.0	-1.0	2.0	1.0	1.0	6.0	3.0	3.0
Node 6	100.0	4.0	1.0	3.0	6.0	7.0	-1.0	5.0	5.0	0.0	0.0	-4.0	4.0	15.0	9.0	6.0
Node 7	53.0	3.0	3.0	0.0	0.0	-1.0	1.0	-1.0	0.0	-1.0	0.0	-2.0	2.0	2.0	0.0	2.0
Node 8	<50	0.0	-3.0	3.0	1.0	-1.0	2.0	0.0	-2.0	2.0	0.0	-3.0	3.0	1.0	-9.0	10.0
Node 9	<50	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Node 10	100.0	3.3	0.0	3.3	2.5	2.0	0.5	-0.8	0.0	-0.8	1.0	1.0	0.0	6.0	3.0	3.0
Node 11	100.0	1.0	0.0	1.0	12.0	14.0	-2.0	9.0	3.0	6.0	2.0	0.0	2.0	24.0	17.0	7.0
Node 12	53.0	-1.0	-3.0	2.0	6.0	0.0	6.0	0.0	1.0	-1.0	-2.0	-2.0	0.0	3.0	-4.0	7.0
Node 13	56.0	-2.0	-1.0	-1.0	2.0	-5.0	7.0	1.0	0.0	1.0	3.0	1.0	2.0	4.0	-5.0	9.0
Node 14	<50	0.0	-3.0	3.0	1.0	-5.0	6.0	0.0	-1.0	1.0	1.0	-1.0	2.0	2.0	-10.0	12.0
Node 15	68.0	0.0	0.0	0.0	1.0	1.0	0.0	-1.0	-1.0	0.0	1.0	1.0	0.0	1.0	1.0	0.0
Node 16	100.0	0.0	0.0	0.0	14.0	11.0	3.0	5.0	5.0	0.0	0.0	0.0	0.0	19.0	16.0	3.0
Node 17	63.0	-1.0	1.0	-2.0	6.0	0.0	6.0	0.0	-2.0	2.0	-2.0	0.0	-2.0	3.0	-1.0	4.0
Node 18	73.0	6.0	5.0	1.0	-2.0	-5.0	3.0	0.0	-1.0	1.0	1.0	-1.0	2.0	5.0	-2.0	7.0
Node 19	85.0	3.5	2.0	1.5	0.0	-2.0	2.0	0.0	0.0	0.0	3.5	-1.0	4.5	7.0	-1.0	8.0
Node 20	100.0	0.0	0.0	0.0	9.0	5.0	4.0	4.0	4.0	0.0	4.0	3.0	1.0	17.0	12.0	5.0
Node 21	97.0	3.0	3.0	0.0	0.0	0.0	0.0	0.0	1.0	-1.0	5.0	2.0	3.0	8.0	6.0	2.0
Node 22	99.0	2.0	2.0	0.0	6.0	5.0	1.0	1.0	2.0	-1.0	2.0	-1.0	3.0	11.0	8.0	3.0
Node 23	90.0	2.5	2.0	0.5	-7.0	-2.0	-5.0	3.0	1.0	2.0	7.5	0.0	7.5	6.0	1.0	5.0
Node 24	100.0	6.0	6.0	0.0	4.0	11.0	-7.0	8.0	7.0	1.0	15.0	5.0	10.0	33.0	29.0	4.0
Node 25	56.0	1.6	3.0	-1.4	-4.3	-2.0	-2.3	2.7	-1.0	3.7	3.0	0.0	3.0	3.0	0.0	3.0
Node 26	100.0	2.7	3.0	-0.3	16.7	22.0	-5.3	11.7	21.0	-9.3	7.0	3.0	4.0	38.0	49.0	-11.0
Node 27	95.0	3.0	6.0	-3.0	-4.0	-2.0	-2.0	5.0	2.0	3.0	4.0	-1.0	5.0	8.0	5.0	3.0
Node 28	<50	-1.1	0.0	-1.1	-6.8	-1.0	-5.8	3.2	0.0	3.2	5.7	0.0	5.7	1.0	-1.0	2.0
Node 29	100.0	-1.0	0.0	-1.0	-6.2	-1.0	-5.2	3.0	1.0	2.0	5.3	0.0	5.3	1.0	0.0	1.0
Node 30	<50	2.5	4.0	-1.5	17.3	19.0	-1.8	16.0	15.0	1.0	19.3	9.0	10.3	55.0	47.0	8.0
Node 31	<50	-1.2	-5.0	3.8	-7.1	-4.0	-3.1	3.3	-8.0	11.3	6.0	1.0	5.0	1.0	-16.0	17.0
Node 32	<50	-1.2	-7.0	5.8	-7.1	-3.0	-4.1	3.3	-9.0	12.3	6.0	-1.0	7.0	1.0	-20.0	21.0
Node 33	100.0	-1.0	0.0	-1.0	16.0	19.0	-3.0	2.0	2.0	0.0	10.0	7.0	3.0	27.0	28.0	-1.0
Node 34	<50	-1.3	-2.0	0.8	-6.0	1.0	-7.0	2.8	1.0	1.8	5.5	0.0	5.5	1.0	0.0	1.0

