Syst. Biol. 50(6):945–969, 2001

Sequence Alignment of 18S Ribosomal RNA and the Basal Relationships of Adephagan Beetles: Evidence for Monophyly of Aquatic Families and the Placement of Trachypachidae

VEREL L. SHULL,[1] ALFRIED P. VOGLER,[1,2,5] MICHAEL D. BAKER,[3,4] DAVID R. MADDISON,[3] AND PETER M. HAMMOND[1]

[1]*Department of Entomology, The Natural History Museum, London SW7 5BD, UK*
[2]*Department of Biology, Imperial College at Silwood Park, Ascot, Berkshire SL5 7PY, UK*
[3]*Department of Entomology, The University of Arizona, Tucson, Arizona 85721, USA*

Abstract.—Current hypotheses regarding family relationships in the suborder Adephaga (Coleoptera) are conflicting. Here we report full-length 18S ribosomal RNA sequences of 39 adephagans and 13 outgroup taxa. Data analysis focused on the impact of sequence alignment on tree topology, using two principally different approaches. Tree alignments, which seek to minimize indels and substitutions on the tree in a single step, as implemented in an approximate procedure by the computer program POY, were contrasted with a more traditional procedure based on alignments followed by phylogenetic inference based on parsimony, likelihood, and distance analyses. Despite substantial differences between the procedures, phylogenetic conclusions regarding basal relationships within Adephaga and relationships between the four suborders of Coleoptera were broadly similar. The analysis weakly supports monophyly of Adephaga, with Polyphaga usually as its sister, and the two small suborders Myxophaga and Archostemata basal to them. In some analyses, however, Polyphaga was reconstructed as having arisen from within Hydradephaga. Adephaga generally split into two monophyletic groups, corresponding to the terrestrial Geadephaga and the aquatic Hydradephaga, as initially proposed by Crowson in 1955, consistent with a single colonization of the aquatic environment by adephagan ancestors and contradicting the recent proposition of three independent invasions. A monophyletic Hydradephaga is consistently, though not strongly, supported under most analyses, and a parametric bootstrapping test significantly rejects an hypothesis of nonmonophyly. The enigmatic Trachypachidae, which exhibit many similarities to aquatic forms but whose species are entirely terrestrial, were usually recovered as a basal lineage within Geadephaga. Strong evidence opposes the view that terrestrial trachypachids are related to the dytiscoid water beetles. [Adephaga; aquatic beetles; Coleoptera; sequence alignment; small subunit rRNA, tree alignment.]

Adephaga contains more than 30,000 species or almost 10% of all the described species of Coleoptera and is traditionally grouped into 8 to 12 or more families. Numerous morphological characteristics indicate monophyly of the suborder (see Beutel, 1995), and several fundamental traits differ from those of the second large suborder, Polyphaga (Lawrence and Newton, 1982). Although relationships with the small suborders Archostemata and Myxophaga are still the subject of much debate, the division of the majority of extant beetle species into Adephaga and Polyphaga seems to reflect a natural, deep-rooted divergence that few authors dispute (Crowson, 1960; Kukalova-Peck and Lawrence, 1993; Lawrence and Newton, 1982).

[4]Present address: Department of Entomology, 411 Science II Building, Iowa State University, Ames, Iowa 50011, USA.
[5]Corresponding author; E-mail: a.vogler@nhm.ac.uk

Whereas Adephaga can be taken as a well-supported monophyletic group, relationships within Adephaga are highly contentious. Ecologically, they can be subdivided into a series of terrestrial groups, including Carabidae, Paussidae, Rhysodidae, Cicindelidae, and Trachypachidae (referred to as "Geadephaga"), and aquatic groups, including Gyrinidae, Dytiscidae, Noteridae, Hygrobiidae, Haliplidae, and Amphizoidae ("Hydradephaga"). This subdivision is ambiguous for the semiaquatic Amphizoidae and is compromised by the terrestrial Trachypachidae. Although trachypachids inhabit dry places away from open water and generally share the habits and body shape of the terrestrial Carabidae, they resemble Hydradephaga in several features generally considered to be adaptations to aquatic life. Further, the aquatic groups are ecologically and functionally heterogeneous and include species crawling on submerged vegetation feeding either on algae (Haliplidae) or on oligochaetes

Downloaded from http://sysbio.oxfordjournals.org/ at Zoological institute RAS on March 20, 2015

(Hygrobiidae), fast-swimming predatory forms such as the "true" diving beetles (Dytiscidae), and the highly specialized whirlygig beetles (Gyrinidae), which hunt on the water surface. Modes of swimming and stroke also differ between groups, with Hygrobiidae and Haliplidae using all three pairs of legs, whereas Dytiscidae and Noteridae use only the hind legs to produce the main thrust and the middle legs to maneuver. Haliplids and hygrobiids exhibit alternating leg movements, but noterids and dytiscids move the pairs of legs simultaneously. This diversity of aquatic life styles therefore raises the question about the common origin of the aquatic groups (Beutel, 1995; I. Ribera, pers. commun.)

The two principal hypotheses relating to basal relationships in Adephaga (Fig. 1) differ in the phylogenetic relevance of the separation of terrestrial and aquatic taxa. Crowson (1955, 1960) was the first to propose the monophyly of the hydradephagan families and to treat them as sister to the terrestrial Geadephaga (Fig. 1A). He further suggested that Trachypachidae represented a relict group that shares morphological characters with both Hydradephaga and Geadephaga and therefore concluded that this family is derived from a terrestrial ancestor that is sister to the (derived) Hydradephaga. Crowson concluded that only two groups could be readily distinguished within Hydradephaga: the "lower" families (Amphizoidae, Hygrobiidae, and Haliplidae) and the dytiscoid families (Gyrinidae, Noteridae, and Dytiscidae), the latter being a product of a subsequent radiation from an ancestral group of dytiscoid families. Within Geadephaga, which is dominated by the species-rich Carabidae (ground beetles; over 24,000 described species), Crowson placed Rhysodidae, a small group of species feeding on slime molds associated with rotten wood (Bell, 1994), as the sister to the remaining geadephagans (minus Trachypachidae).

Downloaded from http://sysbio.oxfordjournals.org/ at Zoological institute RAS on March 20, 2015

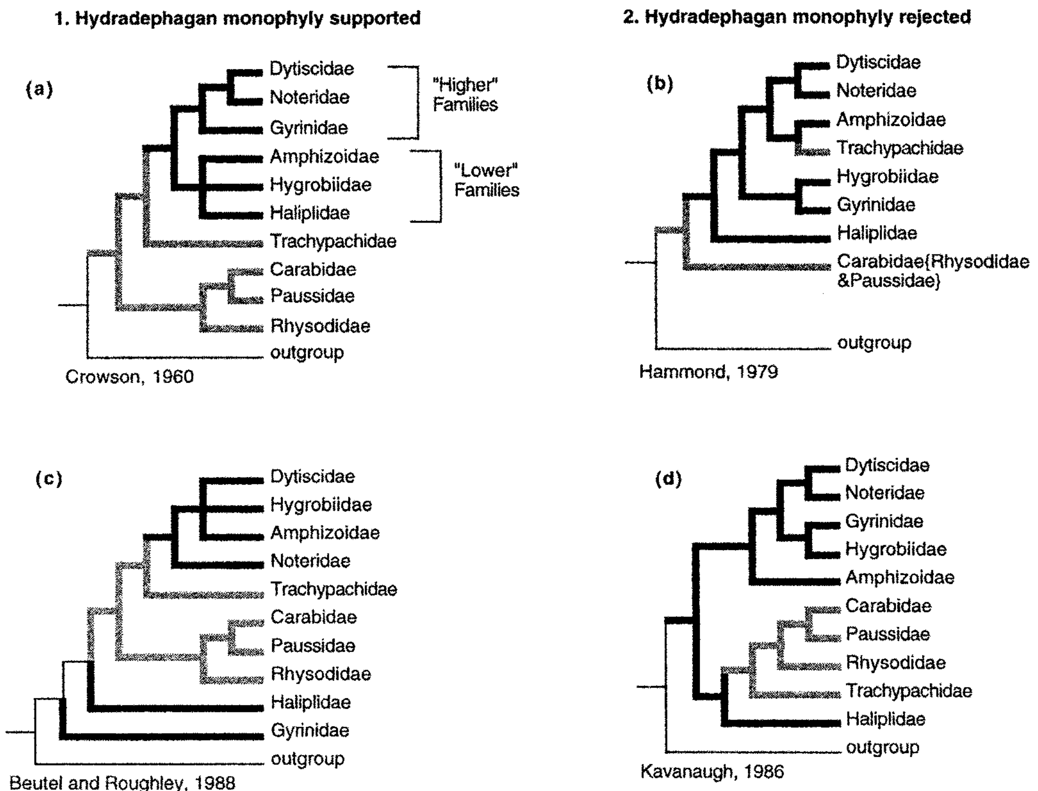

FIGURE 1. Relationships of Adephagan families as proposed by (a) Crowson (1960), (b) Hammond (1979), (c) Kavanaugh (1986), and (d) Beutel and Roughley (1988). Thick black lines denote hydradephagan taxa, thick gray lines represent geadephagan taxa, and thin black lines signify other Coleoptera and outgroups. Cicindelidae have generally been considered a grouping within the Carabidae and are not shown separately in these trees.

He also redefined Paussidae to include the Ozaeninae and concluded that this family was the sister to the remaining Carabidae. Finally, he considered the Cicindelidae (tiger beetles) a subfamily within the Carabidae, but of unclear phylogenetic associations.

Several authors concur with Crowson's scheme but they disagree with the position of Trachypachidae as the sister to Geadephaga. Hammond (1979) considered Hydradephaga to include Trachypachidae (rendering Hydradephaga sensu Crowson paraphyletic) (Fig. 1B). Some authors have generally agreed with Hammond, concluding that Trachypachidae form a monophyletic group together with the dytiscoid (Noteridae, Amphizoidae, Hygrobiidae, and Dytiscidae) complex (Bell, 1966; Ward, 1979; Roughley, 1981).

The second basic hypothesis of basal adephagan relationships rejects the monophyly of Hydradephaga. These proposals consider one or more of the hydradephagan families to be sister of Geadephaga. One of these (Kavanaugh, 1986) proposed Haliplidae to be sister to Geadephaga, with Trachypachidae at the base of the latter (Fig. 1D). Beutel and colleagues, on the basis of a numerical cladistic analysis using adult and larval characters (Beutel and Roughley, 1988; Beutel, 1993, 1995, 1999; Beutel and Haas, 1996), proposed that the surface-swimming Gyrinidae are the most basal group of Adephaga, followed by Haliplidae, and an ancestor of Dytiscoidea and Geadephaga (Fig. 1C). Although on this tree it is most-parsimonious to presume aquatic life style to be ancestral within Adephaga, Beutel and colleagues follow Crowson (1960) who stated that, "No serious coleopterist has ever suggested that terrestrial caraboids are derived from the aquatic ones; it is universally assumed that the derivation has been in the reverse sense." Therefore, Beutel's cladograms invoke three independent invasions into aquatic environments (as first suggested by Bell, 1966, and supported by Ponomarenko, 1977) from riparian ancestors (Beutel, 1999).

Given the difficulties in resolving phylogenetic relationships in Adephaga with morphological characters, we conducted a molecular study to address the question of adephagan higher-level relationships. We used 18S rRNA, a molecule that in recent years has turned into the marker of choice for analyzing higher-level phylogenies in beetles (Vogler and Pearson, 1996; Farrell, 1998; Maddison et al., 1999a,b). Although this molecule contains tracts of both slowly and quickly evolving sequences and is potentially useful for resolving relationships over a wide hierarchical range (Hillis and Dixon, 1991), phylogenetic inferences are compromised by extreme rate heterogeneity and differences in nucleotide composition and sequence length, in particular in the hypervariable regions or expansion segments (Hancock et al., 1988; Tautz et al., 1988). However, given the relative ease of data collection with universal polymerase chain reaction (PCR) primers, an ever-increasing database of 18S rRNA sequences throughout the tree of life, including Coleoptera, requires development and test of appropriate methods for alignment and phylogenetic reconstruction. In this paper, we explore the effect of different alignment strategies, in particular focusing on tree alignment for aligning sequences concurrent with the tree search (Sankoff, 1975; Wheeler, 1996), and compare the results with those of more traditional procedures in which the tree search is performed on a prior alignment.

Methodological Issues of Sequence Alignment

Tree inference methods that use tree alignments (Sankoff and Cedergren, 1983) are "alignment-producing"; that is, for each phylogenetic tree considered, an evolutionary scenario is reconstructed by which an ancestral sequence is assumed to have evolved into the observed sequences along the tree, thus producing an alignment of bases. Because sequence evolution consists of base substitutions, insertions, and deletions, all of these events would ideally be considered in calculating the tree length or likelihood of a tree (Sankoff et al., 1973; Sankoff, 1975; Sankoff and Rousseau, 1975; Sankoff and Cedergren, 1983; Felsenstein, 1988; Hein, 1989; Thorne and Kishino, 1992; Mitchison and Durbin, 1995; Wheeler, 1996, 1998; Mitchison, 1999). Proposed homologies of bases in extant species would follow from the placement of insertions and deletions on the tree, and these bases would thus be aligned one to another. Thus, for each tree, a tree alignment would be created, an alignment that probably would differ from tree to tree in regions with a rich history of insertions and deletions (Wheeler, 1998).

Downloaded from http://sysbio.oxfordjournals.org/ at Zoological institute RAS on March 20, 2015

In contrast, the standard approach dissociates the alignment of bases (determination of base homologies) from the phylogenetic inference. Alignments are typically done before phylogenetic inference, rather than performed on the individual trees being compared during the search for optimal trees. Moreover, the alignment of sequences produced attempts to minimize substitutions, insertions, and deletions—often by considering one fixed phylogeny or by making pairwise alignments of closely similar sequences or groups of sequences. The tree length or other objective function used to judge a tree is then calculated with this matrix. Thus, the homology of bases remains fixed after the first step, and the quest for the optimal tree does not permit alternative alignments during tree search; that is, insertion and deletions are immovable during the tree inference step. Correspondences of bases between sequences therefore are not determined in the context of the tree, and character substitutions might be assumed (and used as the basis for counts of tree lengths and likelihoods) between bases that are not of common descent.

