*Zoological Journal of the Linnean Society* (2000), *130:* 511–526. With 3 figures

doi:10.1006/zjls.1999.0236, available online at http://www.idealibrary.com on IDEAL®

# Phylogeny of sea skaters, *Halobates* Eschscholtz (Hemiptera, Gerridae), based on mtDNA sequence and morphology

JAKOB DAMGAARD*

*Zoological Museum and Zoological Institute, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen, Denmark*

NILS MØLLER ANDERSEN FLS

*Zoological Museum, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen, Denmark*

LANNA CHENG

*Scripps Institution of Oceanography-0202, 9500 Gilman Drive, University of California, San Diego, La Jolla, CA 92093, U.S.A.*

FELIX A. H. SPERLING

*Department of Environmental Science, Policy and Management, Division of Insect Biology, 201 Wellmann Hall, University of California, Berkeley, CA 94720–3112, U.S.A.*

*Received May 1999; accepted for publication December 1999*

We examined phylogenetic relationships among halobatine water striders (Hemiptera, Gerridae) using molecular and morphological data. The molecular data set was 780 bp DNA sequence data from the 3′ half of the mitochondrial gene encoding *cytochrome oxidase subunit* I from 19 species of sea skaters, *Halobates*, and one species from each of three related genera, *Asclepios annandalei*, *Austrobates rivularis*, and *Eurymetra natalensis*. The morphological data set was a slightly modified version of a previously published data set. Unweighted parsimony analyses of the molecular data set gave one tree with weak support for most branches. Maximum likelihood analysis of the same data set gave a tree with slightly different topology, but reveiled many of the clades found in parsimony analyses of the morphological data set. Parsimony analyses of the combined molecular + morphology data sets gave a better

* Corresponding author: E-mail:jdamgaard@zi.ku.dk



0024–4082/00/120511+16 $35.00/0

resolved and better supported tree than did analyses of any single data set. The phylogeny of *Halobates* presented here allows a more rigorous evaluation of several prior hypotheses about evolutionary processes in marine water striders. In particular, it supports the hypothesis of at least two separate transitions from coastal to oceanic environments.



ADDITIONAL KEY WORDS:—marine insects – molecular systematics – cytochrome oxidase subunit I (COI) – unweighted parsimony – maximum likelihood – total evidence – Bremer support – ecological phylogenetics.

CONTENTS



## INTRODUCTION

Although many insects occur in marine habitats (Cheng, 1976), only five species of sea skaters, genus *Halobates* Eschscholtz (Hemiptera, Gerridae, Halobatinae) have successfully colonized open ocean where they live permanently upon the sea surface. There are an additional 39 described species of *Halobates* in sheltered, nearshore marine waters throughout the tropical Indo-Pacific (Herring, 1961; Andersen & Polhemus, 1976; Cheng, 1985; Andersen, 1982, 1991b, 1999). After being ignored for more than a century, our knowledge of the biology, ecology, and distribution of sea skaters has increased substantially during the past few decades (for reviews, see Cheng, 1973, 1975, 1985; Andersen, 1982; Spence & Andersen, 1994). However, many crucial questions remain to be answered, in particular those concerning oceanic species which usually live at some distance from the coast, and occur nearshore only after storms.

The oceanic sea skaters are widely distributed, viz. *Halobates germanus* White (Indian and West Pacific Oceans), *H. sericeus* Eschscholtz (North and South Pacific Ocean), *H. sobrinus* White (eastern tropical Pacific Ocean), *H. splendens* Witlaczil (south-eastern tropical Pacific Ocean), and *H. micans* Eschscholtz (Atlantic, Indian, and Pacific Oceans), occupying areas between the 20°C winter isotherms (Andersen, 1982; Cheng, 1985, 1989b). Most nearshore species have much more restricted distributions, being endemic to particular areas of continental coasts, islands, or groups of islands in the Indo-Pacific (Cheng, 1989a). Adult sea skaters are always wingless, but may be dispersed by surface currents along coasts, chains of islands, or even across stretches of the open sea. This may account for the wide distribution of several nearshore species, namely *H. flaviventris* Eschscholtz, *H. hayanus* White (Indian

and West Pacific Ocean), *H. mariannarum* Esaki (West Pacific Ocean), and *H. hawaiiensis* Usinger (Central Pacific Ocean) (Herring, 1961; Andersen, 1982, 1991b, 1999; Cheng, 1985).

There has been much speculation about the origin and evolution of sea skaters, particularly how, when, and where the oceanic species achieved their unique way of life (Cheng, 1989a; Andersen *et al.*, 1994). Evolutionary questions like these can only be answered through inference based on reliable hypotheses about phylogenetic relationships among the species involved. In his monographic revision of *Halobates*, Herring (1961) recognized several species groups and depicted relationships between them in a 'phylogenetic diagram'. However, some of these groups were explicitly based on plesiomorphic characters and may therefore not be monophyletic in the strict sense. Andersen (1991b) presented the first attempt to reconstruct the phylogeny of *Halobates* using cladistic methods applied to a suite of morphological characters, with emphasis on male and female genital structures. Based on this phylogenetic hypothesis, Andersen (1991b) and Andersen & Weir (1994a, b) partitioned the species of *Halobates* into a number of monophyletic species groups. The most basal group is the subgenus *Hilliella* China (with *H. mjobergi* Hale and *H. lannae* Andersen & Weir, both from Australia), whereas species of the subgenus *Halobates s.s.* fall into two clades, each with about the same number of species. Until quite recently, the sister group of *Halobates* was thought to be the likewise marine genus *Asclepios* (with three species in South and East Asia). However, the recently described, limnic halobatine *Austrobates rivularis* Andersen & Weir (1994a) from northern Australia, has proven to be even more closely related to *Halobates*. All three genera constitute the tribe Halobatini. The other halobatine tribe, Metrocorini, comprises about 80 species living in lotic freshwater habitats throughout the Ethiopian and Oriental regions. The age of the divergence between Halobatini (or at least the genus *Halobates*) and