Node																
Node 35	97.0	-3.0	0.0	-3.0	-6.0	4.0	-10.0	8.0	2.0	6.0	8.0	0.0	8.0	7.0	6.0	1.0
Node 36	67.0	-3.2	-2.0	-1.2	-3.6	2.0	-5.6	2.8	1.0	1.8	6.0	0.0	6.0	2.0	1.0	1.0
Node 37	100.0	-1.0	1.0	-2.0	13.0	22.0	-9.0	4.0	-1.0	5.0	9.0	0.0	9.0	25.0	22.0	3.0
Node 38	87.0	1.0	1.0	1.0	-12.0	-1.0	-11.0	8.0	4.0	4.0	8.0	0.0	8.0	5.0	3.0	2.0
Node 39	100.0	1.0	1.0	0.0	31.3	32.0	-0.7	6.3	5.0	1.3	5.3	3.0	2.3	44.0	41.0	3.0
Node 40	73.0	1.0	2.0	-1.0	1.0	0.0	1.0	1.0	-1.0	2.0	-1.0	-1.0	0.0	2.0	0.0	2.0
Node 41	86.0	-3.0	0.0	-3.0	-4.0	-1.0	-3.0	6.0	2.0	4.0	8.0	0.0	8.0	7.0	1.0	6.0
Node 42	100.0	-0.4	0.0	-0.4	25.4	21.0	4.4	2.9	4.0	-1.1	8.1	3.0	5.1	36.0	28.0	8.0
Node 43	<50	-1.2	1.0	-2.2	-7.1	-10.0	2.9	3.3	1.0	2.3	6.0	0.0	6.0	1.0	-8.0	9.0
Node 44	<50	-1.5	0.0	-1.5	-7.8	-2.0	-5.8	3.5	1.0	2.5	6.8	1.0	5.8	3.0	-6.0	1.0
Node 45	<50	4.5	0.0	4.5	0.0	1.0	-1.0	-5.5	-7.0	1.5	4.0	0.0	4.0	1.0	-2.0	9.0
Node 46	60.0	0.0	-2.0	2.0	1.5	1.0	0.5	-0.5	-1.0	0.5	0.0	0.0	0.0	9.0	10.0	3.0
Node 47	<50	0.0	1.0	-1.0	10.0	10.0	0.0	0.0	0.0	0.0	-1.0	-1.0	0.0	22.0	19.0	-1.0
Node 48	100.0	-1.0	1.0	-2.0	14.4	16.0	-1.6	3.0	0.0	3.0	5.6	2.0	3.6	5.0	0.0	3.0
Node 49	100.0	-4.9	-2.0	-2.9	4.3	3.0	1.3	1.9	0.0	1.9	3.7	-1.0	4.7	11.0	10.0	5.0
Node 50	82.0	2.3	5.0	-2.7	-5.0	0.0	-5.0	6.7	4.0	2.7	7.0	1.0	6.0			1.0
	99.0	**33.6**	**24.0**	**9.6**	**148.9**	**196.0**	**-47.1**	**142.2**	**65.0**	**77.2**	**199.3**	**0.0**	**199.3**	**524.0**	**285.0**	**239.0**

ILD values suggested no obvious pattern of incongruence between data sets except for higher ILD values between morphology and COI and EF-1α data (Table 3).

The absence of major taxonomic discordance among the separate analysis trees (Fig. 2) and negative correlation among branch support values (Table 3) suggests that incongruence was due to homoplasy caused by biases in character change (Table 4) and not the result of different gene histories. The source of homoplasy could lie in saturated characters, but the Tamura–Nei versus Jukes–Cantor plots of branch lengths showed a non-asymptotic relationship for all three molecular partitions and for codon positions (Fig. 3). Thus, these results suggest that the rate of nucleotide change among all sites was similar and saturation is minimal. Yet, there was a declining trend between molecular data branch support (BS) (Fig. 4) and node distance for all three genes and codon positions. In addition, there was an overall negative interaction between the COI and other data (Total HPBS $= -47.1$) which suggested this gene was incompatible with the simultaneous analysis (Table 2). These observations suggest that the data sets contained either few phylogenetically informative or many homoplastic characters for the deeper nodes. Further analysis of HPBS does not support the former hypothesis. For all three data sets, these values were not correlated with node distance, and 16SrRNA and EF-1α showed an increasing trend of HPBS (Fig. 4). Thus phylogenetically informative characters for the deeper nodes were present in the data but concealed by homoplasy.

Dissecting the data further, nucleotide composition, codon usage, and synonymous substitutions were biased and potentially contributed to total homoplasy. Mitochondrial data were AT rich, the average nucleotide composition of AT:GC was 0.63:0.37. For all data, no correlation (COI: $r = 0.3$; 16SrRNA: $r = 0.1$; and EF-1α: $r = -0.08$) was found between AT:GC ratio and HPBS at each node. However, this bias was associated with skewed codon usage in COI; only 15 codons had a relative frequency over 70% among water strider species. Of these codons, 14 contained at least two weak bounding nucleotides. Nucleotide change at first and third positions was variable among codons. For

Table 3
Pairwise analysis of congruence among data sets

	Morphology	COI	EF-1α	16S
Morphology		0.189	0.121	0.021
COI	0.176		0.015	0.01
EF-1α	−0.135	0.036		0.016
16S	−0.218	−0.193	0.611*	

Above the diagonal, ILD/steps of SA tree, Below diagonal Spearman's rank correlation of the PBS.

* $P < 0.0001$.