Unfortunately, no known method can accurately consider substitutions, insertions, and deletions simultaneously on a tree—and thus find the optimal tree alignment or calculate tree length or likelihood of the tree—despite several attempts to develop one (Sankoff, 1975; Sankoff and Rousseau, 1975; Sankoff and Cedergren, 1983; Hein, 1989; Mitchison and Durbin, 1995; Wheeler, 1996, 1998; Schwikowski and Vingron, 1997; Ravi and Kececioglu, 1998; Mitchison, 1999). The search for the optimal alignment on any one tree is as difficult as the search for optimal trees (Woo et al., 1994), and thus for most cases only approximate methods of calculating the tree length or likelihood of a tree are available.

We used two approaches for tree reconstruction, both approximations to an ideal analysis. The first considered substitutions, insertion, and deletion events simultaneously in judging each tree. The application of this strategy is much facilitated with the development of the POY software (Gladstein and Wheeler, 1996), which uses an algorithm that produces an estimated tree alignment and associated tree length and thus permits selecting the shortest trees/alignments among a collection of trees. An unsolved problem with this is that the rank order of a set of trees will depend on the accuracy of the estimated tree lengths. If the rank order of estimated versus true tree lengths of a collection of trees differs, then the trees judged optimal on the basis of POY's estimated tree length may not correspond to the trees judged optimal according to the true tree length. Because the magnitude of the error of the estimated tree length is not known, how severe a problem this might be is unclear. (Although POY does calculate the exact cost of a given tree alignment, tracing back the downpass through the tree and determining the cost of that optimization, this does not solve the problem of determining the optimal tree alignment, that is, the set of correspondences of the bases that gives the lowest cost for a given tree [W. Wheeler, in litt.]). Nonetheless, the method provides a valuable counterpart to the more traditional approach of separation of the analysis into two steps, alignment and phylogeny inference. The approach of POY better addresses the problem of inferring insertions and deletions on a tree but gives only approximate tree lengths; the traditional approach, on the other hand, does not address insertions and deletions appropriately but gives exact tree lengths or likelihood values.

Here we have compared the results of applying both procedures to a dataset of 49 full-length sequences of the 18S rRNA gene for a representative sample of adephagan taxa and outgroups from the three other suborders of Coleoptera and related neuropteroid orders of insects. Both types of analysis produced broadly similar tree topologies that discriminate between some of the proposed hypotheses of adephagan relationships, in particular with regard to the major subdivision into terrestrial and aquatic families. The similarity of results of the two procedures suggests that approximate, parsimony-based, tree alignment methods have matured to be viable alternatives to the traditional approach.

Materials and Methods

Taxon Choice and Sampling Approach

Our goal was to include species from each of the nine adephagan families recognized by Lawrence and Newton (1995) plus the distinct lineages Cicindelidae and Paussidae (Table 1). For the larger adephagan

Downloaded from http://sysbio.oxfordjournals.org/ at Zoological institute RAS on March 20, 2015

TABLE 1. Summary of sampled taxa.

Classification	GenBank accession number	Location and date of specimen collection, or DNA sequence reference	Collector
Neuroptera: Ithonidae			
Oliarces clara Bamks	AF012527	Maddison et al. (1999b)	
Neuroptera: Chrysopidae			
Anisochrysa carnea Stephens	X89482	Chalwatzis et al. (1996)	
Raphidioptera: Raphidiidae			
Phaeostigma notata Linné	X89494	Chalwatzis et al. (1996)	
Megaloptera: Sialidae			
Sialis sp.	X89497	Chalwatzis et al.(1996)	
Archostemata: Cupedidae			
Distocupes sp.	AF201421	Australia	J. Galian
Myxophaga: Torridincolidae			
Torridincola rhodesica Steffan	AF201420	S. Africa: Punmalanga (Eastern Transvaal), Graskop, Waterval River, 24°50′ S 30°52′ E (5 February 1997)	S. Endrödy-Younga
Myxophaga: Hydroscaphidae			
Hydroscapha natans LeConte	AF012525	Maddison et al. (1999b)	
Polyphaga: Hydrophilidae			
Helochares lividus Forstmann	AF201418	Near Minety, Wiltshire, U.K. (30 March 1997)	P. Hammond
Polyphaga: Scarabaeidae			
Dynastes granti Horn	AF002809	Maddison et al. (1999a)	
Polyphaga: Clambidae			
Clambus arnetti Endrödy-Younga	AF012526	Maddison et al. (1999b)	
Polyphaga: Scirtidae			
Cyphon hilaris Nyholm	AF201419	Pen Ponds, Richmond Park, Surrey, U.K.	P. Hammond
Polyphaga: Tenebrionidae			
Tenebrio molitor Linné	X07801	Hendriks et al. (1988)	
Polyphaga: Anthribidae			
Bruchela conformis Suffrian	AF201417	Valence, France (July 1996)	M. Barclay
Adephaga: Gyrinidae			
Gyrininae			
Gyrinus sp. Müller	AF201412	Sepulga Creek, Conecuh Co., Alabama, USA (1994)	S. Oygur
Spanglerogyrinae			
Spanglerogyrus albiventris Folkerts	AF201413	Old Town Creek at Highway 31, 31°27′N 86°49′W, Alabama, USA (17 March 1997)	Maddison, Baker, and Ober
Adephaga: Haliplidae			
Haliplus laminatus Schaller	AF201405	U.K. (1 March 1997)	P. Hammond
Haliplus ruficollis Degeer	AF201406	Ravenscourt Park, London, U.K. (23 October 1996)	P. Hammond
Adephaga: Noteridae			
Noterinae			
Hydrocanthus oblongus Sharp	AF201415	Sepulga Creek, Conecuh Co., Alabama, USA (1994)	S. Oygur
Noterus clavicornis Degeer	AF201416	Bookham Common, Surrey. U.K. (30 October 1996)	M. Barclay
Suphis inflatus LeConte	AF012523	Maddison et al. (1999b)	
Adephaga: Hygrobiidae			
Hygrobia hermanni Fabricius	AF201414	Bookham Common, Surrey, U.K. (11 September 1997)	C. Turner
Adephaga: Dytiscidae			
Copelatinae			
Copelatus kalaharii Gschwendtner	AF201407	Ol Ari Nyiro Ranch, 50 km NW of Rumuruti, Laikipia District, Kenya (25 August 1996)	D. J. Larson

(*Continued on next page*)

Downloaded from http://sysbio.oxfordjournals.org/ at Zoological institute RAS on March 20, 2015

TABLE 1. Continued.

Classification	GenBank accession number	Location and date of specimen collection, or DNA sequence reference	Collector
Laccophilinae			
Laccophilus hyalinus Degeer	AF201410	Bookham Common, Surrey, U.K. (30 October 1996)	M. Barclay
Hydroporinae			
Hydroporus erythrocephalus Linné	AF201409	Niedersachsen, Zwillbrocker Venn, Germany	A. Vogler
Neoporus sp.	AF201411	Sepulga Creek, Conecuh Co., Alabama, USA (1994)	S. Oygur
Dytiscinae			
Cybister fimbriolatus Say	AF201408	Whiting et al. (1997)	
Adephaga: Trachypachidae			
Systolosoma lateritium Négre	AF012522	Maddison et al. (1999b)	
Trachypachus holmbergi Mannerheim	AF201394	Univ. of Alberta, Edmonton, Alberta, Canada (1995)	D. Pollock
Adephaga: Carabidae			
Cychrini			
Scaphinotus petersi catalinae Van Dyke	AF002801	Maddison et al. (1999a)	
Nebriini			
Nebria brevicollis Fabricius	AF201395	Ruislip Lido, London, U.K. (9 April 1996)	P. Hammond
Siagonini			
Siagona jennisoni Dejean	AF012494	Maddison et al. (1999b)	
Loricerini			
Loricera pilicornis Fabricius	AF201396	Ruislip Lido, London, U.K.	P. Hammond
Elaphrini			
Elaphrus cupreus Duftschmid	AF201397	Niedersachsen, Zwillbrocker Venn, Germany (15 July 1994)	A. Vogler
Migadopini			
Antarctonomus complanatus Blanchard	AF012504	Maddison et al. (1999b)	
Omophronini			
Omophron americanus Dejean	AF201398	Sepulga Creek, Conecuh Co., Alabama, USA (1994)	S. Oygur
Scaritini			
Pasimachus californicus Chaudoir	AF201399	Sierra Anchas, Gila Co., Arizona, USA (26 June 1994)	D. F. Brown
Clivinini			
Clivinini sp. (larva)	AF201400	North shore of Lake Okeechobee, Florida, USA (12 September 1976)	P. Hammond
Dyschirius aeneus Dejean	AF201401	Eypes Mouth, Dorset, U.K. (4 August 1997)	R. Booth
Broscini			
Creobius eydouxi Guérun-Ménéville	AF012498	Maddison et al. (1999b)	
Bembidiini			
Bembidion teracolum Say	AF201402	Ruislip Lido, London, U.K. (9 April 1996)	P. Hammond
Psydrini			
Psydrus piceus LeConte	AF002784	Maddison et al. (1999b)	
Mecyclothorax vulcanus Blackburn	AF012482	Maddison et al. (1999b)	
Platynini			
Agonum albipes/marginatum Linné	AF201403	Ruislip Lido, London, U.K. (9 April 1996)	P. Hammond
Chlaeniini			
Chlaenius vestitus Paykull	AF201404	Ruislip Lido, London, U.K. (9 April 1996)	P. Hammond
Adephaga: Paussidae			
Metrius contractus Eschscholtz	AF012515	Maddison et al. (1999b)	
Pachyteles striola sp.	AF012517	Maddison et al. (1999b)	

Downloaded from http://sysbio.oxfordjournals.org/ at Zoological institute RAS on March 20, 2015

TABLE 1. Continued.

Classification	GenBank accession number	Location and date of specimen collection, or DNA sequence reference	Collector
Adephaga: Cicindelidae			
Omus californicus Eschscholtz	AF201392	2 mi S of Camel, Monterey Co., California, USA (20 March 1994)	D. Brzoska
Oxycheila nigroaenea Bates	AF201393	1700 m above sea level, 10 km E of Asni, Ecuador (21 September 1993)	D. L. Pearson
Adephaga: Rhysodidae			
Clinidium calcaratum LeConte	AF012521	Maddison et al. (1999b)	
Omoglymmius hamatus LeConte	AF012520	Maddison et al. (1999b)	

families we attempted to include a wide range of exemplars. Our dataset includes at least one species from the following: both subfamilies of Gyrinidae (Spanglerogyrinae and Gyrininae); one of the two subfamilies of Noteridae (Noterinae is included, Phreatodytinae is not); four of the six subfamilies of Dytiscidae (Copelatinae, Laccophilinae, Hydroporinae, and Dytiscinae are included; Colymbetinae and Aubehydrinae are not); and 15 tribes of Carabidae. A recent and more thorough sampling of Carabidae can be found in Maddison et al. (1999b). Because of the possibility that Adephaga is the most basal suborder of Coleoptera (Kukalova-Peck and Lawrence, 1993), we also included four published sequences of neuropteroid insect orders (Chalwatzis et al., 1996), the presumed sister groups of Coleoptera.

DNA Extraction, Amplification, and Sequencing

All sequences newly reported in this paper, except for *Spanglerogyrus* and *Torridincola*, which were obtained by the protocol described in Maddison et al. (1999b), were acquired as follows. Total DNA was extracted from single fresh, frozen, silica-dried, or ethanol-preserved specimens by a phenol–chloroform extraction method as described previously (Vogler et al., 1993). Only the thorax was used for large beetles, the entire specimen for small beetles. DNA amplification was performed with either AmpliTaq DNA polymerase (Perkin-Elmer) and the magnesium-containing buffer supplied by the manufacturer (100 mM Tris-HCl, pH 8.3 at 25°C, 500 mM KCl, 15 mM $MgCl_2$, and 0.01% w/v gelatin, autoclaved) or Ready-To-Go PCR beads, which contain 1.5 U of Taq polymerase, 10 mM Tris-HCl, pH 9.0 at 25°C, 50 mM KCl, 1.5 mM $MgCl_2$, 200 μM of each dNTP, and stabilizers, including bovine serum albumin (Pharmacia Biotech). Each 18S rRNA sequence was amplified as four overlapping fragments of 500–800 bp. Successful amplification of these smaller fragments was generally more successful than trying to amplify the gene as a single fragment. The primers used for amplification included three primers designed to amplify the 5′ end of the gene: 18S5′ (5′GACAACCTGGTTGATCCTGCCAGT), 300R (5′TCAGGCTCCCTCTCCGG), and 18Sb5.0 (5′TAACCGCAACAACTTTAAT). The six primers used to amplify the middle region of the gene were 18Sai, 18Sbi, 18Sa0.7, 18Sa1.0, and 18Sb0.5 (Whiting et al., 1997) and 18Sb2.5 (5′TCTTTGGCAAAT GCTTTCGC). Finally, four newly designed primers were used to amplify the 3′ end of the gene: 18Sa2.0 (5′ATGGTT GCAAAGCTGAAAC), 18Sa2.4 (5′TCCGAT AACGAACGAGACTC), 18S3′II (5′CATCT AAGGGCATCACAGAC), and 18S3′I (5′CA CCTACGGAAACCTTGTTACGAC). The cycling conditions generally used were: 30 s at 94°C, 30 s at 45–55°C, and 1–2 min at 72°C (repeated for 30 to 40 cycles), followed by 10 min at 72°C. Amplification products were purified with the GeneClean II kit (Bio 101, Inc.). Automated DNA sequencing reagents were supplied by either Applied BioSystems Ltd. (PRISM Ready Reaction Taq Cycle Sequencing, DyeDeoxy Terminator Reaction Kit), or Amersham Life Science, Inc. (Thermo Sequenase Dye Terminator Cycle Sequencing Pre-Mix Kit). Sequencing reactions were purified by ethanol precipitation and then electrophoresed on an ABI377 sequencer. Sequencing errors/ambiguities were edited, and contiguous sequences for a given species were constructed by using Sequencher 3.0 software (Gene Codes Corp.).