TABLE 1. Taxa of halobatine water striders sampled for mtDNA nucleotide sequences, with locality data

| Species | Locality data |
|---|---|
| *Asclepios annandalei* | Singapore, Pulau Ubin. 01 22N, 103 51E, 27.xi.97, K.L. Yeo |
| *Austrobates rivularis* | Australia, Qld., Lydia Creek, 18.iii.1993, P. Zborowski |
| *Eurymetra natalensis* | Tanzania, Usa River E of Arusha, 17.i.1997, P. Gravlund |
| *Halobates alluaudi* | Seychelles, Daros I., 21.iv.1990, R/V Gitte Gry |
| *Halobates bryani* | Fiji, Suva Estuary, 9.i.1986, R. Hauser |
| *Halobates flaviventris* | Seychelles, Daros I., 21.iv.1990, R/V Gitte Gry |
| *Halobates germanus* | Arabian Sea, 14 03N, 50 50E, 9.viii.1995, M. Baars |
| *Halobates hawaiiensis* | Society Is., Moorea, Fare Huahine, 17 43S, 151W, 1.ix.1996, V. Resh |
| *Halobates hayanus* | Papua New Guinea, Port Moresby, Motupore Island, 09 30S, 147 17E, 12.iii.1990, L. Cheng |
| *Halobates mariannarum* | Caroline Is., Pohnpei, Ant Atoll, 07N, 158 13E, 19.iii.1996, R. Reame |
| *Halobates micans* | Arabian Sea, 05 39N, 54 55E, 2.v.1995, M. Baars |
| *Halobates mjobergi* | Australia, Qld., Roonga Point, 17.x.1992, T. Weir & P. Zborowski |
| *Halobates nereis* | W. Caroline Is., Palau, Kamori Channel, Koror, 07 20N, 134 10 E, 18.vii.1995, L. Cheng |
| *Halobates poseidon* | Indian Ocean, Cosmoledo Atoll, 27.iii.1989, D.A. Polhemus |
| *Halobates proavus* | Thailand, Phuket, 26–28.i.1987, N.M. Andersen |
| *Halobates robustus* | Galapagos Is. 00 42S, 91W, Cartago Bay, 19.ii.1982, P. Holdway |
| *Halobates salotae* | Tonga, Neiafu Harbour, Vavau, 18 39S, 173 59W, 2.v.1996, R. Martini |
| *Halobates sericeus* | Central Pacific, 26N, 156 W. 22.viii.1995, T. Villareal |
| *Halobates sexualis* | Sri Lanka, Beruwela, 15.iv.1995, M. Nummelin |
| *Halobates sobrinus* | Eastern Pacific, 06 12N, 81 19W, 21.i.1982, P. Holdway |
| *Halobates splendens* | Eastern Pacific, 03 11S, 96 01W, 26.ii.1982, P. Holdway |
| *Halobates whiteleggei* | Australia, N.S.W., Durras Lake, 10.xii.1994, E.S. Nielsen |

Metrocorini is inferred from a fossil *Halobates* species from Italy to be at least 45 million years ago (Andersen *et al.*, 1994).

Herring recognized an 'open-ocean' group of *Halobates* composed of *H. micans*, *H. sobrinus*, *H. splendens*, *H. germanus*, and *H. sericeus*, and suggested that this group is monophyletic (Herring, 1961: fig. 115). Andersen (1991b) questioned the monophyly of Herring's 'open-ocean' group since his cladistic analysis gave the result that the oceanic species *H. sobrinus*, *H. micans*, and *H. splendens* are more closely related to the nearshore species *H. flaviventris* and *H. hawaiiensis* than to the two other oceanic species, *H. germanus* and *H. sericeus*. This implies two possible evolutionary scenarios: one in which the oceanic way of life evolved only once, and where *H. flaviventris* + *hawaiiensis* have reversed to nearshore habitats; another scenario implies that oceanic habits have been acquired twice, independently of each other.

In the present study, we compare a phylogeny of *Halobates* and allied taxa based on the same morphological characters as used by Andersen (1991b), but scored for new taxa, with a phylogeny based on mitochondrial *cytochrome oxidase subunit* I (COI) sequence data, and combine the two data sets in a 'total evidence' analysis. The results of the phylogenetic analyses are used to discuss possible hypotheses about the evolution of the marine way of life in *Halobates*, in particular the oceanic habit of some species. The implications for studies of the historical biogeography of sea skaters are briefly discussed.

## MATERIAL AND METHODS

### *DNA sequences and protocols*

We sequenced mitochondrial DNA from samples representing 19 named species of the genus *Halobates*, two species belonging to other genera of the tribe Halobatini, viz. *Asclepios annandalei* Distant and *Austrobates rivularis* Andersen & Weir, and one species, *Eurymetra natalensis* Poisson, representing the tribe Metrocorini. The samples of *Halobates* were selected to represent most of the species groups recognized by Andersen (1991b) and Andersen & Weir (1994a, b): *H. mjobergi* Hale (subgenus *Hilliella*), *H. proavus* White (the *H. proavus* group). *H. sexualis* Distant, *H. whiteleggei* Skuse (the *H. regalis* group), *H. hayanus* (the *H. hayanus* group), *H. flaviventris*, *H. germanus*, *H. hawaiiensis*, *H. micans*, *H. sericeus*, *H. sobrinus*, *H. splendens* (the *H. micans* group), *H. nereis* Herring (the *H. matsumurai* group), *H. alluaudi* Bergroth (the *H. alluaudi* group), *H. mariannarum*, *H. salotae* Herring (the *H. mariannarum* group), *H. robustus* Barber, *H. bryani* Herring, and *H. poseidon* Herring. To determine intraspecific variations, we sequenced 10 individuals each of *H. germanus* (an oceanic species) and *H. nereis* (a coastal marine species) collected from adjacent geographic locations. Table 1 shows the locality data of all samples.

For most samples, we extracted DNA from alcohol-preserved adults, but for a few species, live specimens were killed in a −70°C freezer and extracted on site. Except for *H. germanus* and *H. nereis*, we selected males and used only the thorax which contain the powerful leg muscles. By removing the abdomen the risk of contamination with gut contents was also reduced. The head, abdomen, and limbs were stored in alcohol as voucher specimens. They are now deposited at the Zoological Museum, University of Copenhagen.