COI
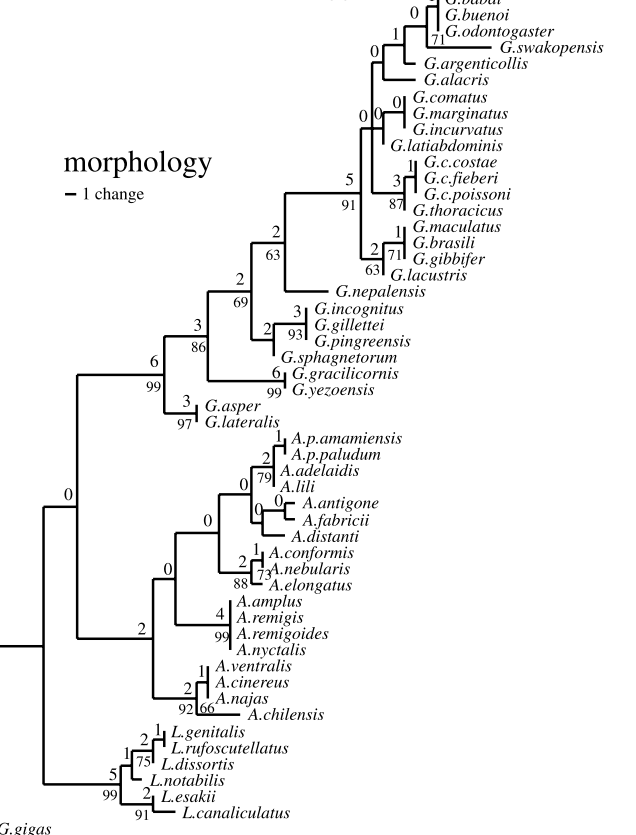— 10 changes

16S
— 1 change
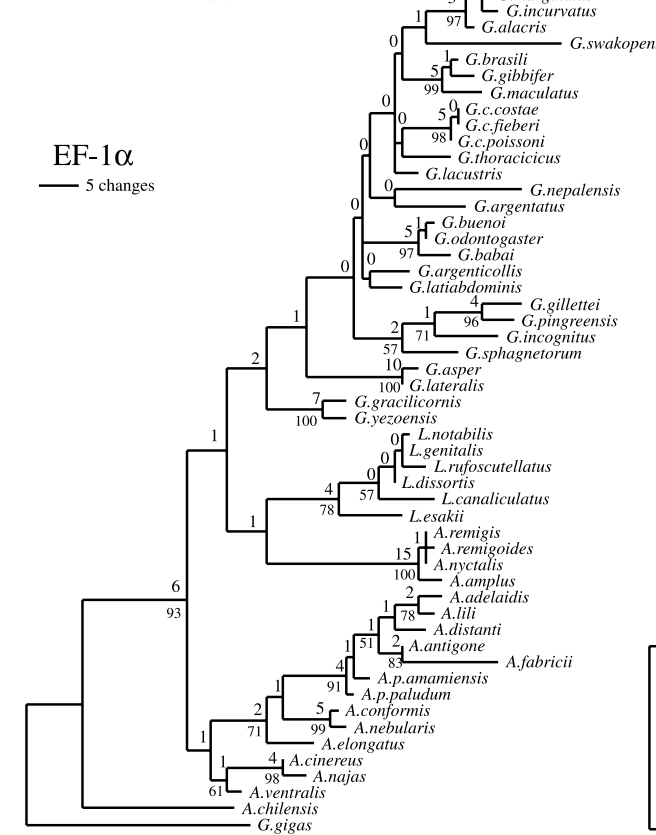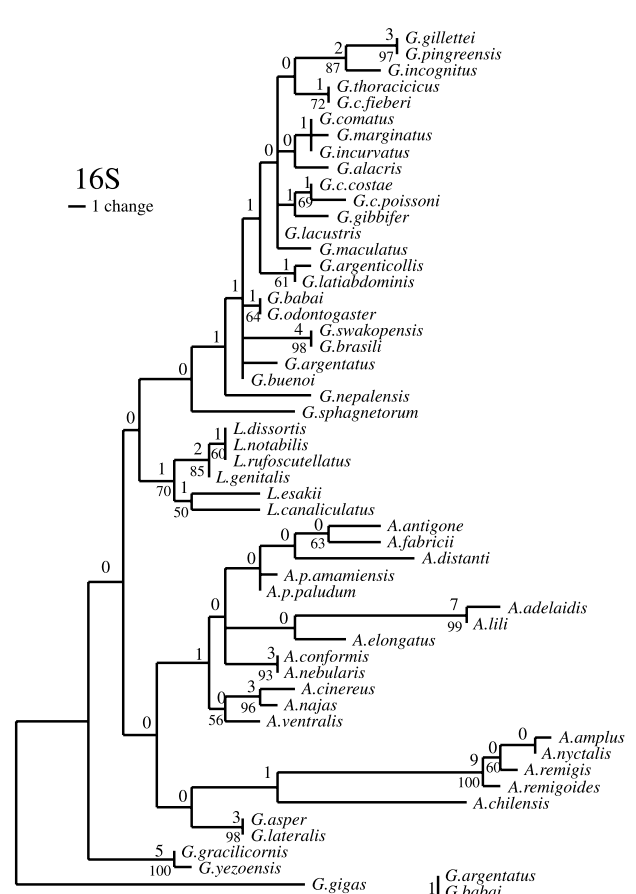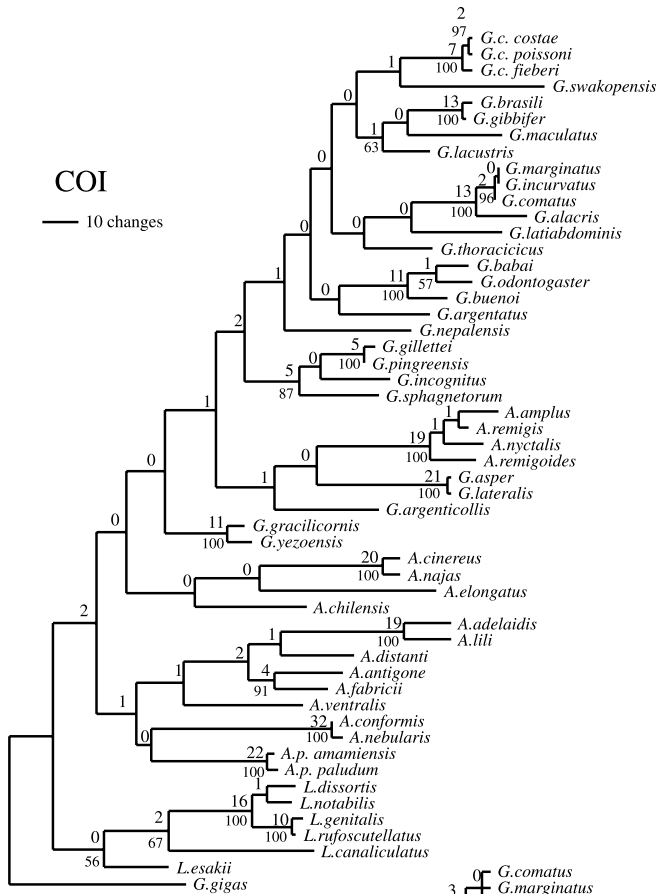
EF-1α
— 5 changes