Downloaded from http://sysbio.oxfordjournals.org/ at Zoological institute RAS on March 20, 2015

The sequences used in this study were based on a single specimen and were complete except for the ends used for PCR and for additional short stretches of sequences missing from the 5′ end (*Omoglymmius hamatus*, *Torridincola rhodesica*) or from the 3′ end of the gene (*Suphis inflatus*, *Agonum albipes*, *T. rhodesica*, and *Distocupes* sp.). In addition, *Haliplus laminatus* and *H. ruficollis* are each missing a 34-base sequence located 3′ of the V4 expansion segment. Generally, amplification was straightforward for geadephagan taxa but more difficult in aquatic beetles, for which amplification frequently failed or produced highly divergent sequences possibly attributable to endogenous microorganisms. Efforts to obtain amplification products from a representative of Amphizoidae failed entirely, despite independent attempts in the laboratories of both A.P.V. and D.R.M. A detailed description of sequences from further aquatic adephagans and potential experimental inconsistencies of the sequences reported here will be discussed elsewhere (Ribera and Vogler, unpubl.), but for the large majority of nucleotide positions included in this study no such problems are expected. GenBank accession numbers are listed in Table 1.

Sequence Alignment and Phylogenetic Analysis

For tree alignment we used a parallel version of POY vers. 2.6 running in a cluster of six Pentium processors at 450 MHz connected in parallel and PVM software. For all searches reported, we present the best (lowest cost) trees from 20 random addition replicates and collecting no more than three shortest trees in each replicate (commands: *–random 20 –maxtrees 3*). To explore different gap costs, we used gap costs below (gap = 1, change = 1) and above (gap = 4) the POY default parameters (gap = 2). Constrained searches were conducted with the *–agree* command by using a constraint file obtained with the program Jack2Hen (available with the POY software). An aligned data matrix can be based on the tree reported by POY, which is itself based on the correspondences of bases among the terminal taxa, using the *–impliedalignment* option. The cost of trees based on this aligned matrix was also assessed in PAUP4.0b2a (Swofford, 1999). Tree lengths found (by using the Treescore option) were identical or very similar to those reported by POY, although further parsimony searches (100 TBR replicates, gaps coded as fifth character state, gap weight set to same value as specified in the POY searches) on this alignment frequently resulted in slightly shorter trees with minor differences in topology. Bremer support was calculated by an approximate method implemented in POY based on a search procedure using TBR swapping.

The complexity of POY searches was reduced by limiting which bases of the full sequence can be homologized (aligned) to each other. As a working hypothesis, we assumed that particular regions of DNA sequence were homologous among all taxa in the matrix, and we delimited these regions according to similarity in the primary structure. In total, we subdivided the primary sequence into 24 such regions. This was done by first subdividing the full-length sequence into three expansion segment regions (corresponding to expansion segments V2, V4, and V6 of Tautz et al., 1988) and four conserved regions flanking these. The expansion segments were further subdivided into three (V2), nine (V4) and seven (V6) segments, each delimited by easily recognizable primary structure motifs common to all taxa. Each of the three expansion segments revealed a central region of high length variability that did not exhibit clearly recognizable primary structure similarity among taxa. Between taxa, the four conserved regions never varied in length by more than ~20 bp, whereas the expansion segments, particularly the central regions of V2, V4, and V6, varied greatly. This difference was 53 bp in V2 (28 bp in *Tenebrio* compared with 81 bp in *Oxycheila*), 285 bp in V4 (28 bp in *Tenebrio* compared with 313 bp in *Metrius*), and 172 bp in V6 (32 bp in *Tenebrio* and *Dynastes* compared with 204 bp in *Pachyteles*). Subdivision of the matrix into such regions of a priori homology also allowed us to include or exclude certain parts of the sequence, in particular the most length-variable regions of the 18S rRNA gene. In practice, for defining the 24 regions of presumed homologies, initial alignments of the raw data were made with Clustal W1.7 (Higgins et al., 1996), using the default parameters of the program. The resulting Clustal matrix was edited

Downloaded from http://sysbio.oxfordjournals.org/ at Zoological institute RAS on March 20, 2015

and subdivided using Se-Al (Rambaut, 1996).

The second type of analysis, with alignment and tree inference separated, were carried out for 10 alignments produced with Clustal W1.7 and 1 alignment "by eye." Different alignments were created by using a range of gap opening:gap extension costs, keeping the cost for substitutions constant. Unless otherwise stated, characters in regions with large deletions or insertions were omitted, as follows: If any sequence had a consecutive run of five or more positions in one sequence (row) that lacked nucleotides (i.e., showed gaps), then all characters corresponding to these positions (columns) were omitted. Visual inspection of those characters thus excluded indicated that this served to exclude columns in the hypervariable regions of the molecule only. The "by eye" alignment was produced by D.R.M., using the secondary structure model of the carabid beetle *Loricera foveata* (Wuyts et al., 2000; see http://rrna.uia.ac.be/secmodel/Lfov_SSU.html). Only those regions that could be easily matched with the secondary structure model of *Loricera* were included in the analysis; this resulted in exclusion of 17.3% of the sequences' nucleotides (as opposed to 12.7–13.6% for the Clustal alignments).

Searches for optimal trees were conducted on each of the alignments by using PAUP*4.0b2a. In parsimony analyses, all characters were treated as unordered, gaps were treated as missing data, and sites were weighted equally. Heuristic searches for most-parsimonious trees used 1,000 random addition sequence replicates and TBR branch swapping. For some analyses, trees were constrained for exploring alternative hypotheses. For analyses with constraints on tree shape, heuristic searches used 200 random addition sequence replicates and TBR branch swapping.

Bremer support, bootstrap, and jackknife parsimony analyses were conducted to measure the support for monophyly of Hydradephaga. Searches for most-parsimonious trees to calculate Bremer support values used 100 random addition sequence replicates and TBR branch rearrangement with trees constrained not to contain the focal node. Bootstrap values were calculated with 200 bootstrap replicates, each with a heuristic search of 10 random addition sequence replicates and TBR branch rearrangement. Jackknife values were calculated with 1,000 jackknife replicates given 36.8% character deletion, with each replicate using a heuristic search involving one random addition sequence replicate and TBR branch rearrangement.

Maximum likelihood analysis began with choice of a model of character evolution and estimation of the values of parameters. For each alignment, a tree for judging models and for inferring parameter values was found by weak heuristic searches for maximum likelihood trees under simple models of evolution. These searches began with a single search with NNI swapping and assumptions of a transition/transversion ratio of 2 and no site-to-site rate variation. The transition/transversion ratio was estimated on the resulting tree, and then SPR swapping on the tree was conducted assuming the estimated transition/transversion ratio. These trees were considered (Maddison et al., 1999b) to be sufficient for inferring parameters of evolutionary models and for judging the fit of models. The likelihood of the one tree from each of these searches was then calculated for evolutionary models of varied complexity by using parameter values estimated from the data. The simplest model examined was a Jukes–Cantor model of nucleotide change, all sites presumed to evolve at the same rate, and base frequencies equal; the most complex was a General Time Reversible (six-parameter) model of nucleotide change, a proportion of sites not free to vary, and the remainder evolving at rates following a gamma distribution, and base frequencies matching the empirically observed values (the GTR + I + Γ model). These and the other 14 models were the same as those examined by Maddison et al. (1999b). As expected, the more complex models yielded greater likelihood values. The most improvement in likelihood came by allowing site-to-site rate variation into the models, rather than increasing the complexity of the nucleotide change model. The most complex model was at least 34 –ln L units less than the next best model. As tested with a likelihood ratio test with four degrees of freedom (see discussion in Maddison et al., 1999b), these models differed significantly in fit at the $P = 0.005$ level, and thus the GTR + I + Γ model was chosen.

Initial likelihood tree searches for the 10 Clustal alignments were conducted with

Downloaded from http://sysbio.oxfordjournals.org/ at Zoological institute RAS on March 20, 2015

8 to 11 random-addition-sequence heuristic searches, SPR rearrangements, and using the more complex model chosen. The maximum likelihood trees for six of these alignments showed Adephaga as monophyletic, whereas in the other four, Hydradephaga was more closely related to Polyphaga than to other adephagans (see Results). According to the diversity of resulting trees, two alignments showing a monophyletic Adephaga (2:1 and 5:1) and two alignments with a paraphyletic Adephaga (7:2 and 10:5) were chosen for more thorough searches and analyses. The more thorough searches for these four alignments and the "by eye" alignment consisted of SPR branch swapping and at least 15 random addition sequence replicates (enough replicates were conducted until the island containing the trees of greatest likelihood was found at least 12 times).

Searches for high-likelihood trees under various topological constraints used the same models of evolution as for unconstrained analyses. Constrained trees of high likelihood were found for the five primary matrices by SPR branch swapping and three random addition sequence replicates.

Minimum evolution distance analyses using LogDet distances were conducted to accommodate various base frequencies among clades (Lake, 1994; Lockhart et al., 1994; Steel, 1994). For each matrix 50 heuristic searches with random-addition-sequence starting trees and TBR branch rearrangement were conducted to find the optimal minimum evolution tree. Optimal constrained trees were sought with the same search strategy. LogDet bootstrap analyses were conducted with 500 replicates, each consisting of a single heuristic search for the minimum evolution tree, except for the "by eye" alignment, for which 2,000 bootstrap replicates were performed.

To test for the lack of monophyly of Hydradephaga, we used parametric bootstrapping (Swofford et al., 1996; Goldman et al., 2000). Sequence data were created through simulated evolution up the branches of a phylogeny in which Hydradephaga was constrained to be not monophyletic, and trees were then inferred from the simulated data; this allowed inference of the expected nature of estimated trees if Hydradephaga was not monophyletic. Because some parameters of the model used in the simulation must be inferred from an observed matrix, and because we had five primary alignments, the test was conducted separately for each of the five alignments. The null hypothesis consisted of the maximum likelihood tree under the constraint that Hydradephaga was not monophyletic, and branch lengths and parameter estimates of a GTR + I + γ model of sequence evolution were estimated by using likelihood from the observed 18S rDNA sequence data. Five hundred simulated data matrices were created under this model, with sequences of length equivalent to those of the observed matrix, by using the Genesis package (D. R. Maddison and W. P. Maddison, 2001) of Mesquite (W. P. Maddison and D. R. Maddison, 2001), which itself used the GTR calculations of PAL (Drummond and Strimmer, 2001). This model does not include insertion or deletion events, and thus all sequences within a simulation are of the same length. The tree length of the most-parsimonious trees under the same constraint as the model tree (nonmonophyly of Hydradephaga) was compared with the length of unconstrained most-parsimonious trees for each of the 500 simulated matrices; for each simulated matrix, both constrained and unconstrained most-parsimonious trees were found by using 20 searches beginning with random addition sequence trees followed by TBR branch rearrangement. This procedure generated a distribution for the test statistic, allowing determination of the probability for observing the equivalent test statistic calculated from the empirical data. This is similar to the likelihood ratio tests of Huelsenbeck et al. (1996), except that parsimony rather than likelihood values are used because of time constraints. Monophyly of Trachypachidae plus Dytiscoidea was tested in the same fashion.

To examine the possibility that outgroup sequences were too divergent for proper resolution and rooting of Adephaga, some analyses were repeated with neuropteroid sequences omitted, and again with all non-adephagans omitted. Four new Clustal alignments (with the same 2:1, 5:1, 7:2, and 10:5 alignment weights used previously) were created for the set of beetle sequences and four more for the set of adephagan sequences. Most-parsimonious trees (using 5,000 heuristic searches with random-addition-sequence starting trees and TBR branch rearrangement), LogDet minimum

Downloaded from http://sysbio.oxfordjournals.org/ at Zoological institute RAS on March 20, 2015

evolution trees (100 heuristic searches), and maximum likelihood trees (enough searches so that the optimal trees were found at least three times, using SPR branch rearrangement) were sought for each of these matrices plus the "by eye" matrix with neuropteroids or nonadephagans removed.

RESULTS

The Data

The 49 sequences in the analysis varied between 1,888 (*Tenebrio*) and 2,396 bp long (*Pachyteles*). The A/T content of the individual sequences varied from 44.6% in *Gyrinus* to 58.3% in *Pachyteles*. The proportion of character differences between species varied from 2%, when certain noterids or dytiscids were compared, to 21% between distantly related taxa. One alignment of the complete matrix (using the default parameters of Clustal W 1.7) produced 2,480 positions, of which 1,190 were variable and 834 parsimony-informative (1,109 and 771, respectively, if only the adephagan ingroup is considered).

Simultaneous Alignment and Tree Search

On the basis of previous studies (see Fig. 1) and our results, we specified 13 groups of interest to this analysis, which we tested to see if these were recovered in the various tree searches, as indicated in Table 2. We started the analysis, using the simultaneous approach of alignment and tree search, initially with the conserved 18S rRNA regions of the gene only and eliminating the variable regions V2, V4, and V6. Under a gap cost:change cost ratio of 1:1, 16 most-parsimonious trees of length 1,075 were identified that recovered nine of the focal groups as monophyletic (Table 2; Fig. 2). The tree showed the basal position of the suborders Archostemata and Myxophaga relative to Adephaga and Polyphaga, which are sister groups. Further, we found Hydradephaga to be monophyletic. The position of Trachypachidae was unresolved in the

Downloaded from http://sysbio.oxfordjournals.org/ at Zoological institute RAS on March 20, 2015

TABLE 2. A summary of monophyletic groups determined from 18S rRNA sequences using POY. Numbers in front of the taxon names correspond to particular nodes of specific interest and are also shown in Figures 2–6. Complete sequences were divided into regions as follows: excluding all expansion segments V2, V4, V6 (all V regions); excluding the central portions of V2, V4, and V6 (central V regions); and excluding the CRPS quartet of Maddison et al. (1999b), which consists of Cicindelidae, Rhysodidae, Paussidae, and Scaritini. Costs are listed as gap:change. M = Monophyletic, P = Para- or Polyphyletic, U = Unresolved.