MtDNA was extracted by using the QiaAmp tissue kit protocol (QIAGEN Inc. Santa Clara, CA) which included at least 2 hours digestion of tissue with Proteinase K. The resulting volume of 300 µl DNA in solution gives plenty of template to work with. Polymerase chain reaction (PCR) amplifications were carried out in a thermal cycler in 51 µl of a cocktail containing 2 µl template); 5 µl of each primer; 14 µl ddH2O; 20 µl DNTP's (GATC 0.5 mM each); and 5 µl Promega PCR-reaction buffer (15 mM $MgCl_2$). After a 'hot start' with 2 min. of denaturation at 94°C, the reaction was paused at 72°C and 0.2 µl *Taq-polymerase* (5U/µl) was added. Amplification parameters for each of the following 35 cycles were as follows: 94°C for 1 min (denaturation), 45°C for 1 min (annealing) and 72°C for 1.5 min (extension). The target sequence was 788 bp from the 3′ half of the COI gene which corresponds to 2184–2971 in *Drosophila yakuba* delimited by the primers Jerry (C1-J-2183) (5′ CAA CAT TTA TTT TGA TTT TTT GG 3′) (reproduced from Simon *et al.*, 1994), and a modified version of C1-N-2968 (Sperling *et al.*, 1997) named K866 (C1-N-2972) (5′GTA TTT CGT TAT AA/T/C/GG AA/GT GTT 3′). Because this sequence was too long to be sequenced in one step, two internal primers were designed for amplification with the end-primers, namely LimH2929 (C1-N-2609) (5′CGA ATA CTG CTC CTA TTG ATA 3′) to work with Jerry, and Nils (C1-J-2387) (5′TCA CCA TCA ATA TTG TGA AC 3′) to work with K866. These gave an overlap of 220 bp. PCR-product was electrophoresed on a 2% NuSieve gel, stained with Ethidium Bromide and sized against a ΦX174/HaeIII (Boehringer Mannheim) DNA ladder under UV-light. PCR products were cleaned of primers and unincorporated nucleotides with a QIAquick PCR Purification Kit (QIAGEN Inc. Santa Clara, CA). Cycle sequencing was done using a Perkin Elmer/ABI Dye Terminator Cycle Sequencing Kit and run on a thermocycler using the profiles recommended by the kit manufacturers. Cycle sequencing products were cleaned using Centrisep columns, and sequenced using a Perkin Elmer ABI377 Automated Sequencer (Applied Biosystems Inc. Foster City, CA). DNA sequence for each species was confirmed with both sense and anti-sense strands. As the COI gene is protein coding with no insertions or deletions and relatively conserved amino acid sequences, alignment and contig construction was unproblematically performed in the program Sequencher (Gene Codes Corporation, Ann Arbor, Michigan). Sequence data are available from GenBank (accession numbers AF200281 through AF200302).

### *Morphological characters*

For morphological analyses, we scored the 22 species for the 64 characters used by Andersen (1991b). The reader should consult this work for definitions of characters and their states. Compared with the data set analysed by Andersen (1991b), nine species of *Halobates* and one species of *Asclepios* species were excluded from our study whereas four species of *Halobates* were replaced with closely related species, viz. *H. maculatus* Schadow by *H. proavus*, *H. peronis* Herring by *H. sexualis*, *H. esakii* Miyamoto by *H. nereis*, and *H. fijiensis* Herring by *H. salotae* with rescored characters. *Eurymetra natalensis* replaced *Metrocoris histrio* (White) as outgroup taxon. Finally, we have also included *Austrobates rivularis* (for details on morphology, see Andersen & Weir, 1994a).

TABLE 2. Morphological character data for 19 species of *Halobates*, *Austrobates rivularis*, *Asclepios annandalei*, and *Eurymetra natalensis*. Definitions of character and character states (0–3) adopted from Andersen (1991b) and Andersen & Weir (1994a). Question mark (?) denotes inapplicable or missing character scores

| | 0123456789 | 1111111111 0123456789 | 2222222222 0123456789 | 3333333333 0123456789 | 44444444444 01234567890 | 5555555555 1234567890 | 666 123 |
|---|---|---|---|---|---|---|---|
| *Halobates mjobergi* | 0000001001 | 0000001011 | 0011012000 | 0000100000 | 00010000000 | 1001101100 | 111 |
| *Halobates robustus* | 1110001000 | 0010102011 | 1011012112 | 0000100000 | 11010200110 | 0001112110 | 111 |
| *Halobates bryani* | 1100001000 | 0010102011 | 1011112012 | 0000100000 | 11010200100 | 0001110110 | 111 |
| *Halobates poseidon* | 1100001000 | 0010102011 | 1011112111 | 0221100100 | 1000100000 | 0001112100 | 011 |
| *Halobates nereis* | 2110101000 | 1022102001 | 1011112112 | 0221101000 | 11011100100 | 0001110110 | 111 |
| *Halobates alluaudi* | 2110111100 | 1022102001 | 1011112112 | 1010101010 | 11010200100 | 0001111110 | 111 |
| *Halobates mariannarum* | 2100101000 | 0021102001 | 1011112122 | 1010101010 | 11010100000 | 0001110110 | 111 |
| *Halobates salotae* | 2110101000 | 1022102001 | 1111012122 | 1010101010 | 11010100100 | 0101112010 | 111 |
| *Halobates proavus* | 1100001000 | 0010002011 | 1011112012 | 0010110000 | 11111000001 | 0011110110 | 011 |
| *Halobates whiteleggei* | 1000111000 | 0010102011 | 1011012111 | 0010100100 | 11112010?01 | 0001112110 | 111 |
| *Halobates sexualis* | 1100101000 | 0010102011 | 1011012111 | 0222100200 | 11122010?01 | 1001112110 | 111 |
| *Halobates hayanus* | 2110001000 | 0010102011 | 1011012000 | 0100100001 | 11211010?01 | 1002112111 | 111 |
| *Halobates germanus* | 2221011000 | 0010102111 | 1011012011 | 0000100001 | 11211010?01 | 1011112011 | 111 |
| *Halobates sericeus* | 2221111000 | 0010112111 | 1011012011 | 0000100100 | 11211010?01 | 0011112011 | 121 |
| *Halobates flaviventris* | 2110101000 | 0011102011 | 1011002000 | 2223100100 | 11110011?11 | 0002112111 | 021 |
| *Halobates hawaiiensis* | 2110101000 | 0011102011 | 1011002000 | 2223100100 | 11110011?11 | 0002112111 | 021 |
| *Halobates sobrinus* | 1221111000 | 0011102111 | 1011002000 | 2000110200 | 11100011?11 | 1012110100 | 011 |
| *Halobates splendens* | 1221111000 | 0012212111 | 1011012011 | 2223100000 | 11210011?11 | 1012011110 | 121 |
| *Halobates micans* | 1221111000 | 0011212111 | 1011012011 | 2223100100 | 11210011?11 | 1012011111 | 121 |
| *Austrobates rivularis* | 0000000001 | 1000001110 | 0011112000 | 0000000??? | 10000000000 | 0001110110 | 01? |
| *Asclepios annandalei* | 0000000001 | 0000102000 | 0000?01000 | 0010000??? | 10000000000 | 0001100110 | 011 |
| *Eurymetra natalensis* | 0000000010 | 0000000?00 | 0000?00??? | ????000??? | 00010000010 | 0002001110 | 000 |

The data matrix is shown in Table 2, with characters numbered 0–63 and their character states numbered 0–3. Some characters (nos 24, 27–33, 37–39, and 48) are conditionally defined and can only be scored if certain structures are present. These characters are scored as inapplicable (denoted by a question mark (?) in Table 2) in *Eurymetra natalensis*, *Asclepios annandalei*, *Austrobates rivularis*, and in some species of *Halobates*. All multistate characters are treated as non-additive (states unordered). Note: Andersen (1991b) treated 10 out of 23 multistate characters as additive (ordered), but the justification for doing so is in most cases weak.