morphology
— 1 change

Table 4
Nucleotide patterns observed for COI, 16S, and EF-1α for gerrid species

	Matrix size	Variable sites	Inform sites	Number of steps	RI	Mean ti/tv	AT/GC bias
COI							
Total	780	280	248	1407	0.6	1.77	2.56
1st codon pos.	260	41	29	132	0.7	5.94	1.84
2nd codon pos.	261	7	6	10	0.95	14	9.1
3rd codon pos.	260	222	213	1224	0.59	1.54	1.6
16SrRNA							
Total	426	96	64	197	0.78	1.57	2.96
EF-1α							
Total	515	141	94	387	0.77	3.32	1.05
1st codon pos.	172	11	5	12	0.97	1	1.42
2nd codon pos.	172	5	0	NA	NA	0.33	0.78
3rd codon pos.	171	118	89	363	0.76	3.72	1.03
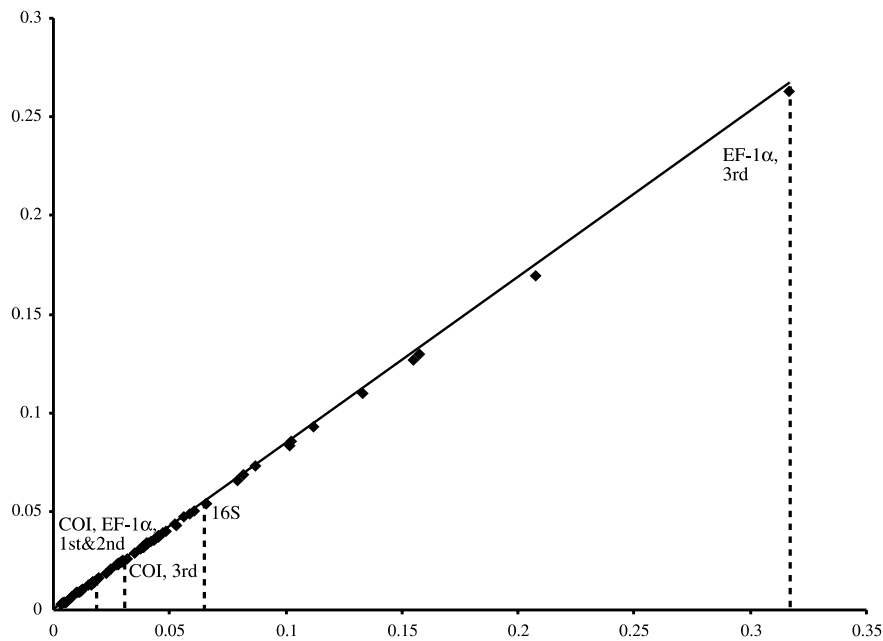
Fig. 3. Scatter plot for COI, 16SrRNA and EF-1α branch lengths constrained to the-SA tree (Fig. 1) and calculated under a Jukes and Cantor (1969) model, which assumes equiprobable character change and Tamura and Nei (1993) model, which accounts for unequal nucleotide frequencies, variation in substitution rates, and transition bias. The linear correlation is interpreted as an indication of non-saturated character change. Dotted lines indicate the maximum value of branch length for data partition.

example, third positions of TTA (Leucine) changed near Jukes–Cantor proportions (0.6 ti/tv) for synonymous (G, 13 times), and non-synonymous (C, 2 and T, 19 times) nucleotides. The non-synonymous change would have conferred an amino acid change to Phenylalanine if not for the concordant change in first position to C, in all cases. In contrast, changes at third position of TTT (Phenylalanine) were synonymous transitions, while no observable change occurred at the first position. In total,

only 49 amino acids changed and the majority was Isoleucine (ATT) to Valine (GTA) transformation via transitions in the first and third positions. Number of nucleotide changes at the COI first ($r = 0.12$) and third ($r = 0.09$) positions do not correlate with HPBS. Although our a posteriori inspection of the data suggests that these biases may confer homoplasy, we do not have any compelling reason to exclude data from the analysis. COI data does provide information for some nodes and

Fig. 2. Phylograms of gerrid species obtained from parsimony analyses of individual data sets and rooted with *G. gigas*. Bremer support values are given above branches and bootstrap support values >50% are given below. Nodes without support values are unresolved in the strict consensus of the shortest trees. COI: 1 of 48 shortest trees of length 1486 (CI = 0.29; RI = 0.6); 16SrRNA: 1 of 463 shortest trees of length 205 (CI = 0.60; RI = 0.78); EF-1α: 1 of 40 shortest trees of length 397 (CI = 0.46; RI = 0.77); and Morphology: 1 of 64 shortest trees of length 149 (CI = 0.29; RI = 0.91).
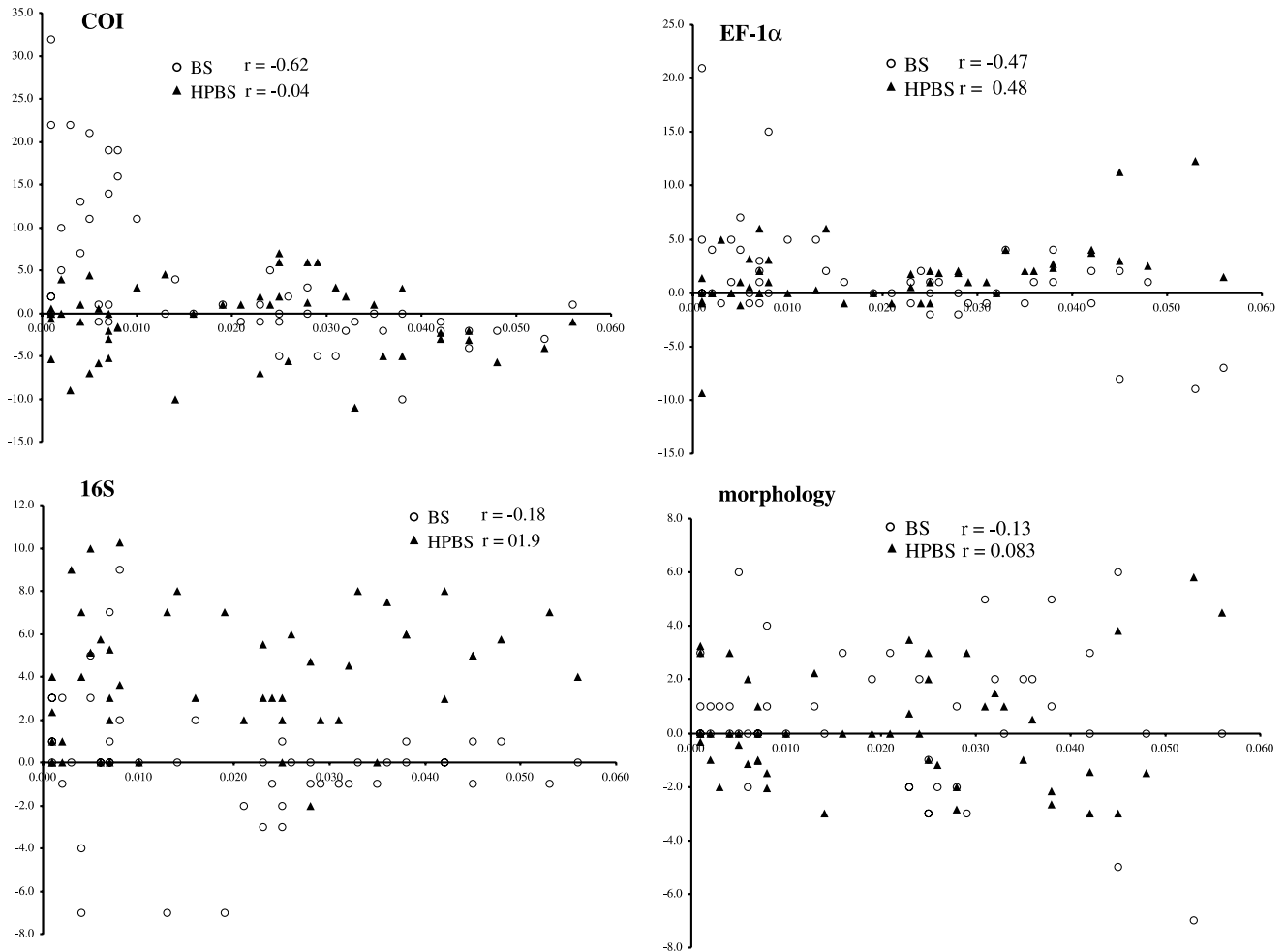
Fig. 4. Relationships between PBS values and nodal distance from tips of tree. Nodal distance is the ML branch length from a given node to the tip of the tree.