	All characters		Excluding all V regions			Excluding central V regions			Excluding central V regions, minus CRPS		
	1:1[a]	2:1	1:1	2:1	4:1	1:1	2:1	4:1	1:1	2:1	4:1
Tree score	6,246	8,339	1,074	1,198	1,376	1,892	2,138	2,579	1,387	1,554	1,840
1 Coleoptera	P	P	M	M	P	M	M	M	M	M	M
2 Polyphaga + Adephaga	P	P	M	(M)[b]	P	M	M	M	M	M	M
3 Polyphaga	M	M	M	P	P	M	M	M	M	M	M
4 Adephaga	P	P	M	(M)[b]	P	M	M	M	M	M	P
5 Hydradephaga + Trachypachidae	P	P	M	(M)[b]	P	P	P	P	P	P	P
6 Hydradephaga	P	P	M	(M)[b]	M	P	M	M	M	U	(M)[c]
7 Hydradephaga minus Haliplidae	P	P	M	(M)[b]	M	P	P	P	P	P	P
8 Noteridae + Hygrodiidae + Dytiscidae	(M)[c]	P	M	(M)[b]	M	M	P	M	P	M	P
9 Hygrodiidae + Dytiscidae	(M)[c]	P	P	P	P	M	M	P	M	M	M
10 Geadephaga (incl. Trachypachydae)	P	P	P	P	P	M	M	M	M	M	P
11 Geadephaga minus Trachypachidae	P	P	M	M	P	M	P	P	P	P	P
12 CRPS quartet	P	P	P	M	P	P	P	P	n/a	n/a	n/a

[a]Gap cost, listed as gap:change.
[b]Monophyletic except for the inclusion of *Torridincola*.
[c]Monophyletic except for the inclusion of Polyphaga.

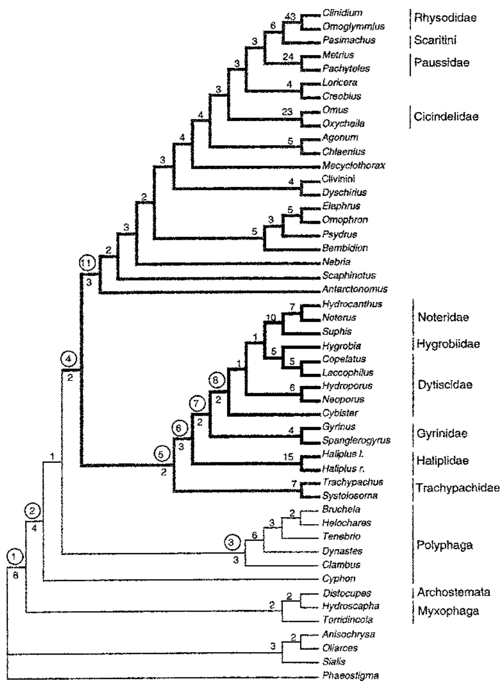

FIGURE 2. Most-parsimonious tree obtained from a POY analysis from the conserved regions (regions 1, 3, 5, and 7) with a gap cost and a change cost of one. The cost of the alignment is 1,074. Numbers on branches represent Bremer support values. The circled numbers refer to particularly relevant nodes in Tables 2 and 3. Shading of branches as in Figure 1.

various trees, being basal to Hydradephaga in some but basal to Geadephaga in others. Within Hydradephaga, we found Haliplidae and Gyrinidae basal to the Dytiscidae and Noteridae; the Dytiscidae also included the single representative of the Hygrobiidae. Within Geadephaga, relationships of tribal-level taxa largely followed the pattern observed by Maddison et al. (1999b). A group of four major taxa (Cicindelidae, Rhysodidae, Paussidae, and Scaritini), the CRPS quartet of Maddison et al. (1999b), which are unlikely to represent a clade in light of previous taxonomic work, was found near the derived Harpalinae but was not monophyletic. Bremer support values for many of the nodes, including the monophyletic Hydradephaga, were weak, but levels of support generally correlated with the stability to different alignment parameters.

We also analyzed the same set of characters under different gap costs and including portions of the V2, V4, and V6 regions (Table 2). The increase in gap cost above gap = 2 generally results in the loss of several monophyletic groups, including the monophyly of Coleoptera, Adephaga, Geadephaga, and others (Table 2). Similarly, tree searches on the complete 18S rRNA sequences including the hypervariable V2, V4, and V6 regions, resulted in the loss of monophyly of all major clades, except Polyphaga ("All characters" in Table 2). Such was the case for a wide range of alignment parameters (not shown), but resulted from inclusion of the middle parts of the hypervariable regions only: When the most variable (central) part of these regions were excluded, many of the monophyletic groups were recognizable again ("Excluding central V regions" in Table 2; Fig. 3). Removing the CRPS quartet from the analysis had no major impact on the topology of trees, although the cost of trees was much reduced by omitting these long sequences, and topologies were less affected by the increase of gap costs ("Minus CRPS" in Table 2)

Because the monophyly of some clades, including the Coleoptera and Adephaga,

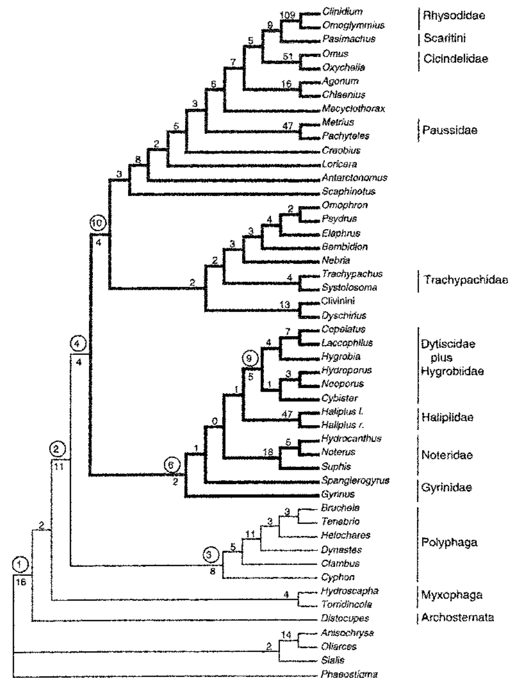

FIGURE 3. One of the three most-parsimonious trees obtained from a POY analysis of all regions except the central portion of the hypervariable regions with a gap cost and a change cost of one. The cost of this tree is 2,138 (also see Table 2). Decay index values are listed below relevant nodes, and suprageneric taxa are listed to the right.

Downloaded from http://sysbio.oxfordjournals.org/ at Zoological institute RAS on March 20, 2015

should not be in doubt, the signal of the hypervariable regions must introduce misleading information into the analysis, which was further investigated. First, to test for spurious alignment between distant parts of the sequence that are not homologous, the total sequence was subdivided into 24 segments to constrain which regions of the gene could assume nucleotide homology (see Materials and Methods). POY analysis on this subdivided data matrix resulted in a slightly shorter tree of 8,199 steps (compared with 8,339 steps when the matrix was divided into seven regions only), but this tree topology also failed to recover many of the well-established groups (not shown). To determine whether the misleading phylogenetic signal could be attributed to any single one of these three regions, the central regions were added in turn to analyses containing the remaining conserved characters. The inclusion of the central part of V2, but not of the much larger V4 and V6 regions, resulted in the recovery of most of the major clades of Table 2. The effect of V2 may be less severe because the number of characters in this short region are too few to affect the phylogenetic signal present in the conserved regions. The most-parsimonious tree obtained from the analysis that included this central V2 character was 2,854 steps; inclusion of the characters from the central portion of V4 and V6 produced trees of 4,918 and 4,564 steps, respectively. Trees produced from either the central portion of the V4 or V6 were entirely spurious, recovering none of the major clades. In addition, the fact that these trees are not congruent with each other, despite presumably sharing the same phylogenetic history, suggests they do not represent a consistent phylogenetic signal and cannot be used for inferences of relationships of Adephaga.

We further tested support for critical nodes and prior hypotheses about relationships in Adephaga by constraining tree searches in POY. Specifically, we were interested in the level of support for a monophyletic Hydradephaga and its relationships to other groups, particularly the Geadephaga and Trachypachidae. We also tested support for the "multiple origin" hypothesis of Beutel (1995) that renders the Hydradephaga polyphyletic. The constraints examined are defined in Table 3. Searches required the least additional cost under constraining for a monophyletic Adephaga and for the sister relationship of Geadephaga (including Trachypachidae) and Hydradephaga [(GT)HA scenario in Table 3]. Constraints that force Trachypachidae or even the remaining Geadephaga with the "higher" aquatic families, as in Beutel's (1995) hypothesis, clearly required higher alignment costs. Moreover, the greatest gap cost (gap:change = 4:1) was the least compatible with any of the constraint scenarios, but the relative increase in cost and hence in the extent of conflict indicated tendencies similar to those under the lower, preferred gap costs. Finally, these searches also indicated the value of the V regions (excluding the central portion), which greatly enhanced the discriminatory power of these analyses, as evident from the larger number of extra steps in the unfavored scenarios.

Downloaded from http://sysbio.oxfordjournals.org/ at Zoological institute RAS on March 20, 2015

TABLE 3. Searches for tree alignments under different constraints. The cost of the alignments under the following constraints are tabulated, given the gap costs indicated, with and without the V regions (central portions excluded) A: Adephaga monophyletic; (GT)HA: Geadephaga (incuding Trachypachidae) monophyletic, Hydradephaga monophyletic, and Adephaga monophyletic, (HT)GA: Hydradephaga + Trachypachidae monophyletic, Geadephaga (excluding Trachypachidae) monophyletic, and Adephaga monophyletic; TD: Trachypachidae + Dytiscidae + Hygrobiidae + Noteridae monophyletic, to the exclusion of all other taxa, including Gyrinide and Haliplidae; Beutel: matching most of the family-level relationships of Beutel (1992), with (Gyrinidae (Haliplidae (Geadephaga excluding Trachypachidae), (Trachypachidae (Noteridae (Hygrobiidae, Dytiscidae))))); PH: Polyphaga + Hydradephaga monophyletic.

	Gap: change cost	Minimum cost in unconstrained POY search	A	(GT)HA	(HT)GA	TD	Beutel	PH
			Increase in parsimony tree length in constrained search					
Excl V regions	1:1	1,074	0	2	3	5	4	4
	2:1	1,198	−1	0	0	4	4	3
	4:1	1,376	12	12	11	17	20	18
Incl partial V regions	1:1	1,892	2	−1	12	9	13	0
	2:1	2,138	4	0	9	15	13	14
	4:1	2,579	0	0	9	12	15	2

Analyses with Alignment and Tree Inference Separated

In the analysis of Clustal-aligned matrices, most-parsimonious trees for each of the 10 alignments that excluded gap runs of five or more were found on one to four islands (Maddison, 1991), each island being found between 47 and 578 times. The strict consensus of these is shown in Figure 4. For all matrices the trees showed Adephaga to be monophyletic, and Hydradephaga was monophyletic in all but one of the matrices (that with a 3:2 gap opening:gap extension cost). For six of the alignments, Geadephaga, including Trachypachidae, was monophyletic; in none of the alignments were trachypachids sister to part or all of Dytiscoidea, as has been proposed by several authors (Bell, 1966; Hammond, 1979; Ward, 1979; Roughley, 1981). Details of the monophyly or paraphyly of various taxa are shown for the four primary Clustal alignments in Table 4.

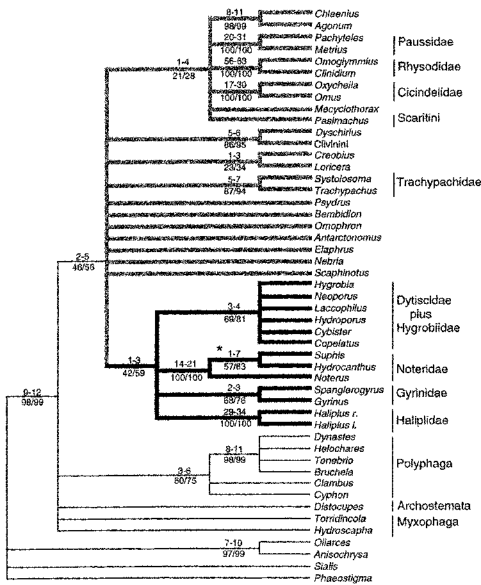

FIGURE 4. Strict consensus tree of the most-parsimonious trees for the four alignments based on a gap opening:gap extension costs of 2:1, 5:1, 7:2, and 10:5, with characters participating in gap runs of five or more excluded. This is almost the strict consensus tree for the trees of greatest likelihood for those four alignments under the constraint that Adephaga is monophyletic; the only difference is that the likelihood consensus tree does not have relationships within Noteridae resolved (that is, the branch marked with an asterisk is absent). Numbers above a branch indicate the range in Bremer support values for the clade over these four alignments. The first number below the clade is the lowest bootstrap value for that clade over the four alignments, the second number below the clade is the lowest jackknife value.

The 22 most-parsimonious trees for the "by eye" alignment were present in one island found 4,999 times. These trees showed a monophyletic Hydradephaga, with Polyphaga as its sister.

If all characters (including the highly length-variable regions) are included in the analysis, the most-parsimonious trees for all 10 Clustal alignments show Adephaga as paraphyletic; in all but the 20:5 alignment, the two neuropterans, *Anisochrysa* and *Oliarces*, both move within beetles, as sister to various geadephagan groups; in the 20:5 alignment the myxophagan *Torridincola* is within Carabidae. In total, only one to three of the focal groups are monophyletic, broadly confirming the results from the POY analysis that revealed high inconsistency of the alignment-variable regions with the other data.

Maximum likelihood analyses on these Clustal-generated alignments generally resulted in trees that showed many of the features observed in the parsimony reconstructions, as shown for the 5:1 alignment in Figure 5. The heuristic searches for trees of high likelihood revealed multiple SPR peaks. For example, the trees for the 2:1 alignment formed four peaks in the likelihood treescape. One suboptimal peak (–ln L = 11531.325) showed Adephaga, Hydradephaga, and Geadephaga to be monophyletic, with Trachypachidae as sister to the rest of Geadephaga; another suboptimal peak (–ln L = 11531.228) showed Polyphaga as the sister to a monophyletic Hydradephaga, and these two as sister to a monophyletic Geadephaga; the optimal peak (–ln L = 11531.215) showed Polyphaga as sister to a monophyletic Hydradephaga, and Geadephaga was paraphyletic, with Trachypachidae as sister to Polyphaga + Hydradephaga. That these trees of such different shape differ by no more than 0.11 –ln L units is surprising, and speaks to relatively weak support for basal relationships of Adephaga with these alignments.