### *Phylogenetic analyses*

Phylogenetic reconstructions were obtained by two methods: maximum parsimony and maximum likelihood. Unweighted parsimony analyses of various data sets were chiefly performed using the programs PAUP 3.1 (Swofford, 1993) and PAUP* 4.0b2 (Swofford, 1998) in combination with MacClade 3.05 (Maddison & Maddison, 1992). Since the number of taxa and the size of the data matrix precludes branch-and-bound searches, we carried out heuristic searches with 20 random-taxon-addition iterations. Clade stability was estimated using two different parameters: bootstrap and branch support (aka Bremer support or decay index; Bremer, 1994). Bootstrap values were generated in PAUP from 500 replicates, each with 10 random-addition heuristic searches. Branch support values were obtained in PAUP by using the 'converse constraints' approach to obtain branch support for the most stable clades (Bremer, 1994). For the analysis of combined data sets, partitioned branch support values for data partitions were calculated following the method described by Baker & DeSalle (1997) and Baker *et al.* (1998). To examine whether significant incongruence exists between data partitions based on codon positions, and between the nucleotide data set and the morphological data set, we conducted Incongruence Length Difference (ILD) tests (Farris *et al.*, 1995) using the 'con-test' command in the program DADA (Nixon, 1994) with 100 iterations, five 'autospin' random-additions searches per iteration, and the bb* command of Hennig86 (Farris, 1988). Maximum likelihood analyses were conducted in PAUP* (Swofford, 1998). The Hasegawa-Kishino-Yano (Hasegawa *et al.*, 1985) model of nucleotide substitution was implemented using observed nucleotide frequencies, two substitution types (transition/transversion ratio initially estimated by MacClade 3.05 from the 50% bootstrap consensus tree, and a discrete gamma distribution (d$\Gamma$) to account for rate variation among sites ($\alpha$). For the heuristic search, the 50% parsimony bootstrap tree was used as the starting tree for NNI branch swapping under likelihood settings. When $\alpha$ was estimated, this value was used for estimation of a new TI/TV ratio. At the completion of this search, the estimated $\alpha$ and TI/TV ratios were used for a heuristic search using NNI branch swapping to find the tree with the highest ln-likelihood. A molecular clock was not enforced.

## RESULTS

### *Nucleotides*

Of the 788 bp of COI amplified for the 22 species of halobatine water striders, 780 bp were unambiguously sequenced, except for *H. mjobergi* (771 bp). Of these,

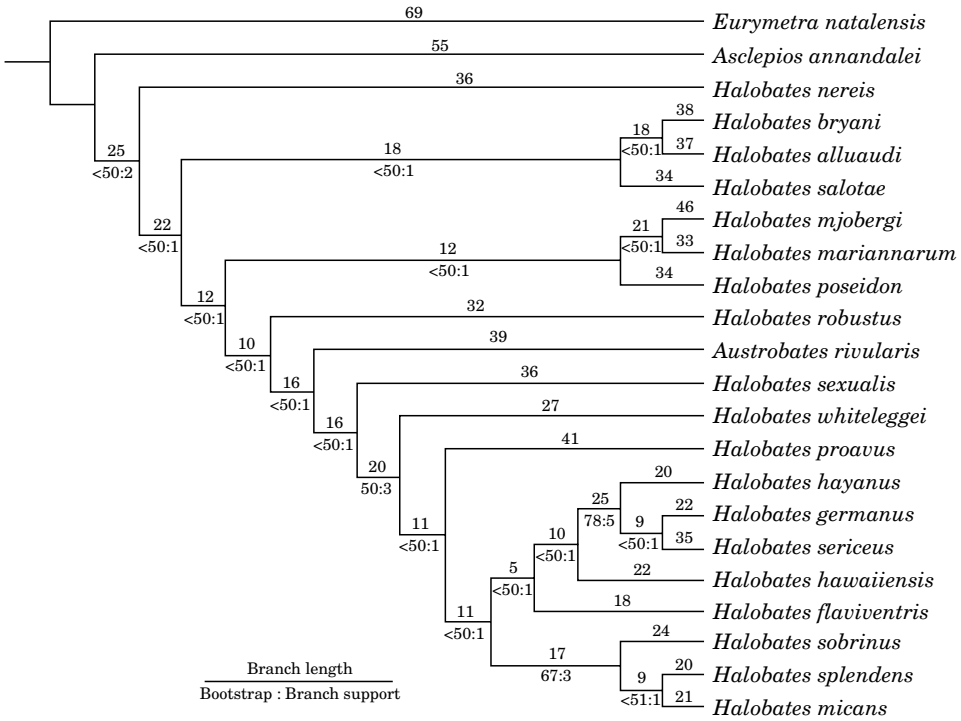


Figure 1. Single most parsimonious tree resulting from an unweighted parsimony analysis of 780 bp of the 3′ half of COI conducted in PAUP* (Swofford, 1998) using a heuristic search with 20 random-addition replicates. Length = 1026, CI = 0.332, RI = 0.340. Numbers above branches indicate branch length. The first number below each branch indicates bootstrap support from 500 pseudoreplicates with 10 random-addition heuristic searches per pseudoreplicate in PAUP*. The second number below each branch is branch support.

282 (36.2%) vary in comparisons across all taxa, 274 (35.1%) vary within the Halobatini, and 257 (32.9%) vary within *Halobates*. The number of phylogenetic informative characters are as follows: Halobatinae (206); Halobatini (199); and *Halobates* (183). Of the 282 variable sites, 225 (79.8%) are third codon positions versus 57 (20.2%) for first + second positions. The ratio of transitions to transversions is 1.81 over all sites (3.45 for first + second codon position; 1.68 for third codon positions. The nucleotide sequences are A/T rich, the A + T content ranging from 67.1 to 74.4%. Similar frequencies were found in mitochondrial protein coding genes in other hemimetabolous insects (Simon *et al.*, 1994), including *Limnoporus*, a genus of limnic water striders (Sperling *et al.*, 1997). Nucleotide frequencies for A range from 34.2 to 36.7%, for T from 32.9 to 37.7%, for C from 14.1 to 17.8%, and for G from 12.8 to 14.5%.