the interaction of the data reduces homoplasy at the deeper nodes as observed with the HPBS values (Figs. 4 and 5).

Discussion

Taxonomic implications

Limnoporus is monophyletic (node 50), and the SA tree generally agrees with current generic taxonomy (Andersen and Spence, 1992; Sperling et al., 1997). Although *Limnoporus dissortis* and *Limnoporus notabilis* are known to hybridize extensively in western Canada (Spence, 1990), the two species have not been previously regarded as sister species. *Gerris* is also monophyletic (node 27). Damgaard and Sperling, 2001 suggested a sister taxa relationship of the subgenera *Gerriselloides* and *Gerris* s.str., but our analysis found *Gerriselloides* (node 24) as basal subgenus and a sister group relationship between subgenera *Macrogerris* (node 24) and

Gerris s.str. (node 23) as originally proposed by Andersen (1993).

Within *Gerris* s.str., the *gillettei*-group (node 22) is basal and the monotypic *nepalensis*-group was sister group to the remaining *Gerris* species (node 19), as previously reported (Andersen, 1993; Damgaard and Sperling, 2001). This group was left highly unresolved by Damgaard and Sperling (2001), but our results indicate that the *odontogaster*-group (node 17) is at the base of the clade. Andersen (1993) placed *G. latiabdominis* in the *marginatus*-group, but we agree with Damgaard and Sperling (2001) that it should be excluded from this clade. The sister species relationship between *G. latiabdominis* and *G. argenticollis* (node 13) is poorly supported, but since these species vary drastically in morphology (Andersen, 1993) they should both be considered as monotypic species groups. Finally, in contrast to Damgaard and Sperling (2001), we found a monophyletic *thoracicus*-group (node 7), which is sister to the *lacustris* + *swakopensis*-groups (node 4). The *marginatus*-group (node 11) was at the base of this clade.
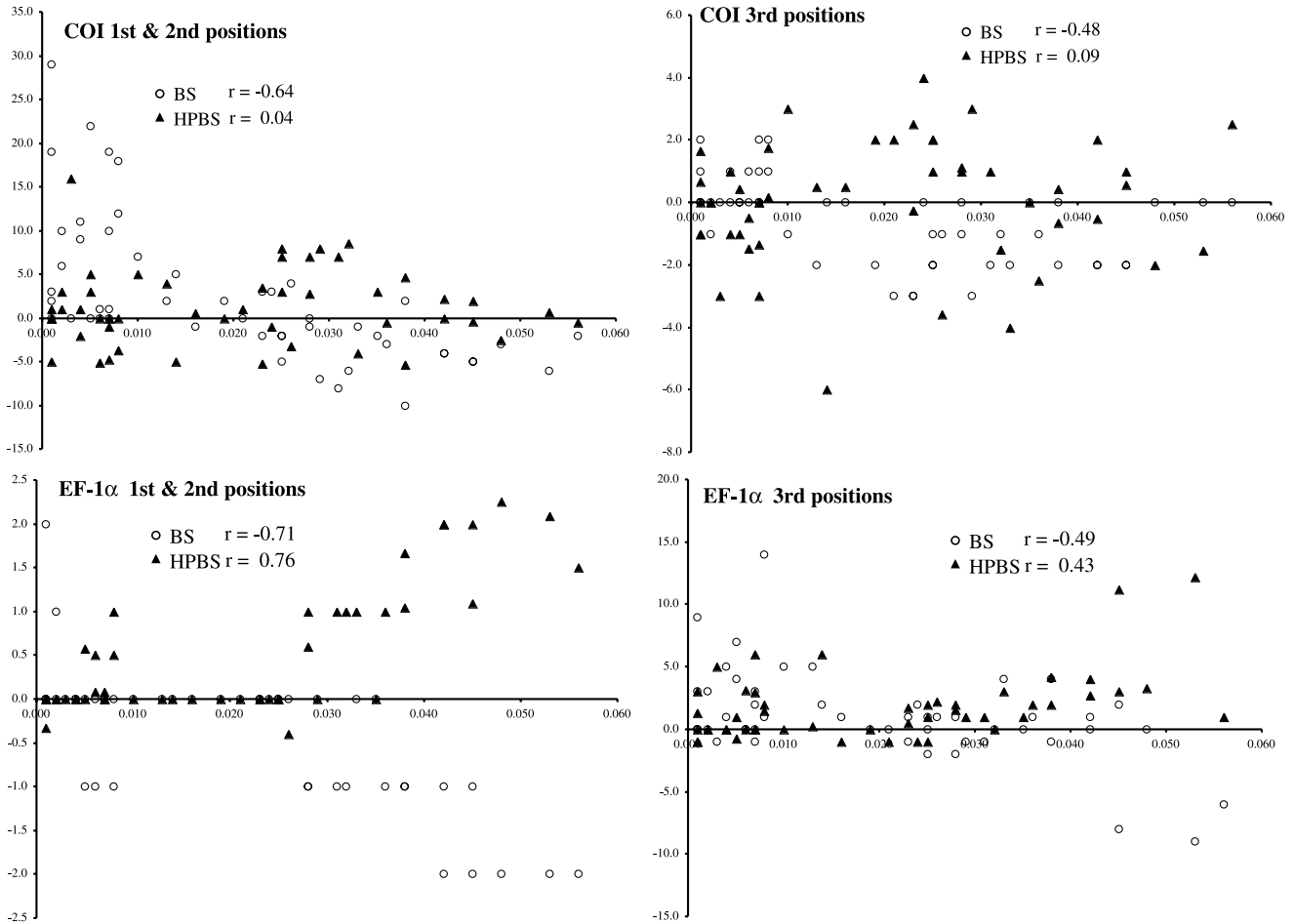
Fig. 5. Relationships between PBS values and nodal distance from tip of tree for COI and EF-1α codon positions. Nodal distance is the ML branch length from a given node to the tip of the tree.