We further used these alignments in constrained parsimony and maximum likelihood analyses to examine the consequences of various alternative proposals of adephagan relationships. The constraints examined are the same as those discussed in the tree

Downloaded from http://sysbio.oxfordjournals.org/ at Zoological institute RAS on March 20, 2015

TABLE 4. Summary of monophyletic groups for the standard parsimony and maximum likelihood analyses. Tree score is tree length for the parsimony analyses and -ln likelihood for the likelihood analyses. M, monophyletic; P, paraphyletic; U, unresolved. In the maximum likelihood analyses the estimated percentage of characters that are not free to vary ranged between 0.49 and 0.53, and the gamma shape parameter varied from 0.43 to 0.47. The values for the General Time Reversible rate matrix are (in PAUP notation): 2:1 alignment, (1.906 5.1865839 2.589 0.44280579 10.217); 5:1 alignment, (1.852 4.208 2.449 0.6309 8.538); 7:2 alignment, (1.273 3.0786951 1.875 0.0627 6.387); 10:5 alignment, (1.171 2.730 1.854 0.558 5.406).

	Most-parsimonious trees excluding gap runs ≥5				Most-parsimoniouis trees including all characters				Maximum likelihood trees excluding gap runs ≥5			
	2:1[a]	5:1	7:2	10:5	2:1	5:1	7:2	10:5	2:1	5:1	7:2	10:5
Tree score	1,785	1,924	2,268	2,349	4,741	5,561	5,981	6,301	11,531	12,049	13,337	13,688
1 Coleoptera	M	M	M	M	P	P	P	P	M	M	M	M
2 Polyphaga + Adephaga	M	M	M	P	P	P	P	P	M	M	M	M
3 Polyphaga	M	M	M	M	M	P	M	M	M	M	M	M
4 Adephaga	M	M	M	M	P	P	P	P	P	M	P	P
5 Hydradephaga + Trachypachidae	P	U	U	P	P	P	P	P	P	P	P	P
6 Hydradephaga	M	M	M	M	M	M	P	P	M	M	M	P
7 Hydradephaga minus Haliplidae	U	P	U	P	M	P	P	P	P	P	P	P
8 Noteridae + Hygrobiidae + Dytiscidae	P	P	P	M	P	P	P	P	P	P	P	M
9 Hygrobiidae + Dytiscidae	M	M	M	M	M	M	M	M	M	M	M	M
10 Dytiscidae	P	P	P	P	P	P	P	P	P	P	P	P
11 Geadephaga	M	U	U	P	P	P	P	P	P	M	P	P
12 Geadephaga minus Trachypachidae	M	U	M	P	P	P	P	P	P	P	P	P
13 CRPS quartet	P	M	U	U	P	P	P	U	P	P	P	P

[a] Gap costs, indicating the ratio of gap opening cost:gap extension cost values used in the Clustal alignment.

Downloaded from http://sysbio.oxfordjournals.org/ at Zoological institute RAS on March 20, 2015

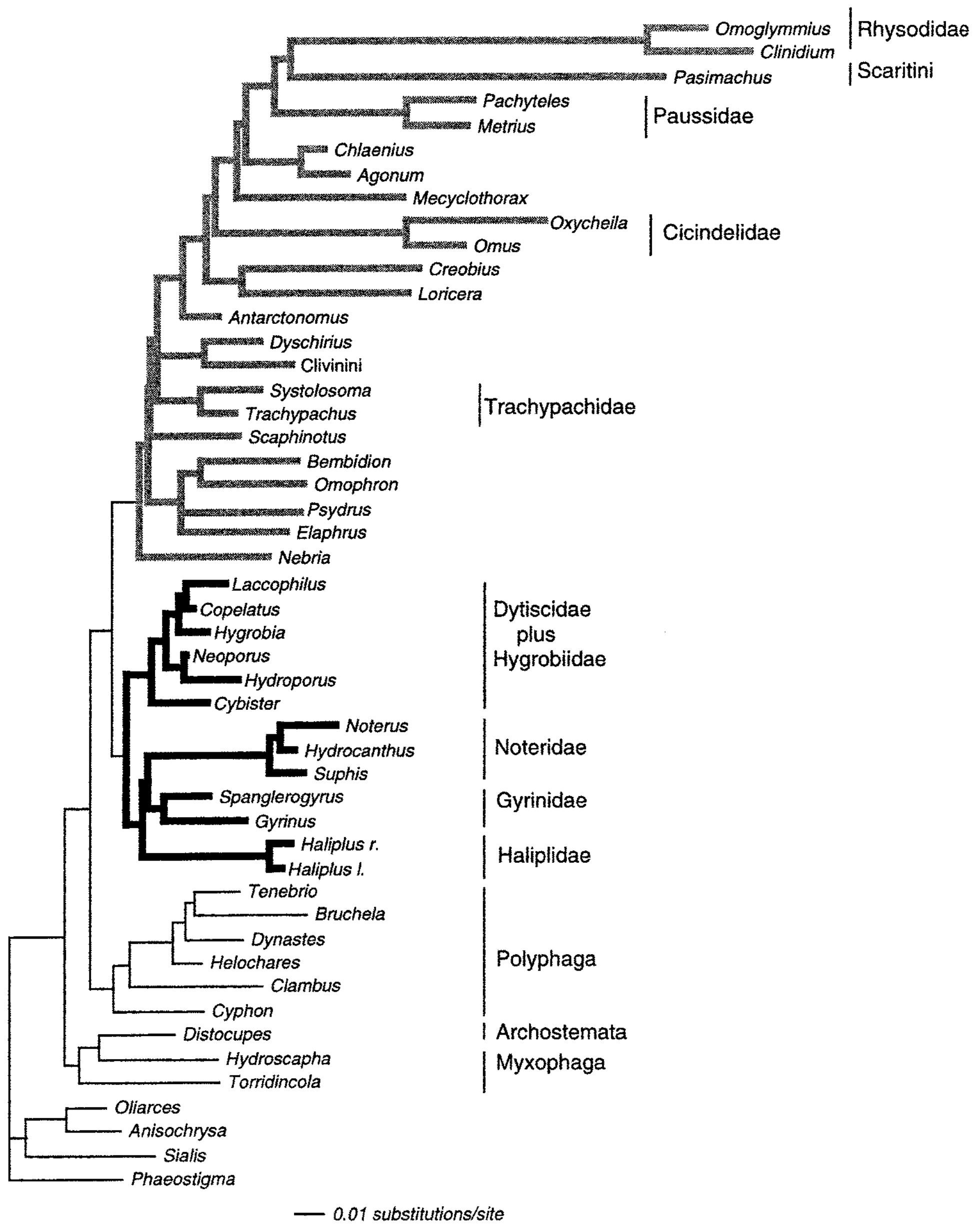

FIGURE 5. Maximum likelihood tree for the 5:1 alignment, with characters participating in gap runs of five or more excluded.

Downloaded from http://sysbio.oxfordjournals.org/ at Zoological institute RAS on March 20, 2015

alignment section, above, and as outlined in Table 3. Consistent with the tree alignment results, these analyses (Table 5) generally support a monophyletic Adephaga and the sister relationship of Geadephaga (including Trachypachidae) and Hydradephaga [the (GT)HA scenario in Table 5], under a wide range of alignment conditions. Hypotheses forcing the Trachypachidae or the remaining Geadephaga with the Dytiscoidea (Beutel, 1995) are less favored under both parsimony and ML methods.

With the neuropteroid outgroups removed, Hydradephaga is monophyletic in most analyses (Table 6). In a few of these analyses, Hydradephaga is not monophyletic because it includes Polyphaga. If Adephaga is constrained to be monophyletic, Hydradephaga is monophyletic in every analysis except the likelihood analysis for the "by eye"

TABLE 5. Comparisons of trees under different constraints. The third and later columns indicate the increase in parsimony tree length or the increase in -ln likelihood for trees with constraints on their shape. Constraints are as defined in Table 3; in addition, TDB defines a backbone constraint in which Trachypachidae + Dytiscidae + Hygrobiidae + Noteridae is monophyletic, with Gyrinidae and Haliplidae removed from the constraint trees (thus, their position in the tree is entirely unconstrained).

Alignment	Unconstrained	A	(GT)HA	(HT)GA	TDB	TD	Beutel	PH
Parsimony analysis				Increase in parsimony tree length				
2:1	1,785	0	0	1	1	9	17	2
5:1	1,924	0	0	0	0	6	14	4
7:2	2,268	0	0	3	3	8	18	3
10:5	2,349	0	1	3	1	8	15	5
"By eye"	1,254	3	3	6	6	9	14	0
ML analysis				Increase in-ln likelihood				
2:1	11,531.21512	0.11	0.11	9.18	9.06	18.60	37.25	0
5:1	12,048.73719	0	0	10.07	10.07	19.80	37.62	0.05
7:2	13,337.25132	6.15	6.15	14.43	12.94	20.84	49.09	0
10:5	13,687.52626	8.93	8.93	13.17	10.55	27.47	43.49	0
"By eye"	9,149.09903	10.62	11.13	21.25	18.25	30.24	45.72	0

matrix (Table 6). If only Adephaga is included in the analysis, then Hydradephaga is always separated fully from Geadephaga, with no mixing.

Monophyly of Hydradephaga, consistent with the results from POY, is thus one of the most notable results of these analyses, although support for this result is not strong (Table 7). The decay indices are low (1, 1, 2, 3, and 1 for the 2:1, 5:1, 7:2, 10:5, and "by eye" analyses, respectively), the parsimony bootstrap values are low (51, 47, 43, 42, and 33, respectively), and the jackknife values are moderate (64, 61, 60, 59, and 42). However, the parametric bootstrapping tests rejected nonmonophyly of Hydradephaga for all five alignments (Table 7). This may seem counterintuitive, given the low values of decay indices, but the simulation studies indicate that even values of only 1–3 for Hydradephaga are unlikely if Hydradephaga is not monophyletic.

Under most alignment parameters Hydradephaga shows a strong affinity with Polyphaga, particularly in the "by eye" alignment and under maximum likelihood optimizations, which renders Hydradephaga paraphyletic in some cases. This is unexpected, and could be in part a spurious result of the GC bias they both share (Fig. 6). Because the method is robust to base composition bias, we used a minimum evolution method with LogDet distances to test for the effect of GC bias. For all five matrices, optimal LogDet minimum evolution trees (found in at least 13 of 50 replicates) showed a monophyletic Adephaga, Hydradephaga, and Geadephaga (including Trachypachidae), suggesting that the monophyly of these groups was not a result of artificial grouping

Downloaded from http://sysbio.oxfordjournals.org/ at Zoological institute RAS on March 20, 2015

TABLE 6. Status of Hydradephaga in optimal trees from various analyses for five alignments. All optimal trees showed either monophyletic Hydradephaga (H), Hydradephaga paraphyletic containing Polyphaga (HP), or Hydradephaga paraphyletic containing monophyletic Geadephaga (HG). No optimal trees showed Trachypachidae + Dytiscoidea monophyletic.

Taxa included	Analysis	No constraints					Adephaga constrained to be monophyletic				
		2:1	5:1	7:2	10:5	"By eye"	2:1	5:1	7:2	10:5	"By eye"
All taxa	Parsimony	H	H	H	H	H	H	H	H	H	H,HG
	MLE	H	H	H	HP	HP	H	H	H	H	HG
	LogDet	H	H	H	H	H	H	H	H	H	H
Coleoptera	Parsimony	H,HP	H,HP	H	H	H	H	H	H	H	H
	MLE	HP	HP	H	H	HP	H	H	H	H	HG
	LogDet	H	H	H	H	H	H	H	H	H	H
Adephaga	Parsimony	H	H	H	H	H	—	—	—	—	—
	MLE	H	H	H	H	H	—	—	—	—	—
	LogDet	H	H	H	H	H	—	—	—	—	—

TABLE 7. Support for monophyly of Hydradephaga for five alignments under different analyses. The first two rows give the bootstrap and decay values for Hydradephaga with parsimony. PB test shows the *P*-value from the parametric bootstrapping test, that is, the probability, estimated by repeatedly simulating data on a tree with Hydradephaga not monophyletic, of having observed a value for (length of shortest tree with Hydradephaga not monophyletic—length of shortest unconstrained tree) as extreme as that empirically observed. MLE decay indicates the likelihood decay index, that is, the difference in −ln likelihood between the maximum likelihood tree with Hydradephaga not monophyletic and that with Hydradephaga monophyletic; a negative value indicates that trees of greatest likelihood have Hydradephaga not monophyletic. LogDet bootstrap is the nonparametric bootstrap value for Hydradephaga in the LogDet analysis.

	2:1	5:1	7:2	10:5	"By eye"
Parsimony bootstrap	51	47	43	42	33
Parsimony decay	1	1	2	3	1
PB test	0.038	0.014	0.018	<0.002	0.004
MLE decay	1.51	2.16	0.54	−4.59	−2.09
LogDet bootstrap	83	70	76	63	72

of taxa having similar base composition. However, the LogDet analysis did eliminate the close association of Polyphaga and Hydradephaga; in the "by eye" alignment, the LogDet analyses (Fig. 6) did not show both groups to be as closely related as in the parsimony analysis. Bootstrap values in

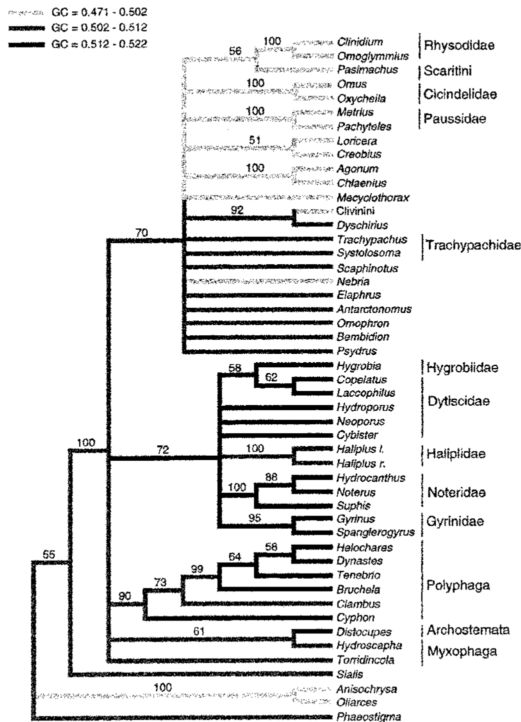

FIGURE 6. Majority-rule tree from 2,000 bootstrap replicates with trees inferred from single-replicate heuristic search for the minimum evolution distance tree by using LogDet distances and the "by eye" matrix. The shade of the branch indicates the fraction of included data that consists of C or G, the darker branches being more GC-rich (see key).

the LogDet analyses for a monophyletic Hydradephaga range from 63 to 83 and thus are much higher than in the parsimony analyses (Table 7), also suggesting that the lack of monophyly of Hydradephaga in some other analyses is an artificial result of the GC bias they share with Polyphaga.