From a heuristic search of the unweighted nucleotide data set, we obtained one most parsimonious tree (MPT), 1026 steps long. The ln-likelihood value for this tree is 5297.9772. Figure 1 depicts this tree with branch lengths, bootstrap, and branch support values attached (see Table 3 for tree statistics). The clades supported by the

TABLE 3. Statistics for MPTs generated from parsimony analyses of morphology, COI mtDNA, amino acids, and combined morphology + COI mtDNA data sets. PIC = phylogenetic informative characters, CI = consistency index; RI = retention index

| Data sets | Nos of characters | Nos of PIC | Nos of MPTs | Length | CI | RI |
|---|---|---|---|---|---|---|
| Morphology | 64 | 58 | 1 | 179 | 0.453 | 0.664 |
| COI mtDNA | 780 | 206 | 1 | 1026 | 0.332 | 0.340 |
| Amino Acids | 260 | 11 | 26 | 72 | 0.447 | 0.553 |
| Combined | 844 | 264 | 1 | 1224 | 0.344 | 0.398 |

TABLE 4. Support for monophyly of species groups of *Halobates* in parsimony analyses of morphology, COI mtDNA, amino acids (AA), and combined morphology + COI mtDNA data sets and in a maximum likelihood analysis of COI mtDNA. X = group present, ? = unresolved

| Taxa | Unweighted parsimony | | | | Maximum Likelihood |
|---|---|---|---|---|---|
| | Morph. | COI | AA | Comb. | |
| 1. *H. micans* + *splendens* | X | X | X | X | X |
| 2. *H. flaviventris* + *hawaiiensis* | X | | ? | | |
| 3. *H. sobrinus* + 1 | | X | ? | X | X |
| 4. 2 + 3 | X | | ? | X | X |
| 5. *H. sericeus* + *germanus* | | X | | X | |
| 6. *H. hayanus* + 5 | | X | | X | X |
| 7. 4 + 6 | X | X | | X | X |
| 8. *H. proavus* + *whiteleggei* + 7 | X | X | X | X | X |
| 9. *H. sexualis* + 8 | X | | X | X | |
| 10. *H. mariannarum* + *salotae* | X | | ? | | |
| 11. *H. nereis* + *alluaudi* + 10 | X | | ? | X | |
| 12. *H. bryani* + *robustus* + *poseidon* + 11 | X | | | X | X |
| 13. *Halobates s.s.* | X | | ? | X | |
| 14. *Halobates* | X | | ? | X | X |
| 15. *Halobates* + *Austrobates* | X | X | X | X | X |

nucleotide data are listed in Table 4. Only the clades containing *H. hayanus* + *germanus* + *sericeus* and *H. sobrinus* + *splendens* + *micans* are reasonably well supported. The data neither support the monophyly of the subgenus *Halobates s.s*, nor the monophyly of the genus *Halobates*.

Nucleotide sequences were translated into amino acid sequences using the extended mtDNA code for *Drosophila* (Maddison & Maddison, 1992), and 40 out of 260 residues (15.4%) varied across all taxa with 11 (4.2%) being phylogenetically informative. An parsimony analysis of this data set gave 26 MPT's, each 72 steps long (see Table 3 for tree statistics). A strict consensus tree is highly unresolved, containing only the clades nos 1, 8, 9, and 15 (Table 4) and a presumably spurious clade *Halobates sericeus* + *whiteleggei*. The nucleotide sequence data were also used in generating a tree based on maximum likelihood. A TI/TV ratio of 3.194 and an among site variation of 0.157 gave the highest In-likelihood value 5275.5960. This tree was 19 steps longer than the MPT and contained clades nos. 1, 3, 6, 7, 8, and 15 (Table 4), also found in the parsimony analysis based on COI alone, as well as clades nos 4, 12, and 14 which were not found by the parsimony analysis. The

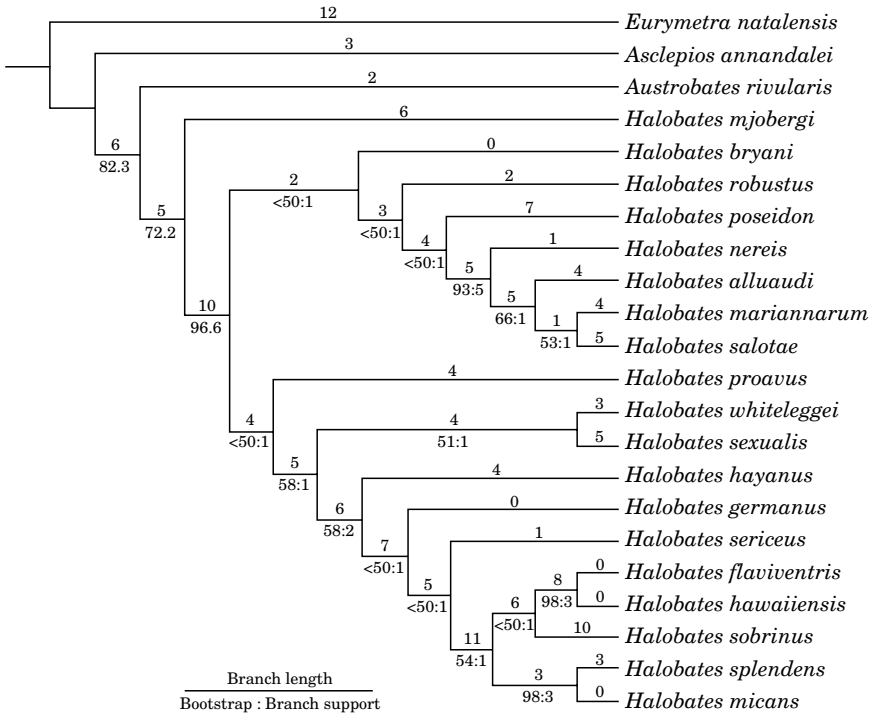


Figure 2. Single most parsimonious tree resulting from an unweighted parsimony analysis of 64 morphological characters conducted in PAUP* (Swofford, 1998) using a heuristic search with 20 random-addition replicates. Length = 179, CI = 0.453, RI = 0.664. Format for numbers as in Fig. 1.

maximum likelihood tree is 1045 steps long or 19 steps longer than the tree generated by maximum parsimony (Fig. 1).

### *Morphology*

An unweighted parsimony analysis of the morphological character matrix (Table 2) yielded only one MPT (Fig. 2), 179 steps long (see Table 3 for tree statistics). The following clades are well supported (in terms of bootstrap and branch support values; see Fig. 2): *Halobates micans* + *splendens*, *H. flaviventris* + *hawaiiensis* (identical scores for all characters), *H. hayanus* + the '*H. micans* group' (see above), *H. nereis* + *alluaudi* + *mariannarum* + *salotae*, *Halobates s.s.* (excluding *H. (Hilliella) mjobergi*), *Halobates*, and *Austrobates* + *Halobates*. The same clades are also supported by reliable morphological synapomorphies (Andersen, 1991b; Andersen & Weir, 1994a).