Aquarius is paraphyletic and *A. chilensis* and the *remigis*-group (node 31) are more closely related to *Gerris* than to other members of the genus (node 44) as mentioned above. The South American *A. chilensis* was provisionally placed in the otherwise western Palaearctic *najas*-group due to remarkable similarities in morphology and ecology (Andersen, 1990: pp. 43, 74–75). Presence of pretarsal aroliae in *A. chilensis* was regarded by Andersen as a possible character reversal, because this character is shared only with more basal members of the gerrine water striders. However, our study indicates that the similarities between *A. chilensis* and other members of the *najas*-group (node 43) are homoplasies probably due to similar ecological constraints associated with life on lotic waters. Matsuda (1960, p. 179) noted that members of the *remigis*-group differ from other species of *Aquarius* due to their relatively shorter antennal segments, longer metasternum, modified male terminalia, etc. Our study supports the exclusion of the *remigis*-group from *Aquarius*. The current placement of *A. chilensis* and the *remigis*-group (Fig. 1, nodes 31, 32) is not well supported and inclusion of more basal gerrine

water striders are needed for the resolution of these taxa. The relationships among the remaining species groups of *Aquarius* (node 44) resemble the traditional taxonomy (Andersen, 1990; Damgaard et al., 2000). The *najas*-group (node 43) was basal to the sister species groups, *conformis* (node 40) and *paludum* (node 38), even though our analysis failed to reconstruct a sister species relationship between *A. paludum* and (*A. adelaidis*/*Aquarius lili*) as suggested by Polhemus and Polhemus (1994). The taxonomy of genera and species groups based on these results are summarized in Table 1. However, broader taxonomic, classification and biogeographic implications of these new finding will be treated in detail in another publication (Damgaard and Cognato, submitted).

Data interaction and patterns of homoplasy

Phylogenetic incongruence among gene trees and species trees is often cited as a potential problem that will stymie phylogenetic reconstruction (Bull et al., 1993; Maddison, 1997). Gross discordance with previous

taxonomy and/or among mitochondrial and nuclear gene trees is often the first warning of this problem. However, real examples of this type of conflict are rare (Brower et al., 1996; Sota and Vogler, 2001). Poor taxonomy and/or biases in nucleotide evolution (Baker and DeSalle, 1997) may often falsely indicate gene tree-species tree discordance. Evaluation of homoplasy patterns within a simultaneous analysis can identify the origin of incongruence (Baker et al., 2001; Cognato and Vogler, 2001; Gatesy et al., 1999). In data where true discordance between gene trees and species trees exists, correlated opposing branch support values or a significant difference in tree lengths will be observed between those data sets that reflect separate species (morphology and nuclear genes) and gene phylogenies (mitochondrial genes) (Sota and Vogler, 2001). As with *Ips* bark beetles (Cognato and Vogler, 2001), none of these patterns are observed in our data. Separate analysis trees are generally congruent with each other and previous taxonomy (Fig. 2). Significant heterogeneity between data sets was not exclusively observed between the mitochondrial and the nuclear genes or morphology data. In addition, pairwise comparisons of PBS values are not negatively correlated and a positive correlation is observed between nuclear (EF-1α) and mitochondrial (16SrRNA) genes (Table 3). Interestingly, positive correlation of branch support between the above genes has been observed for bark beetle species (Cognato and Vogler, 2001). Thus, the SA tree is a robust hypothesis of species phylogeny and observed incongruence is due to both morphological convergence and bias in nucleotide evolution especially with COI.

Surprisingly, the morphological data provides little support for the SA tree and, in fact, the simultaneous analysis reveals less support than compared to its separate analysis (PBS values, Table 2). In part, this homoplasy derives from the paraphyly of *Aquarius*. Constraints of the water surface habitat may have precluded widely divergent change of structural morphology for at least 50 million years (Andersen et al., 1993) leading to the recognition of *Aquarius* based on homoplastic characters.

Utility of gene partitions in phylogenetic reconstruction has been an ongoing debate in theoretical and empirical investigation (e.g., Friedlander et al., 1994; Graybeal, 1994; Olmstead and Sweere, 1994; Yang, 1998). Nucleotide saturation is often used as evidence for excessive homoplasy in a particular data partition and these data are either removed from subsequent phylogenetic analyses (e.g., Cunningham, 1997; Dowton and Austin, 2001) or down-weighted (e.g., Mitchell et al., 1997; Reed and Sperling, 1999). The present analysis and others (Cognato and Vogler, 2001; Olmstead and Sweere, 1994; Wenzel and Siddall, 1999) demonstrate that saturation is a poor predictor of homoplasy. For example, the gerrid COI saturation plot (Fig. 3) indicates that the data are not overwhelmed by homoplasy, yet these data have an overall negative effect on the simultaneous analysis. The opposite situation was observed for *Ips* where a COI saturation plot indicated excessive homoplasy yet the interaction of these data in simultaneous analysis was positive (Cognato and Vogler, 2001). The utility of this gene could only be determined after cladistic analyses, however, the overall negative effect of COI in the gerrid cladogram does not justify its exclusion. In addition, measures of topological concordance and total PBS have shown that phylogenetic information is present in highly divergent data (Baker et al., 2001; Björklund, 1999; Kallersjö et al., 1999). Our study confirms these observations. However, our data does not confirm a superior utility of nuclear versus mitochondrial sequence data (Baker et al., 2001). Standardized total PBS values for the data partitions show that the water strider SA tree is supported most by 16sRNA followed by EF-1α, morphology and COI. This similar pattern was observed in *Ips* where most support was derived from 16SrRNA and EF-1α, followed by COI and non-molecular data (Cognato and Vogler, 2001). We also applied HPBS values to measure the conflict or support that resulted from the simultaneous analysis at each node though few studies have addressed character information with this measure (Cognato and Vogler, 2001; Gatesy et al., 1999). As in the previous published data sets, we found varying amounts of net hidden support within the gene partitions. For particular clades, hidden support or conflict was not correlated with node or taxonomic level (Fig. 4). In addition, we failed to associate codon or nucleotide composition biases with HPBS values. Thus, the utility of a gene was unpredictable before conducting a simultaneous analysis.