None of these analyses provided evidence for a relationship between Trachypachidae and Dytiscoidea (T + D); in fact, they speak against such a relationship (Table 8). Trees having these two grouped together are six to nine steps longer than trees without, and between 10^8 and 10^{13} times less likely (i.e., –ln likelihood scores are 18.6 and 30.2 more than for trees without this grouping). The parametric bootstrapping tests rejected monophyly of T + D for four of the five alignments. The much higher *P*-value for the 7:2 alignment (0.146; Table 8) than for the remaining alignments ($\leq$0.004) is unexpected but was verified through replication. The result appears to be caused by the unusually long branches estimated using the 7:2 alignment. If the model tree inferred under the 7:2 alignment is used in the simulation, but all other parameters, including branch lengths, are inferred with the "by eye" alignment, then the *P*-value is <0.002; however, if branch lengths instead are estimated with the 7:2 alignment, then the *P*-value is 0.148. This illustrates the sensitivity of parametric bootstrapping to details of the model used.

In summary, both principal types of analysis support similar topologies. Variation of parameters also affects the tree topologies in similar ways under either analysis. Critical is the removal of the highly length-variable central portion of the V2, V4, and V6 regions and the use of a relatively low cost

Downloaded from http://sysbio.oxfordjournals.org/ at Zoological institute RAS on March 20, 2015

TABLE 8. Evidence against the monophyly of Trachypachidae + Dytiscoidea (T + D) for five alignments under different analyses. Parsimony decay indicates the decay index for tree length, that is, the length of the most-parsimonious tree with T + D not monophyletic minus the length with T + D monophyletic. PB test gives the *P*-value from the parametric boostrapping test, that is, the probability, estimated by repeatedly simulating data on a tree with T + D monophyletic, of having observed a value of (length of shortest tree with T + D monophyletic—length of shortest unconstrained tree) as extreme as that empirically observed. MLE decay refers to the decay index for likelihood, that is, the value of −ln likelihood of the maximum likelihood tree without T + D minus the value with T + D.

	2:1	5:1	7:2	10:5	"By eye"
Parsimony decay	−9	−6	−8	−8	−9
PB test	0.004	<0.002	0.146	0.004	<0.002
MLE decay	−18.60	−19.80	−20.84	−27.47	−30.24

for insertions/deletions (gap cost). Under these conditions, a summary tree can be derived that conveys agreement and ambiguity in our conclusions on basal adephagan relationships (Fig. 7), showing the widely supported monophyly of Adephaga, Geadephaga (including Trachypachidae), and Hydradephaga, as well as the uncertainty about relationships within either of the two adephagan branches. Within Hydradephaga, we could identify only the monophyly of Dytiscidae plus Hygrobiidae. Within Geadephaga, the placement of Rhysodidae, Cicindelidae, and Paussidae relative to the Carabidae remains unresolved.

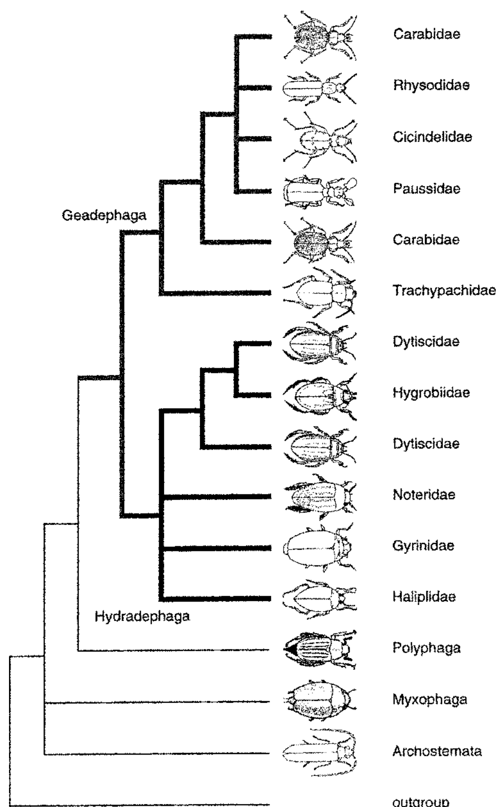

FIGURE 7. Summary presentation of relationships of adephagan families as suggested by our molecular data.

DISCUSSION

Alignment and Phylogenetic Analysis

The two principally different approaches to the alignment of length-variable rRNA sequences—the two-step generation of an alignment matrix followed by a separate tree search, and the one-step simultaneous approach—both resulted in broadly similar conclusions about the phylogenetic relationships of major adephagan groups. This is encouraging, as the simultaneous procedure is a less commonly used and relatively underexplored method. Although satisfying, obtaining similar results with different algorithmic procedures is not necessarily an indication that the trees found are reliable in reflecting phylogeny; the congruence of the trees does not clearly constitute a measure of support but instead may simply reflect data that mislead all methods in the same way (Felsenstein, 1981; but see Kim, 1993).

We argued earlier in this paper, as have others cited previously, that the ideal analysis would involve simultaneous inference of the alignment (i.e., insertions and deletions) and the tree. This argument is based on the presumption that the alignment one would infer differs from one candidate phylogeny to another. This is clearly true for our data. Figure 8 shows the dramatically different alignments of the V2 region when other regions (the hypervariable V4 and V6 regions) are excluded or included in the analysis. As our analysis established, the hypervariable regions contain phylogenetic signal entirely inconsistent

Downloaded from http://sysbio.oxfordjournals.org/ at Zoological institute RAS on March 20, 2015

(a)

```
CCTGCTAACTAGACG-----T---AC--A-----T--ACTCG-AT-TT------TCG
CCTGCTAAATAGACG-----T---AT--A-----A--ACTCG-AT-TT-AACGATCG
CCTGCTAAATAGACG-----T---TT-ATC-AAAA--ACCGGTATCACGAAGGCTTG
CCTGCTAACTAGACG-----T---TT-AT--TAAA--ACCGGTATCACGAAGGCTCG
CCTGCTAACTAGGCG-T---A---CT-----T--T--ACCGGTATCTCGAAGGCTCA
CCTGCTAACTAGGCG-T---T---AC-----T--T--ACCGGTATCTCGAAGGCTCG
CCTGCTAACTAGGCG-T---A---TT-----T--A--ACCGGTATCCTGAAGGCTCG
CCTGCTAACTAGGCG-T---A---CT-----T--A--ACCGGTATCTAGAAGGCTCG
CCTGCTAACTAGGCG-T---A---AC-----T--T--ACCGGTATCTTGAAGGCTTG
CCTGCTAAATAGGCG-T---A---CT-----C--T--ACCGGTATCTCGAAGGCTCA
CCTATTAAATAGACG-T---T---AC-----T--T--ACCGGTATCTCGAAGG-TCG
CCTGTTAACTAGGCG-T---T---AC-----T--T--ACCGGTATCTCGAAGGCTCG
CCTGCTAACTAGGCG-T---A---AC-----T--T--ACCGGTATCTCGAAGGCTCG
CCTGCTAACTAGGCG-T---T----------C--T---CCGGTATCTAGAAGGCTCG
CCTACTAAATAGACG-----T---ACGAT--T--T--TCCGGTATCTTGAAGGCGCA
CCTATTAACTAGGCGTT---AC--AC-----T--T--TCCGGTATCTCGAAGGCCCA
CCTGCTAAATAGGCG-T---A---AC-----T--T--TCCGGTATCTCGAAGGCTCA
CCTGCTAACTAGGCG-T---T---AC-----T--T--ACCGGTATCTCGAAGGCTCG
CCTGCTAAATAGGCG-T---T---A------T--T--ACCGGTATCCTGAAGGC-CG
CCTGCTAACTAGGCG-T---T---AC-----T--T--ACCGGTATCTCGAAGGCTCG
CCTGCTAACTAGGCG-TA--A---CC-AA--A--T--ACCGGTATCTTGAAGGCTCC
CCTACTAAATAGACG-T---T---AC-AC--T--T--ACCAGTATCTCGAAGGCTCA
CCTACTAAATAGACG-T---T---AC-AC--T--T--TCTAGTATCTCGAAGGCTCA
CCTGCTAACTAGGCG-----T---AC-----T--T--TACGGCATCTCGAAGGCTCC
CCTGCTAACTAGGCG-T---T---AC-----T--T--TCCGGCATCTCGAAGGCTCC
CCTGCTAACTAGGCG-T---T---AC-----T--T--TCCGGCATCTCGAAGGCTCG
CCTGCTAACTAGGCG-T---A---AC-----T--T--TCCGGCATCTCGAAGGCTCG
CCTGCTAAATAGGCG-T---T---AC-----T--T--TCCGGCATCTCGAAGGCTCG
CCTGCTAAATAGGCG-T---T---AC-----T--T--TCCGGCATCTCGAAGGCTCG
CCTGCTAAATAGGCG-T---T---AC-----T--T--TCCGGCATCTCGAAGGCTCG
CCTGCTAACTAGGCG-T---A---AC----AT--T--TCCGGTATCTGGAAGGCTCG
CCTGCTAAATAGGCG-T---A---CC-----T--G--ACCGGTATCTGGAAGGCTCG
CCTGCTAACTAGGCG-T---T---AC-----T--T--TCCGGCATCTCGAAGGCTCG
CCTGCTAACTAGGCG-T---A---AC--------T--TCCGACATCCGGAAGGCTCG
CCTGCTAACTAGGCG-T---A---AC--------T--TCCGACATCCGGAAGGCTCG
CCTGCTAACTAGGCG-T---A---AC--------T--TCCGACATCCGGAAGGCTCG
CCTGCTAACTAGGCG-T-------A---------T--TTCGACATCCCAAAGGCCCG
CCTGCTAAATAGGCG-T-------A---------T--TTCGACATCCCAAAGGCCCG
CCTGCTAACTAGGCG-T-------A------T--T--TTCGACATCTCAAAGGCCCG
CCTGCTAACTAGGCG-T-------A------T--T--CTCGACATCTCAAAGGCCCG
CCTGCTAACTAGGCG-T-------A------T--T--TCCGGCATCTCGAAGGGTAG
CCTGCTAAATAGGCG-T-------A------T--T--ACCGGCATCTTAAAGGCTCG
CCTGATAACTAGGCG-T---A-T-AT-----T--AA-ACCGGTATCTCAGAGG-TCG
CCTGCTAACTAGGCG-T-AAA-C-AA-----T--TT-ACCGGTATCCCGAAGGCCCG
CCTGCTAACTAGGCG-T---A-C-AC-----T--AA-TCCGGTATCTCGAAGGCCCG
CCTGCTAACTAGGCG-T---A-CAAT-----T--CA-TCCGGTATCTC-AAGGCCCG
CCTGCTAAATAGACG-T---A-C--T-----T--TA--CCGGCATCTC-AAGGCCCA
CCTGCTAAATAGACG-T---A-C--C-----T--TA--CCGGCATCTC-AAGGCCCA
CCTGCTAAATAGACG-T---A-G-TT-----C--TATTCCGGTATCTC-AAGGCCCG
```

(b)

```
CCTGCTAAATAGACG------T--A-------CAT-A-C-TCGATTT---------TCG
CCTGCTAAATAGACG------T--A-------TAA-A-C-TCGATTTAACG-A---TCG
CCTGCTAAATAGACGT--T--T--A----TC-AAA-A-A-CCGGTATCACGAAGGCTTG
CCTGCTAACTAGACGT--T--T--A----T--TAA-A-A-CCGGTATCACGAAGGCTCG
CCTGCTAACTAGGCG---T--A--C----T--T---T-A-CCGGTATCTCGAAGGCTCA
CCTGCTAACTAGGCG---T--T--A----C--T---T-A-CCGGTATCTCGAAGGCTCG
CCTGCTAACTAGGCG---T--A--T----T--T---A-A-CCGGTATCCTGAAGGCTCG
CCTGCTAACTAGGCG---T--A--C----T--T---A-A-CCGGTATCTAGAAGGCTCG
CCTGCTAACTAGGCG---T--A--A----C--T---T-A-CCGGTATCTTGAAGGCTTG
CCTGCTAAATAGGCG---T--A--C----T--C---T-A-CCGGTATCTCGAAGGCTCA
CCTATTAAATAGACG---T--T--A----C--T---T-A-CCGGTATCTCGAAGG-TCG
CCTGTTAACTAGGCG---T--T--A----C--T---T-A-CCGGTATCTCGAAGGCTCG
CCTGCTAACTAGGCG---T--A--A----C--T---T-A-CCGGTATCTCGAAGGCTCG
CCTGCTAACTAGGCG---T----------T--C---T---CCGGTATCTAGAAGGCTCG
CCTACTAAATAGACG---T--A--CGA--T--T---T-T-CCGGTATCTTGAAGGCGCA
CCTATTAACTAGGCG-T-T--AC-A----C--T---T-T-CCGGTATCTCGAAGGCCCA
CCTGCTAAATAGGCG---T--A--A----C--T---T-T-CCGGTATCTCGAAGGCTCA
CCTGCTAACTAGGCG---T--T--A----C--T---T-A-CCGGTATCTCGAAGGCTCG
CCTGCTAAATAGGCG---T--T--A-------T---T-A-CCGGTATCCTGAAGG-CCG
CCTGCTAACTAGGCG---T--T--A----C--T---T-A-CCGGTATCTCGAAGGCTCG
CCTGCTAACTAGGCG---T--A--A----C--CAA-ATA-CCGGTATCTTGAAGGCTCC
CCTACTAAATAGACG--TT--A-CA----C--T---T-A-CCAGTATCTCGAAGGCTCA
CCTACTAAATAGACG--TT--A-CA----C--T---T-T-CTAGTATCTCGAAGGCTCA
CCTGCTAACTAGGCG------T--A----C--T---T-T-ACGGCATCTCGAAGGCTCC
CCTGCTAACTAGGCG---T--T--A----C--T---T-T-CCGGCATCTCGAAGGCTCC
CCTGCTAACTAGGCG---T--T--A----C--T---T-T-CCGGCATCTCGAAGGCTCG
CCTGCTAACTAGGCG---T--A--A----C--T---T-T-CCGGCATCTCGAAGGCTCG
CCTGCTAAATAGGCG---T--T--A----C--T---T-T-CCGGCATCTCGAAGGCTCG
CCTGCTAAATAGGCG---T--T--A----C--T---T-T-CCGGCATCTCGAAGGCTCG
CCTGCTAAATAGGCG---T--T--A----C--T---T-T-CCGGCATCTCGAAGGCTCG
CCTGCTAACTAGGCG---T--A--A----C-AT---T-T-CCGGTATCTGGAAGGCTCG
CCTGCTAAATAGGCG---T--A--C----C--T---G-A-CCGGTATCTGGAAGGCTCG
CCTGCTAACTAGGCG---T--T--A----C--T---T-T-CCGGCATCTCGAAGGCTCG
CCTGCTAACTAGGCG---T--A--A----C--T---T---CCGACATCCGGAAGGCTCG
CCTGCTAACTAGGCG---T--A--A----C--T---T---CCGACATCCGGAAGGCTCG
CCTGCTAACTAGGCG---T--A--A----C--T---T---CCGACATCCGGAAGGCTCG
CCTGCTAACTAGGCG------T--A-------T---T---TCGACATCCCAAAGGCCCG
CCTGCTAAATAGGCG------T--A-------T---T---TCGACATCCCAAAGGCCCG
CCTGCTAACTAGGCG------T--A-------T---T-T-TCGACATCTCAAAGGCCCG
CCTGCTAACTAGGCG------T--A-------T---T-C-TCGACATCTCAAAGGCCCG
CCTGCTAACTAGGCG------T--A-------T---T-T-CCGGCATCTCGAAGGGTAG
CCTGCTAAATAGGCG------T--A-------T---T-A-CCGGCATCTTAAAGGCTCG
CCTGATAACTAGGCG---T--AT-A----T--T--AA-A-CCGGTATCTCAGAGG-TCG
CCTGCTAACTAGGCG---TAAAC-A----A--T--TT-A-CCGGTATCCCGAAGGCCCG
CCTGCTAACTAGGCG---T--AC-A----C--T--AA-T-CCGGTATCTCGAAGGCCCG
CCTGCTAACTAGGCG---T--A--C--AAT--T---C-ATCCGGTAT-CTCAAGGCCCG
CCTGCTAAATAGACG---T--A--C----T--T---T-A-CCGGCATC-TCAAGGCCCA
CCTGCTAAATAGACG---T--A--C----C--T---T-A-CCGGCAT-CTCAAGGCCCA
CCTGCTAAATAGACG---T-AGT-T----C--T--AT-T-CCGGTATCTC-AAGGCCCG
```