### *Combined data*

Finally, we performed unweighted parsimony analysis on the combined morphology and nucleotide sequence data set. This analysis yielded one MPT, 1224

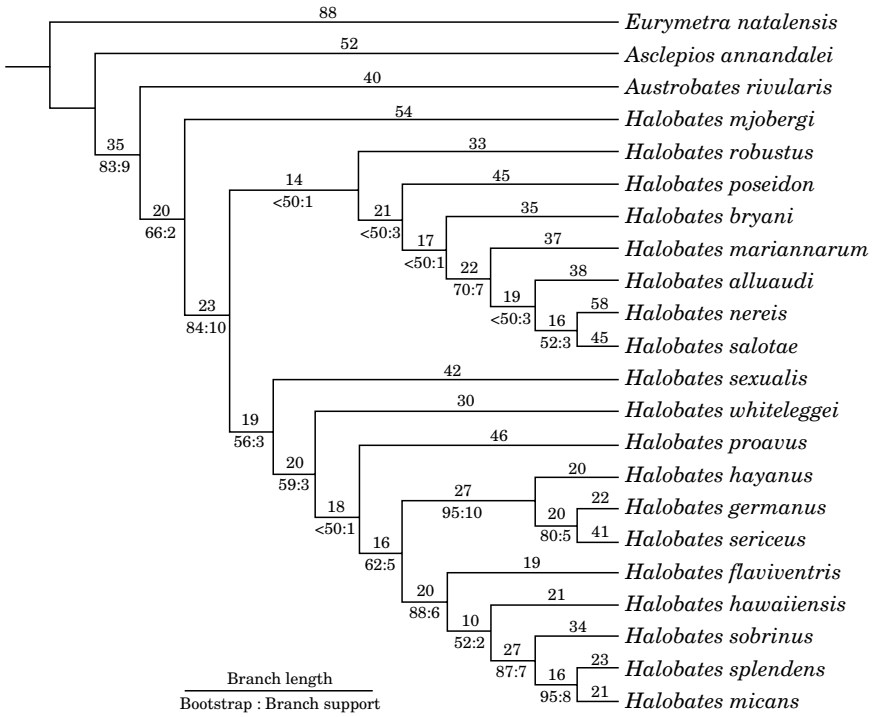


Figure 3. Single most parsimonious tree resulting from an unweighted parsimony analysis of the combined molecular and morphological data sets conducted in PAUP* (Swofford, 1998) using a heuristic search with 20 random-addition replicates. Length = 1224, CI = 0.344, RI = 0.398. Format for numbers as in Fig. 1.

steps long. This tree is depicted in Figure 3, with branch lengths, bootstrap, and branch support values attached (see Table 3 for tree statistics). The topology of the combined tree is more similar to that of the 'morphology' tree (Fig. 2) than to the topology of the 'nucleotide sequence' tree (Fig. 1). Clade nos 1, 3, 4, 5, 6, 7, 11, 13, and 15 (Table 4) are well supported, both in terms of bootstrap and branch support values. To examine whether significant incongruence exists between the morphological and molecular data sets, we conducted an Incongruence Length Difference (ILD) test (Farris *et al.*, 1995). The result ($P = 0.20$) indicates that the two data sets are not significantly incongruent. In order to assess the degree of support provided by each data set when analysed together, we calculated the partitioned branch support (= partitioned Bremer support; Baker & DeSalle, 1997; Baker *et al.*, 1998) for both data sets with reference to the combined tree (Fig. 3). The partitioned branch support values (Table 5) suggest that both data sets contribute to the overall branch support values for most clades. We therefore argue that the incongruence among our data partitions is isolated to specific relationships and the 'false' signal created by this incongruence (e.g. the non-monophyly of *Halobates* and spurious placement of *Austrobates* in Fig. 1) is overcome by a combined analysis. Following the argumentation presented by Baker *et al.* (1998), we believe that the simultaneous analyses of our morphological and molecular data sets are both justified and desirable (see also Baker & DeSalle, 1997; Sperling *et al.*, 1997; Baker *et al.*, 1998; Remsen & DeSalle, 1998).

TABLE 5. Overall and partitioned branch support (PBS) for the tree resulting from a parsimony analysis of combined morphology + COI mtDNA data sets (Fig. 3). Taxa are the same species groups as in Table 4. Further explanations in text. n/a = non-applicable

| Taxa | Combined | Morphology | COI mtDNA | PBS ratio |
|---|---|---|---|---|
| 1. *H. micans* + *splendens* | 8 | 6 | 2 | 3.00 |
| 2. *H. flaviventris* + *hawaiiensis* | n/a | n/a | n/a | n/a |
| 3. *H. sobrinus* + 1 | 7 | 0 | 7 | 0 |
| 4. 2 + 3 | 6 | 3.3 | 2.7 | 1.22 |
| 5. *H. sericeus* + *germanus* | 5 | 5 | 0 | indef. |
| 6. *H. hayanus* + 5 | 10 | −1.5 | 11.5 | −0.13 |
| 7. 4 + 6 | 5 | 1.8 | 3.2 | 0.56 |
| 8. *H. proavus* + *whiteleggei* + 7 | 3 | −3 | 6 | −0.50 |
| 9. *H. sexualis* + 8 | 3 | 0.5 | 2.5 | 0.2 |
| 10. *H. mariannarum* + *salotae* | n/a | n/a | n/a | n/a |
| 11. *H. nereis* + *alluaudi* + 10 | 7 | 5.5 | 1.5 | 3.67 |
| 12. *H. bryani* + *robustus* + *poseidon* + 11 | 1 | 0.7 | 0.3 | 2.33 |
| 13. *Halobates s.s.* | 10 | 12 | −2 | −6.00 |
| 14. *Halobates* | 2 | 3 | −1 | −3.00 |
| 15. *Halobates* + *Austrobates* | 9 | 5.3 | 3.7 | 1.43 |

## DISCUSSION

The phylogenetic relationships inferred from analyses of COI sequence data (Fig. 1) are not entirely congruent with those obtained from parallel analyses of morphological characters (Fig. 2). In the single MPT obtained by parsimony analysis (Fig. 1), the position of *Austrobates rivularis* within *Halobates* indicates that the molecular data weaken the monophyly of the latter genus which is otherwise supported by several morphological synapomorphies (Andersen, 1991b; Andersen & Weir, 1994a). In the same tree, *H. (Hilliella) mjobergi*, the most basal species of *Halobates*, is placed among members of the subgenus *Halobates s.s.*, whereas *H. nereis* assumes a basal position. However, as indicated by boostrap and branch support values, these branches are weakly supported. In the maximum likelihood tree, *H. (Hilliella) mjobergi* is moved to a basal position as the sister group to *Halobates s.s.* + *Austrobates*, and *H. nereis* is moved to a position within *Halobates* that is also supported by morphological data. This suggest that the maximum likelihood model is more resistant to error caused by homoplasy than is maximum parsimony, possibly because the former is designed to account for multiple changes on long branches (Kuhner & Felsenstein, 1994; Huelsenbeck, 1995).