In addition, "global" measures of homoplasy do not predict phylogenetic utility (Sanderson and Donoghue, 1996) or justify exclusion of data from simultaneous analysis (Yoder et al., 2001). They are summaries of the concordance among secondary homologies (De Pinna, 1991) for the parsimonious reconstruction of a given data set. Additional data and a different parsimonious solution can elevate seemingly homoplastic characters into secondary homologies. Phylogenetic utility of data partitions may only be revealed by branch support measures for the parsimonious cladogram. For example, our COI data is deemed "incompatible" based on a total negative HPBS but positive HPBS values for individual nodes are observed. Like ILD, branch support measures are data dependent, but the contribution of a data partition is revealed at the node level. Hence, character biases of particular partitions do not preclude simultaneous analyses. Some data provide more information than others, but every character is potentially informative regardless of its origin. Prediction of phylogenetic utility based on data type, saturation curves, or "global" measures of homoplasy is ill advised.

Acknowledgments

References

Andersen, N.M., 1975. The *Limnogonus* and *Neogerris* of the Old World with character analysis and a reclassification of the Gerrinae (Hemiptera: Gerridae). Entomologica Scandinavia Supplement 7, 1–96.

Andersen, N.M., 1990. Phylogeny and taxonomy of water striders, genus *Aquarius* Schellenberg (Insecta, Hemiptera, Gerridae) with a new species from Australia. Steenstrupia 16, 37–81.

Andersen, N.M., 1993. Classification, phylogeny, and zoogeography of the pond skater genus *Gerris* Fabricius (Hemiptera: Gerridae). Canadian Journal of Zoology 71, 2473–2508.

Andersen, N.M., 1995. Cladistics, historical biogeography, and a check list of gerrine water striders (Hemiptera, Gerridae) of the World. Steenstrupia 21, 93–123.

Andersen, N.M., Spence, J.R., 1992. Classification and phylogeny of the Holarctic water strider genus *Limnoporus* Stål (Hemiptera, Gerridae). Canadian Journal of Zoology 70, 753–785.

Andersen, N.M., Spence, J.R., Wilson, M.V.H., 1993. 50 million years of structural stasis in water striders (Hemiptera, Gerridae). American Entomologist 39, 174–176.

Baker, R.H., DeSalle, R., 1997. Multiple sources of character information and the phylogeny of Hawaiian drosophilids. Systematic Biology 46, 654–673.

Baker, R.H., Yu, X., DeSalle, R., 1998. Assessing the relative contribution of molecular and morphological characters in simultaneous analysis trees. Molecular Phylogenetics and Evolution 9, 427–436.

Baker, R.H., Wilkinson, G.S., DeSalle, R., 2001. Phylogenetic utility of different types of molecular data used to infer evolutionary relationships among stalked-eyed flies (Diopsidae). Systematic Biology 50, 87–105.

Björklund, M., 1999. Are third positions really that bad? A test using vertebrate cytochrome *b*. Cladistics 15, 191–197.

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

Brower, A.V.Z., DeSalle, R., Vogler, A., 1996. Gene trees, species trees, and systematics: a cladistic perspective. Annual Review of Ecology Systematics 27, 423–450.

Bull, J.J., Huelsenbeck, J.P., Cunningham, C.W., Swofford, D.L., Waddell, P.J., 1993. Partitioning and combining data in phylogenetic analysis. Systematic Biology 42, 384–397.

Caterino, M.S., Vogler, A.P., 2002. The phylogeny of the Histeroidea (Coleoptera: Staphyliniformia). Cladistics 18, 394–415.

Cognato, A.I., Vogler, A.P., 2001. Exploring data interaction and nucleotide alignment in a multiple gene analysis of *Ips* (Coleoptera: Scolytidae). Systematic Biology 50, 758–780.

Cunningham, C.W., 1997. Can three incongruence tests predict when data should be combined? Molecular Biology and Evolution 14, 733–740.

Damgaard, J., Andersen, N.M., Sperling, F.A.H., 2000. Phylogeny of the water strider genus *Aquarius* Schellenberg (Heteroptera: Gerridae) based on mitochondrial and nuclear DNA and morphology. Insect Systematics and Evolution 31, 71–90.

Damgaard, J., Sperling, F.A.H., 2001. Phylogeny of the water strider genus *Gerris* Fabricius (Heteroptera; Gerridae) based on COI mtDNA, EF-1α and morphology. Systematic Entomology 26, 241–254.

Danforth, B.N., Ji, S., 1998. Elongation factor-1α occurs as two copies in bees: implications for phylogenetic analysis of EF-1α sequences in insects. Molecular Biology and Evolution 15, 225–235.

Debry, R.W., 2001. Improving interpretation of the decay index for DNA sequence data. Systematic Biology 50, 742–752.

De Pinna, M.C.C., 1991. Concepts and tests of homology in the cladistic paradigm. Cladistics 7, 367–394.

Dolphin, K., Belshaw, R., Orme, C.D.L., Quicke, D.L.J., 2000. Noise and incongruence: interpreting results of the incongruence length difference test. Molecular Phylogenetics and Evolution 17, 401–406.

Dowton, M., Austin, A.D., 2001. Simultaneous analysis of 16S, 28S, COI and morphology in the Hymenoptera: Apocrita-evolutionary transitions among parasitic wasps. Biological Journal of the Linnean Society 74, 87–111.

Doyle, J.J., 1992. Gene trees and species trees: molecular systematics as one character taxonomy. Systematic Botany 17, 144–163.

Doyle, J.J., 1997. Trees within trees: gene and species, molecules and morphology. Systematic Biology 46, 537–553.

Farris, J.S., Kaller, M., Klugesjö, A.G., Butl, C., 1995. Testing significance of congruence. Cladistics 10, 315–320.

Friedlander, T.P., Regier, J.C., Mitter, C., 1994. Phylogenetic information content of five nuclear gene sequences in animals: initial assessment of character sets from concordance and divergence studies. Systematic Biology 43, 511–525.

Gallant, S.L., Fairbairn, D.J., 1996. A new species of *Aquarius* from the southeastern United States with electrophoretic analyses of the clade containing *Gerris*, *Limnoporus*, and *Aquarius* (Hemiptera: Gerridae). Systematics 89, 637–644.