FIGURE 8. Printout from tree searches conducted by POY to illustrate the operation of the program. Tree searches were performed on the full data set of 24 segments of the gene, excluding (a) and including (b) the central portion of region V2, V4, and V6. The segment of the peripheral V4 region shown represents the correspondences of individual nucleotide positions inferred in the tree alignment. Note the differences between both matrices as the result of the exclusion/inclusion of data external to this DNA segment and the changes in the implied homology between bases in either analysis. This printout was obtained by using the *impliedalignment* command.

with the remainder of the gene. The presence of this region implies a different optimal tree and consequently different alignments for the remainder of the data, including the V2 region shown in Figure 8.

Despite our preference for a simultaneous approach, the limitations on available implementations of that method argues for the continued use of the traditional, two-step approach. For example, tests of character changes and explorations of evolutionary models that best fit the data are possible with the traditional approach but are not currently possible with the simultaneous approach. These are useful for assessing the types of changes and their frequency and potentially can help to build models of evolution for rRNA genes, which may be critical for their use in phylogeny inference. Although one can expect that fitting models of evolution for rRNA molecules with their notoriously complex rate variation between taxa and along the length of the sequence will be extremely difficult, exploring the most important parameters is one possible way to advance, given the overwhelming problems with the use of these molecules for inferring relationships of deep relationships of organisms (Philippe and Laurent, 1999).

Exclusion of Alignment-Variable Regions

In no case did we obtain meaningful trees when including the complete length of the rRNA in the analysis. The phylogenetic signal was strongly disturbed by the central parts of the hypervariable regions V4 and V6, which, used on their own, resulted in

Downloaded from http://sysbio.oxfordjournals.org/ at Zoological institute RAS on March 20, 2015

trees with no resemblance to those from the conserved regions or to any of the traditional hypotheses of adephagan relationships. This is not surprising, given that these regions are highly incongruent with the remainder of the gene phylogeny, even at hierarchical levels much lower than those under consideration here, for example, within Cicindelidae (Vogler et al., 1997). Evidence indicates that these hypervariable regions exhibit a strong tendency to converge on similar primary and secondary structure motifs as a result of slippage replication, the predominant mutational process affecting rRNA genes (Hancock and Vogler, 2000). This bias could result in nonhomology of similar sequences in these regions, justifying their exclusion from the phylogenetic analysis.

However, which bases should be excluded or retained is not immediately obvious. Procedures for removing data of questionable homology have been devised that make this decision according to whether or not a position is sensitive to alignment (Gatesy et al., 1993). Although defensible for their objectivity and stringency to deal with the problem of alignment ambiguity, these procedures are impractical for two reasons: many informative sites are removed unnecessarily, and the methods are agnostic with regard to the homology of the base pairs that are removed. One need not assume a priori that a given base does not have a homolog in another taxon if the base is in a region that could be aligned in various ways—although homology may be more difficult to establish in such alignment-variable regions and may require testing different parameters for finding these homologies. An ideal method would not exclude the information in alignment-ambiguous regions but would attempt to infer the homologies by using a tree-alignment approach that can take into account the ambiguities and uncertainties of alignment and thus be able to use the data therein. Without such a procedure, however, another viable approach would be to exclude those regions for which our procedures appear to fail in the attempt to infer homologous sites.

We tested the most length-variable portions of the molecule for homology by assessing congruence of the phylogenetic signal with other parts of the molecule and for the recovery of well-established nodes (Table 2). Alignment-variable regions were either defined by secondary structure (the large stem-loop structures defining the three hypervariable regions V2, V4, and V6), or primary structure (the central portions of the hypervariable regions, which exhibit regions of no discernible sequence similarity bounded by motifs clearly recognizable as similar across all taxa). Only for these latter regions, but not for the adjacent, less length-variable parts of the V regions, could homology not be established on the basis of congruence criteria, and these regions were deemed phylogenetically misleading.

For the two-step analysis separating alignment and tree search, no detailed congruence study of regions of 18S rDNA was conducted. Because of the difficulty in the homology assignment suggested by the POY analysis, we therefore removed some characters from the hypervariable regions. For the Clustal alignments, we removed all characters that were included in runs of five or more consecutive internal gaps in any one sequence, which resulted in a repeatable (although arbitrary) criterion to remove from the matrix most of the complex length variation and thus a large proportion of the alignment-sensitive bases. In this regard, implementation of this procedure is similar to the "cull" procedure of Gatesy et al. (1993) but less stringent; many of the smaller indels, which potentially contribute to the phylogenetic signal, are retained. This should avoid some of the problems of a stringent "cull" procedure but has the disadvantage of relying on a (potentially inappropriate) alignment. For the "by eye" alignment, we removed those sites for which secondary structure locations could not be easily determined by inspection. Both approaches effectively removed similar proportions of the data matrix from the preferred analysis, except the "by eye" alignment omitted notably more data.

The most highly length-variable regions also happen to be those that are "nonhomologous," given the criterion of congruence, and therefore they are removed under both protocols used here. This fact may be due to the peculiarities of the evolutionary process of rRNA genes: Insertion/deletions in the alignment-variable regions apparently are introduced with high frequency and occur at multiple foci in the sequence, leading to alignment ambiguity in the affected regions. Repeated insertion and deletions affecting these regions and the apparently

Downloaded from http://sysbio.oxfordjournals.org/ at Zoological institute RAS on March 20, 2015

slippage-derived nature of this mutational process result in convergent bias of sequence evolution in the alignment-variable regions (Hancock and Vogler, 2000). Nonhomology is a by-product of the slippage-derived insertion/deletions and thus is concentrated in the alignment-ambiguous regions. Removing alignment ambiguity therefore also removes regions of nonhomology.

Suppose the removal of the entire hypervariable regions or of all regions with significant length variation is too conservative and removes regions for which homology could still be assigned. In particular, the use of secondary structural elements has been promoted as a way of unequivocal alignment in cases where similarity based on sequence identity in primary sequence is insufficient to determine correspondences of bases (Kjer, 1995; Hickson et al., 1996). However, the "by eye" alignment and subsequent exclusion of regions resulted in more, not fewer, regions being excluded, because more of the alignment was judged ambiguous by inspection than by the methods used for excluding data with the other alignments. Although a more detailed manual inspection of the sequences might have resulted in discovery of possible secondary structure assignments of more of the data, these manual alignment methods are themselves not ideal. First, they utilize models of sequence evolution that are not stated fully explicitly (the mental [i.e., non-algorithmic] decisions may use implicit data or beliefs) and for which counterevidence may exist (Hancock and Vogler, 2000). Second, just as mental inference of phylogenies could not easily yield the diversity of equally-supported trees made evident by the advent of numerical search methods, mental inference of alignments is unlikely to yield the diversity of equally-supported alignments that is hinted at by the variation in alignments under different parameter values in Clustal. Finally, mental inference does not conduct a full tree alignment as one would like, and to the extent that a phylogeny is considered during the process, the resulting alignment could be biased toward a prior belief about the phylogeny.

Yet, the apparent phylogenetic stability of major secondary structural elements provides a good starting point for building preliminary hypothesis of homology, which can be tested subsequently in the light of the phylogeny derived from it. Our delimitation of 24 discrete regions of clearly recognizable similarity for which homology is accepted a priori (although mostly defined by similarity in primary structure rather than secondary structure) is an implementation of this principle and has already resulted in improvements in the POY search with the cost of trees reduced (8,338 vs. 8,199) when aligned under the same conditions but constraining the alignable bases in this way. Conceivably, further constraints of this kind would result in even shorter trees, indicating that defining such elements of (primary or secondary structure) similarity would indeed improve homology assignment.

Phylogeny of Adephaga

A summary of our conclusions is given in Figure 7. Our analyses weakly support a basal separation of adephagan families in the terrestrial Geadephaga and the aquatic Hydradephaga, largely consistent with Crowson's (1955) proposal. The peculiar family Trachypachidae resides somewhere intermediate between these groups, being favored as sister to Geadephaga in several analyses, but in one case its position is also consistent with a grouping as the sister to Hydradephaga (Fig. 2). Relationships within Geadephaga have been discussed previously (Maddison et al., 1999b), and the inclusion of a wider range of hydradephagan taxa does not change relationships within this group, including the peculiar position of the CRPS quartet of Maddison et al. (1999b), an unexpected grouping of Cicindelidae, Rhysodidae, Paussidae, and Scaritini, in a derived position near the Harpalinae. Despite hopes that these presumably basal geadephagan groups would insert in a more basal position once the remainder of Adephaga had been sampled more thoroughly (Maddison et al., 1999b), this was clearly not the case. Nevertheless, the current position of the CRPS quartet is probably a spurious result of convergent mutational processes producing long insertions in the hypervariable regions and increased frequency of nucleotide changes throughout the gene. The CRPS quartet still represents a serious problem for the molecular systematics of Adephaga and cannot be resolved without additional evidence.

Downloaded from http://sysbio.oxfordjournals.org/ at Zoological institute RAS on March 20, 2015

The monophyly of Hydradephaga is supported under a wide range of parameters. This finding is in contrast to several current hypotheses, in particular those of Beutel (1993, 1995, 1999), who proposed polyphyly of the aquatic families and three independent invasions of the aquatic adaptive zone by Gyrinidae, Haliplidae, and the dytiscoid families, respectively. Although more recent studies (Beutel and Haas, 1996; Beutel, 1999) have found Haliplidae and dytiscoids to be derived from a common ancestor (although paraphyletic), Beutel and coworkers continue to argue that the differences in aquatic lifestyles is evidence against a common aquatic ancestor. Although this cannot be excluded on the basis of phylogenetic reconstruction alone, the monophyly of all aquatic families would be most-parsimoniously interpreted as the evidence for a single colonization of the water, followed by diversification of modes of propagation, feeding style, and body shapes.

Trachypachidae may prove to be of great importance in this context. They exhibit a number of characters that link them with the aquatic lineages, the dytiscoids in particular. Several features that are not considered adaptations for swimming link trachypachid adults and larvae with hydradephagans. Perhaps most striking of these are the wings and wing-folding-associated structures detailed by Hammond (1979). However, several of their shared features are structural elements that are presumably adaptive for swimming, such as the fused metasternum and the presence of a postprocoxal bridge, both of which presumably immobilize the hind and front part of the body, respectively, and thereby improve the efficiency of the stroke exerted by the legs during swimming. In addition, Trachypachidae members roughly fit the overall body plan of aquatic beetles, a smooth perimeter that does not strongly subdivide the thoracic and abdominal parts of the body and a smooth transition of the posterior border of the pronotum and anteriorly truncate elytra. Also, their antenna are entirely glabrous, in common with all aquatics but unlike the pubescent antenna of Geadephaga (Hammond, 1979). Despite their aquatic "adaptations," trachypachids generally inhabit dry habitats, where these traits might be thought to be disadvantageous because they may slow running ability and overall mobility on land. However, D.R.M. has observed *Trachypachus holmbergi* (in Edmonton, Alberta), *T. gibbsii* (in Sonora Pass, California), and *T. slevini* (Moolack Beach, Oregon) rapidly burrowing through very loose, dry substrate (loam, pine duff, and exfoliated clay particles, respectively); the adaptations required for such behavior might not be too dissimilar from that required for "burrowing" through water.