The COI gene appears to be excellent for phylogenetic comparisons of not too distantly related species of insects such as the *Drosophila buzzatti* species complex (Spicer, 1995), bumble bees (Pedersen, 1996), attine ants (Wetterer *et al.*, 1998), and species groups of water striders (Sperling *et al.*, 1997; Damgaard *et al.*, 2000), but less so for comparisons between genera and higher taxa. The limitations of COI depend on its place in the mitochondrial membrane which causes functional constraints on the number of first and second position sites which are free to vary in order to maintain the hydrophobic and hydrophilic amino acids. However, third codon positions are much more susceptible to the problems of multiple hits, a characteristic accentuated in mitochondria due to the expanded codon recognition pattern that allows a single tRNA species to decode four codons (Gray, 1989). For example, as transitions generally accumulate more rapidly than transversions,

the strength of their phylogenetic signal often decreases with increased sequence divergence due to multiple substitutions, which erase prior history and introduce homoplasy. Thus, in deep rooted phylogenies there is the potential for considerable homoplasy in third positions, especially due to saturation of A-T transversions. Since the tribe Halobatini (or at least the genus *Halobates*) diverged from its limnic sister group, Metrocorini, at least 45 million years ago (as inferred from a fossil *Halobates* species; Andersen *et al.*, 1994), a more complete understanding of the phylogeny in the tribe may call for the use of another, more conserved gene.

When the molecular and morphological data sets are combined (Fig. 3), the phylogenetic relationships between *Austrobates*, *Halobates* (*Hilliella*), and *H.* (*Halobates s.s.*) are largely congruent with those inferred from morphological characters (Fig. 2). The nucleotide sequence data support the monophyly of a group composed by the oceanic species *Halobates sobrinus*, *H. splendens*, and *H. micans*. Another well supported group is composed by the oceanic species *H. germanus* and *H. sericeus*, and the nearshore species *H. hayanus*. Finally, the molecular data strongly support the hypothesis proposed by Andersen (1991b), that *H. flaviventris* (and its morphological sibling species *H. hawaiiensis*) is more closely related to some open ocean species than to other nearshore *Halobates*.

Establishing a reliable phylogeny for sea skaters helps answer important questions about the evolution of sea skaters and, in particular, of the oceanic way of life in some species. We believe that a phylogenetic reconstruction based on both molecular and morphological data is reliable enough to justify such generalizations. The phylogeny presented in Figure 3 suggests that ancestral Halobatini lived in both limnic and marine habitats as proposed by Andersen (1991b) and Andersen & Weir (1994a). *Asclepios* apparently preferred coastal marine habitats whereas *Austrobates* and *Halobates* evolved from their euryhaline ancestors to inhabit limnic and marine habitats, respectively. The oceanic way of life in some *Halobates* species probably evolved at least twice. First, at the base of the clade composed by the oceanic species *H. germanus* and *H. sericeus*, and second, at the base of the clade composed by the oceanic species *H. sobrinus*, *H. splendens*, and *H. micans*. This hypothesis is more parsimonious than one including only one basal transition to the open ocean, and three independent reversals to nearshore habitats in *H. hayanus*, *H. flaviventris*, and in *H. hawaiiensis*.

Historical biogeography also depends upon reliable phylogenetic reconstructions (i.e. Andersen, 1991a, 1998, 1999), but biogeographical scenarios are further complicated by the interplay between events of vicariance, dispersal, and extinction (Andersen *et al.*, 1994; Andersen, 1999). The known distribution of *Halobates* covers the Indo-Pacific region (excluding the presence of *H. micans* in the Atlantic Ocean). This is not only because of the widespread occurrence of a few, open-ocean species. There are endemic, coastal species of *Halobates* in such distantly separated places as the Red Sea and the Galapagos Islands (Herring, 1961). It is therefore hardly possible to delimit a place of origin for *Halobates* based alone upon the present distribution of the genus. The sister-group of *Halobates* is the genus *Austrobates* which is endemic to Cape York Peninsula, northeastern Australia. Within *Halobates* the sister-group of all other species is two species belonging to the subgenus *H. (Hilliella)* (Andersen & Weir, 1994b) which are distributed along the coasts of tropical northern Australia. The phylogeny presented here supports the hypothesis that *Halobates* diverged from its sister-group *Austrobates* somewhere in the area which now constitutes the northernmost part of the Australian continent. Recent fossil evidence (Andersen

*et al.*, 1994) indicates that *Halobates* had evolved before the Middle Eocene (45 Mya), when Australia was part of eastern Gondwanaland.

The oceanic sea skaters probably evolved in the Indo-Pacific since the nearshore relatives of the oceanic species (*Halobates flaviventris*, *hayanus*, and *hawaiiensis*) are widespread within this area. It is noteworthy that the pantropical *Halobates micans* is closely related to two species, *H. sobrinus* and *H. splendens*, both of which are confined to the eastern tropical Pacific Ocean. *H. micans* probably separated from *H. splendens* in the central or eastern part of the Pacific Ocean and later dispersed across the Pacific to the Indian as well as the Atlantic Oceans. Future studies of relationships between populations of *H. micans* are needed to track possible routes of dispersal.

## ACKNOWLEDGEMENTS

The bulk of this research was carried out at the Department of ESPM, University of California, Berkeley, during the tenure of a Research Fellowship for J.D. We are indebted to Andrew M. Shedlock and Orion Jankowski for assistance in developing methods for extraction and sequencing; Mike Caterino, Anthony Cognato and May Kuo for technical advice and support; and Peter Arctander, Peter Gravlund, Rudolf Meier, and Bo Vest Pedersen for valuable comments on an earlier version of the manuscript. We are also grateful to the following for collecting specimens for this study: M. Baars (NIOZ, Texel, The Netherlands), P. Gravlund (Zoological Museum University of Copenhagen, Denmark), F. Martini (Hawaii), E.S. Nielsen, T. A. Weir and P. Zborowski (CSIRO Division of Entomology, Canberra, Australia), M. Nummelin (Zoological Institute, University of Helsinki), D.A. Polhemus (Smithsonian Institution, Washington, D.C.), R. Ream (University of Washington, Seattle), V. Resh (University of California, Berkeley), T. Villareal (University of Massachusetts, Boston), and C. M. Yang (Zoological Reference Collections, National University of Singapore). This work was supported by grants from the Danish Natural Science Research Council (Grant No. 9502155 to J. D. and N. M. A.), the US Office of Naval Research (Grant No. N10014–95–1–0142 to L. C.), and The California Agriculture Experiment Station to F.A.H.S.