Gatesy, J., O'Grady, P., Baker, R.H., 1999. Corroboration among data sets in simultaneous analysis: hidden support for phylogenetic relationships among higher level artiodactyl taxa. Cladistics 15, 271–313.

Goloboff, P., 1993. NONA. Program and Documentation. Version 2.0.

Graybeal, A., 1994. Evaluating the phylogenetic utility of genes: a search for genes informative about deep divergences among vertebrates. Systematic Biology 43, 174–193.

Hungerford, H.B., Matsuda, R., 1960. Keys to subfamilies, tribes, genera and subgenera of the gerridae of the World. Kansas University Scientific Bulletin 41, 3–23.

Jukes, T.H., Cantor, C.R., 1969. Evolution of protein molecules. In: Munro, H.N. (Ed.), Mammalian Protein Metabolism. Academic Press, New York, pp. 21–132.

Kallersjö, M., Albert, V.A., Farris, J.S., 1999. Homoplasy increases phylogenetic signal. Cladistics 8, 275–287.

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

Maddison, W.P., 1997. Gene trees in species trees. Systematic Biology 46, 523–536.

Maddison, W.P., Maddison, D.R., 1992. MacClade: Analysis of Phylogeny and Character Evolution, Version 3.05. Sinauer Associates, Sunderland, MA.

Matsuda, R., 1960. Morphology, evolution and a classification of the Gerridae (Hemiptera–Heteroptera). University of Kansas Scientific Bulletin 41, 25–632.

Mitchell, A.S., Cho, S., Regier, J.C., Mitter, C., Poole, R.W., Matthews, M., 1997. Phylogenetic utility of elongation factor-1α in Noctuoidea (Insecta: Lepidoptera): the limits of synonymous substitutions. Molecular Biology and Evolution 14, 381–390.

Miyamoto, M.M., Fitch, W.M., 1995. Testing species phylogenies and phylogenetic methods with congruence. Systematic Biology 44, 64–76.

Moran, N.A., Matthew, E.K., Gelsey, M.J., Murphy, T.G., Scholes, E.A., 1999. Phylogenetics and evolution of the aphid genus *Uroleucon* based on mitochondrial and nuclear DNA sequences. Systematic Entomology 24, 85–93.

Nixon, K.C., Carpenter, J.M., 1996. On simultaneous analysis. Cladistics 12, 221–241.

Olmstead, R., Reeves, P.A., Yen, A.C., 1998. Patterns of sequence evolution and implications for parsimony analysis of chloroplast DNA. In: Soltis, D.E., Soltis, P.S., Doyle, J.J. (Eds.), Molecular Systematics of Plants II: DNA Sequencing. Kluwer, Boston, pp. 164–187.

Olmstead, R., Sweere, J., 1994. Combining data in phylogenetic systematics: an empirical approach using three molecular data sets in Solenaceae. Systematic Biology 43, 467–481.

Pamilo, P., Nei, M., 1988. Relationships between gene trees and species trees. Molecular Biology and Evolution 5, 568–583.

Phillips, A., Janies, D., Wheeler, W.C., 2000. Multiple sequence alignment in phylogenetic analysis. Molecular Phylogenetics and Evolution 16, 317–330.

Polhemus, D.A., Polhemus, J.T., 1994. A new species of *Aquarius* (Heteroptera: Gerridae) from Timor, with notes on Timorese zoogeography. Proceedings of the Entomological Society of Washington 96, 54–62.

Reed, R.D., Sperling, F.A.H., 1999. Interaction of process partitions in phylogenetic analysis: an example from the swallowtail butterfly genus *Papillio*. Molecular Biology and Evolution 16, 286–297.

Sanderson, M.J., Donoghue, M.J., 1996. The relationship between homoplasy and confidence in a phylogenetic tree. In: Sanderson, M.J., Hufford, L. (Eds.), Homoplasy: The Recurrence of Similarity in Evolution. Academic Press, New York, pp. 67–89.

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., Flook, P., 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Annals of the Entomological Society of America 87, 651–701.

Soltis, D.E., Soltis, P.S., Mort, M.E., Chase, M.W., Savolainen, V., Hoot, S.B., Morton, C.M., 1998. Inferring complex phylogenies using parsimony: an empirical approach using three large DNA data sets for angiosperms. Systematic Biology 47, 32–42.

Sorenson, M.D., 1999. TreeRot, version 2. Boston University, Boston, MA.

Sota, T., Vogler, A.P., 2001. Incongruence of mitochondrial and nuclear gene trees in the carabid beetles *Ohomopterus*. Systematic Biology 50, 39–59.

Spence, J.R., 1990. Introgressive hybridization in Heteroptera: The example of *Limnoporus* Stål (Gerridae) species in western Canada. Canadian Journal of Zoology 68, 1770–1782.

Sperling, F.A.H., Spence, J.R., Andersen, N.M., 1997. Mitochondrial DNA, allozymes, morphology, and hybrid compatibility in *Limnoporus* water striders (Heteroptera: Gerridae): Do they all track species phylogenies? Annals of the Entomological Society of America 90, 401–415.

Swofford, D.L., 1991. When are phylogeny estimates from molecular and morphological data incongruent. In: Miyamoto, M.M., Cracraft, J. (Eds.), Phylogenetic Analysis of DNA Sequences. Oxford University Press, New York.

Swofford, D.L., 1998. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, MA.

Swofford, D.L., Olsen, G.J., Waddell, P.J., Hillis, D.M., 1996. Phylogenetic inference. In: Hillis, D.M., Moritz, C., Mable, B.K. (Eds.), Molecular Systematics. Sinauer, Sunderland, MA, pp. 407–514.

Tamura, K., Nei, M., 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Molecular Biology and Evolution 10, 512–526.

Templeton, A.R., 1983. Phylogenetic inference from endonuclease cleavage site maps with particular reference to the humans and apes. Evolution 37, 221–244.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research 24, 4876–4882.

Wenzel, J.W., Siddall, M.E., 1999. Noise. Cladistics 15, 51–64.

Yang, Z., 1998. On the best evolutionary rate for phylogenetic analysis. Systematic Biology 47, 125–133.

Yoder, A.D., Irwin, J.A., Payseur, B.A., 2001. Failure of the ILD to determine data combinability for slow loris phylogeny. Systematic Biology 50, 408–425.