## REFERENCES

**Andersen NM. 1982.** The semiaquatic bugs (Hemiptera, Gerromorpha). Phylogeny, adaptations, biogeography, and classification. *Entomonograph* **3:** 1–455.

**Andersen NM. 1991a.** Cladistic biogeography of marine water striders (Hemiptera, Gerromorpha) in the Indo-Pacific. *Australian Systematic Botany* **4:** 151–163.

**Andersen NM. 1991b.** Marine insects: genital morphology, phylogeny and evolution of sea skaters, genus *Halobates* (Hemiptera, Gerridae). *Zoological Journal of the Linnean Society* **103:** 21–60.

**Andersen NM. 1998.** Marine water striders (Heteroptera, Gerromorpha) of the Indo-Pacific: cladistic biogeography and Cenozoic palaeogeography. In: Hall R, Holdway JD, eds. *Biogeography and Geological Evolution of SE Asia.* Leiden: Backhuys Publishers, 341–354.

**Andersen NM, 1999.** The evolution of marine insects: phylogenetic and geographical aspects of species diversity in marine water striders. *Ecography* **22:** 98–112.

**Andersen NM, Farma A, Minelli A, Piccoli G. 1994.** A fossil *Halobates* from the Mediterranean and the origin of sea skaters (Hemiptera, Gerridae). *Zoological Journal of the Linnean Society* **112:** 479–489.

**Andersen NM, Polhemus JT. 1976.** Water-striders (Hemiptera: Gerridae, Veliidae, etc.). In: Cheng L, ed. *Marine Insects*. Amsterdam: North-Holland Publishing Company, 187–224.

**Andersen NM, Weir TA. 1994a.** *Austrobates rivularis* gen. et sp. nov., a freshwater relative of *Halobates* (Hemiptera, Gerridae) with a new perspective on the evolution of sea skaters. *Invertebrate Taxonomy* **8:** 1–15.

**Andersen NM, Weir TA. 1994b.** The sea skaters, genus *Halobates* Eschscholtz (Hemiptera, Gerridae), of Australia: taxonomy, phylogeny, and zoogeography. *Invertebrate Taxonomy* **8:** 861–909.

**Baker RH, DeSalle R. 1997.** Multiple sources of character information and the phylogeny of Hawaiian Drosophilids. *Systematic Biology* **46:** 654–673.

**Baker RH, 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.

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

**Cheng L. 1973.** *Halobates. Oceanography and Marine Biology, Annual Review* **11:** 223–235.

**Cheng L. 1975.** Marine pleuston – Animals at the sea-air interface. *Oceanography and Marine Biology, Annual Review* **13:** 181–212.

**Cheng L (Ed). 1976.** *Marine Insects*. Amsterdam: North-Holland Publishing Company.

**Cheng L. 1985.** Biology of *Halobates* (Heteroptera: Gerridae). *Annual Review of Entomology* **30:** 111–135.

**Cheng L. 1989a.** Biogeography and phylogeny of the sea-skater *Halobates*. *Chinese Journal of Oceanography and Limnology* **7:** 233–239.

**Cheng L. 1989b.** Factors limiting the distribution of *Halobates* species. In: Ryland JS, Tyler PA, eds. *Reproduction, Genetics and Distribution of Marine* Organisms. 23rd European Marine Biology Symposium. Fredensborg, Denmark: Olsen & Olsen, 357–362.

**Damgaard J, Andersen NM, Sperling FAH. 2000.** Phylogeny of the water strider genus *Aquarius* Schellenberg (Heteroptera: Gerridae) based on mitochondrial and nuclear DNA sequences and morphology. *Insect Systematics & Evolution* **31:** 71–90.

**Farris JS. 1988.** *Hennig86, version 1.5*. Copenhagen: N.M. Andersen, Zoological Museum, University of Copenhagen.

**Farris JS, Källersjö M, Kluge AG, Bult C. 1995.** Constructing a significance test for incongruence. *Systematic Biology* **44:** 570–572.

**Gray MW. 1989.** Origin and Evolution of mitochondrial DNA. *Annual Review of Cellular Biology* **5:** 25–50.

**Hasegawa M, Kishino M, Yano T. 1985.** Dating the human-ape split by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* **22:** 160–174.

**Herring JL. 1961.** The genus *Halobates* (Hemiptera: Gerridae). *Pacific Insects* **3:** 223–305.

**Huelsenbeck JP. 1995.** Performance of phylogenetic methods in simulation. *Systematic Biology* **44:** 17–48.

**Kuhner MK, Felsenstein J. 1994.** A simulation comparison of phylogeny algorithms under equal and unequal evlutionary rates. *Molecular Biology and Evolution* **11:** 459–468.

**Maddison WP, Maddison DR. 1992.** *MacClade: Analysis of Phylogeny and Character Evolution, version 3.05*. Sunderland, Massachusetts: Sinauer Associates.

**Nixon KC. 1994.** *DADA, version 1.1*. Ithaca, New York: Nixon K, Cornell University.

**Pedersen BV. 1996.** A phylogenetic analysis of cuckoo bumblebees (*Psithyrus*, Lepeletier) and bumblebees (*Bombus*, Latreille) inferred from sequences of the mitochondrial gene cytochrome oxidase I. *Molecular Phylogenetics and Evolution* **5:** 289–297.

**Remsen J, DeSalle R. 1998.** Character congruence of multiple data partitions and the origin of the Hawaiian Drosophilidae. *Molecular Phylogenetics and Evolution* **9:** 225–235.

**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 PCR primers. *Annals of the Entomological Society of America* **87:** 651–701.

**Spence JR, Andersen NM. 1994.** Biology of water striders: Interactions between systematics and ecology. *Annual Review of Entomology* **39:** 101–128.

**Sperling FAH, Spence JR, Andersen NM. 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.

**Spicer GS. 1995.** Phylogenetic utility of the mitochondrial cytochrome oxidase gene: Molecular evolution of the *Drosophila buzzatii* species complex. *Journal of Molecular Evolution* **41:** 749–759.

**Swofford DL. 1993.** *PAUP*. Phylogenetic Analysis Using Parsimony, version 3.1.1*. Champaign, Illinois: llinois Natural History Survey.

**Swofford DL. 1998.** *PAUP**. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sunderland, Massachusetts: Sinauer Associates.

**Wetterer JK, Schultz TR, Meier R. 1998.** Phylogeny of fungus-growing ants (tribe Attini) based on mtDNA sequence and morphology. *Molecular Phylogenetics and Evolution* **9:** 42–47.