Perhaps the most notable result from our analyses is the strong support against the idea that trachypachids are related to dytiscoid water beetles. Our data speak against the belief that these "aquatic" features of trachypachids are inherited from a recent common ancestor with dytiscoids, but whether these traits are derived independently, or could have been inherited from a more distant common ancestor, is still unclear. Also unclear is whether modern trachypachids had an aquatic ancestor, because the extent of extinction within the group clouds our ability to determine their history. Trachypachidae is represented only by six extant, described species in two genera, the North American *Trachypachus* and the south-temperate *Systolosoma*, which are clearly monophyletic. However, the family was more diverse and widespread in fossil deposits of the Mesozoic (Ponomarenko, 1977). The current diversity of trachypachids thus represents only a small section of a formerly much larger group that could have included species with life styles different from those of the extant taxa, some of which could have been aquatic. The phylogenetic position at the base of the two main Adephagan subgroups suggests that trachypachids represent a third major branch, which today is represented by only a few relictual taxa of widely disjunct distribution. To the extent that the living taxa are a biased sample of past trachypachid diversity, we may not be able to see the richness of their history. Nonetheless, with the information we do have, it is most-parsimonious to presume that trachypachids are not ancestrally aquatic, and that only one major switch between land and water has occurred in the known history of Adephaga.

ACKNOWLEDGMENTS

This work was funded by the Leverhulme Trust (F/696/H, to A.P.V., P.M.H., and V.L.S.) and the National Science Foundation (grant DEB-9420219 to D.R.M.). We thank Ignacio Ribera for discussions on aquatic beetles

Downloaded from http://sysbio.oxfordjournals.org/ at Zoological institute RAS on March 20, 2015

and Ward Wheeler for discussions on sequence alignment and support with POY software. We are also grateful to Karl Kjer, Chris Simon, Richard Olmstead, and an anonymous referee for their help during the reviewing process. Sule Oygur provided advice and specimens in the early stages of this study. We thank the colleagues listed in Table 1 for specimens. We thank Richard Grenyer and Adam Lewin for supporting the Beowulf cluster in the Department of Entomology, The Natural History Museum. The technical support of the DNA Sequencing Facility of the Natural History Museum is gratefully acknowledged.

References

BELL, R. T. 1966. *Trachypachus* and the origin of Hydradephaga. Coleopts Bull. 20:107–112.

BELL, R. T. 1994. Beetles that cannot bite: Functional morphology of the head of adult rhysodines (Coleoptera: Carabidae or Rhysodidae). Can. Entomol. 126:667–672.

BEUTEL, R. G. 1993. Phylogenetic analysis of Adephaga (Coleoptera) based on characters of the larval head. Syst. Entomol. 18:127–147.

BEUTEL, R. G. 1995. The Adephaga (Coleoptera); phylogeny and evolutionary history. Pages 173–218 *in* Biology, phylogeny, and classification of the Coleoptera (J. Pakaluk and S. A. Slipinski, eds.). Museum Zoologii PAN, Warzawa.

BEUTEL, R. G. 1999. Trachypachidae and the phylogeny of Adephaga (Coleoptera). Pages 81–106 *in* Phylogeny and classification of Caraboidea (Coleoptera: Adephaga) (G. E. Ball, A. Casale, and A. Vigna-Taglianti, eds.). Museo Regionale di Scienze Naturali, Atti, Torino, Italy.

BEUTEL, R. G., AND A. HAAS. 1996. Phylogenetic analysis of larval and adult characters of Adephaga (Coleoptera) using cladistic computer programs. Entomol. Scand. 27:197–205.

BEUTEL, R. G., AND R. E. ROUGHLEY. 1988. On the systematic position of the family Gyrinidae (Coleoptera: Adephaga). Z. Zool. Syst. Evolutungsforsch. 26:380–400.

CHALWATZIS, N., J. HAUF, Y. VAN DE PEER, R. KINZELBACH, AND F. K. ZIMMERMANN. 1996. 18S ribosomal RNA genes in insects: Primary structure of the genes and molecular phylogeny of the Holometabola. Ann. Entomol. Soc. Am. 89:788–803.

CROWSON, R. A. 1955. The natural classification of the families of Coleoptera. Nathaniel Loyd & Co., London.

CROWSON, R. A. 1960. The phylogeny of Coleoptera. Annu. Rev. Entomol. 5:111–134.

DRUMMOND, A., AND K. STRIMMER. 2001. PAL: Phylogenetic Analysis Library, version 1.2. http://www.pal-project.org.

FARRELL, B. D. 1998. "Inordinate fondness" explained: Why are there so many beetles? Science 281:555–559.

FELSENSTEIN, J. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. J. Mol. Evol. 17:368–376.

FELSENSTEIN, J. 1988. Phylogenies from molecular sequences: Inference and reliability. Annu. Rev. Genet. 22:521–565.

GATESY, J., R. DESALLE, AND W. C. WHEELER. 1993. Alignment-ambiguous nucleotide sites and the exclusion of systematic data. Mol. Phylogenet. Evol. 2:152–157.

GLADSTEIN, D., AND W. C. WHEELER. 1996. POY. Program and documentation. American Museum of Natural History. Freely available from ftp.amnh.org/people/wheeler/poy.

GOLDMAN, N., J. P. ANDERSON, AND A. G. RODRIGO. 2000. Likelihood-based tests of topologies in phylogenetics. Syst. Biol. 49:652–670.

HAMMOND, P. M. 1979. Wing-folding mechanisms of beetles with special reference to investigations of Adephagan phylogeny. Pages 113–180 *in* Carabid beetles; their evolution, natural history, and classification (T. L. Erwin, G. E. Ball, D. R. Whitehead, and A. Halpern, eds.). W. Junk, The Hague.

HANCOCK, J. M., D. TAUTZ, AND G. A. DOVER. 1988. Evolution of the secondary structures and compensatory mutations of the ribosomal RNAs of *Drosophila melanogaster*. Mol. Biol. Evol. 5:393–414.

HANCOCK, J. M., AND A. P. VOGLER. 2000. How slippage-derived sequences are incorporated into RNA secondary structure. Mol. Phylogenet. Evol. 14:366–374.

HEIN, J. 1989. A new method that simultaneously aligns and reconstructs ancestral sequences for any number of homologous sequences, when the phylogeny is given. Mol. Biol. Evol. 6:649–668.

HICKSON, R. E., C. SIMON, A. COOPER, G. S. SPICER, J. SULLIVAN, AND D. PENNY. 1996. Conserved sequence motifs, alignment, and secondary structure for the third domain of animal 12S rRNA. Mol. Biol. Evol. 13:150–169.

HIGGINS, D. G., J. D. THOMPSON, AND T. J. GIBSON. 1996. Using CLUSTAL for multiple sequence alignment. Methods Enzymol. 266:383–401.

HILLIS, D. M., AND M. T. DIXON. 1991. Ribosomal DNA: Molecular evolution and phylogenetic inference. Q. Rev. Biol. 66:411–453.

HUELSENBECK, J. P., D. M. HILLIS, AND R. NIELSEN. 1996. A likelihood-ratio test of monophyly. Syst. Biol. 45:546–558.

KAVANAUGH, D. 1986. A systematic review of Amphizoid beetles (Amphizoidae: Coleoptera) and their phylogenetic relationships to other Adephaga. Proc. Calif. Acad. Sci. 44:67–109.

KIM, J. 1993. Improving the accuracy of phylogenetic estimation by combining different methods. Syst. Biol. 42:331–340.

KJER, K. M. 1995. Use of rRNA secondary structure in phylogenetic studies to identify homologous positions: An example of alignment and data presentation from the frogs. Mol. Phylogenet. Evol. 4:314–330.

KUKALOVA-PECK, J., AND J. F. LAWRENCE. 1993. Evolution of the hind wing in Coleoptera. Can. Entomol. 125:181–258.

LAKE, J. A. 1994. Reconstructing evolutionary trees from DNA and protein sequences: Paralinear distances. Proc. Natl. Acad. Sci. 91:1455–1459.

LAWRENCE, J. F., AND A. F. NEWTON. 1982. Evolution and classification of beetles. Annu. Rev. Ecol. Syst. 13:261–290.

LAWRENCE, J. F., AND A. F. NEWTON. 1995. Families and subfamilies of Coleoptera (with selected genera, notes, references and data on family-group names). Pages 779–1092 *in* Biology, phylogeny, and classification of Coleoptera (J. Pakaluk and S. A. Slipinski, eds.). Museum i Instytut Zoologii PAN, Warzawa.

Downloaded from http://sysbio.oxfordjournals.org/ at Zoological institute RAS on March 20, 2015

LOCKHART, P. J., M. A. STEEL, M. D. HENDY, AND D. PENNY. 1994. Recovering evolutionary trees under a more realistic model of sequence evolution. Mol. Biol. Evol. 11:605–612.

MADDISON, D. R. 1991. The discovery and importance of multiple islands of most-parsimonious trees. Syst. Zool. 40:315–328.

MADDISON, D. R., M. D. BAKER, AND K. A. OBER. 1999a. A preliminary phylogenetic analysis of 18S ribosomal DNA of carabid beetles (Insecta: Coleoptera). Pages 229–250 *in* Phylogeny and classification of Caraboidea (Coleoptera: Adephaga) (G. E. Ball, A. Casale, and A. Vigna-Taglianti, eds.). Museo Regionale di Scienze Naturali, Atti, Torino, Italy.

MADDISON, D. R., M. D. BAKER, AND K. A. OBER. 1999b. Phylogeny of carabid beetles as inferred from 18S ribosomal DNA (Coleoptera: Carabidae). Syst. Entomol. 24:103–138.

MADDISON, D. R., AND W. P. MADDISON. 2001. Genesis: Models of character evolution. A package of modules for Mesquite version 0.5. http://mesquiteproject.org/mesquite/genesis/genesis. html.

MADDISON, W. P., AND D. R. MADDISON. 2001. Mesquite: A modular system for evolutionary analyses, version 0.96. http://mesquiteproject.org.

MITCHISON, G., AND R. DURBIN. 1995. Tree-based maximal likelihood substitution matrices and hidden Markov models. J. Mol. Evol. 41:1139–1151.

MITCHISON, G. J. 1999. A probabilistic treatment of phylogeny and sequence alignment. J. Mol. Evol. 49:11–22.

PHILIPPE, H., AND J. LAURENT. 1999. How good are deep phylogenetic trees? Curr. Opin. Gen. Dev. 8:315–328.

PONOMARENKO, A. G. 1977. Suborder Adephaga. Pages 1–104 *in* Mesozoic beetles (L. V. Arnoldi, V. V. Zherikin, L. M. Nikritin, and A. G. Ponomarenko, eds.). Trudi Paleontologiceskogo Instituta, Academija Nauk SSSR.

RAMBAUT, A. 1996. Se-Al Sequence alignment editor, version 1.0a1. ftp://evolve.zo.ox.ac.uk/packages/Se-Al/Se-Al10a1.hqx.

RAVI, R., AND J. D. KECECIOGLU. 1998. Approximation algorithms for multiple sequence alignment under a fixed evolutionary tree. Discrete Appl. Math. 88:355–366.

ROUGHLEY, R. E. 1981. Trachyphidae and Hydradephaga (Coleoptera), a monophyletic unit? Pan-Pac. Entomol. 57:273–285.

SANKOFF, D. 1975. Minimal mutation trees of sequences. SIAM J. Appl. Math. 28:35–42.

SANKOFF, D. D., AND R. J. CEDERGREN. 1983. Simultaneous comparison of three or more sequences related by a tree. Pages 253–263 *in* Time warps, string edits, and macromolecules: The theory and practice of sequence comparison (D. Sankoff and J. B. Kruskal, eds.). Addison-Wesley, Reading, Massachusetts.

SANKOFF, D., C. MOREL, AND R. J. CEDERGREN. 1973. Evolution of 5S RNA and the non-randomness of base replacement. Nature (New Biol.) 245:232–234.

SANKOFF, D. D., AND P. ROUSSEAU. 1975. Locating the vertices of a Steiner tree in arbitrary space. Math. Prog. 9:240–246.

SCHWIKOWSKI, B., AND M. VINGRON. 1997. The deferred path heuristic for the generalized tree alignment problem. J. Comput. Biol. 4:415–431.

STEEL, M. A. 1994. Recovering a tree from the Markov leaf colourations it generates under a Markov model. Appl. Math. Lett. 7:19–23.

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

SWOFFORD, D. L. 1999. PAUP*. Phylogenetic Analysis Using Parsimony. Version 4. Sinauer Associates, Sunderland, Massachusetts.

TAUTZ, D., J. M. HANCOCK, D. A. WEBB, C. TAUTZ, AND G. A. DOVER. 1988. Complete sequence of the rRNA genes of *Drosophila melanogaster*. Mol. Biol. Evol. 5:366–376.

THORNE, J. L., AND H. KISHINO. 1992. Freeing phylogenies from artifacts of alignment. Mol. Biol. Evol. 9:1148–1162.

VOGLER, A. P., R. DESALLE, T. ASSMANN, C. B. KNISLEY, AND T. D. SCHULTZ. 1993. Molecular population genetics of the endangered tiger beetle, *Cicindela dorsalis* (Coleoptera: Cicindelidae). Ann. Entomol. Soc. Am. 86:142–152.

VOGLER, A. P., AND D. L. PEARSON. 1996. A molecular phylogeny of the tiger beetles (Cicindelidae): Congruence of mitochondrial and nuclear rDNA data sets. Mol. Phylogenet. Evol. 6:321–338.

VOGLER, A. P., A. WELSH, AND J. M. HANCOCK. 1997. Phylogenetic analysis of slippage-like sequence variation in the V4 rRNA expansion segment in tiger beetles (Cicindelidae). Mol. Biol. Evol. 14:6–19.

WARD, R. D. 1979. Metathoracic wing structures as phylogenetic indicators in the Adephaga (Coleoptera). Pages 181–191 *in* Carabid beetles; their evolution, natural history, and classification (T. L. Erwin, G. E. Ball, D. R. Whitehead, and A. Halpern, eds.). W. Junk, The Hague.

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

WHEELER, W. C. 1998. Alignment characters, dynamic programming and heuristic solutions. Pages 243–251 *in* Molecular approaches to ecology and evolution (R. DeSalle and B. Schierwater, eds.). Birkhäuser Verlag, Basel.

WHITING, M. F., J. C. CARPENTER, Q. D. WHEELER, AND W. C. WHEELER. 1997. The Strepsiptera problem: Phylogeny of the holometabolous insect orders inferred from 18S and 28S ribosomal DNA sequences and morphology. Syst. Biol. 46:1–68.

WOO, S. S., J. JIANG, B. S. GILL, A. H. PATERSON, AND R. A. WING. 1994. Construction and characterization of a bacterial artificial chromosome library of *Sorghum bicolor*. Nucleic Acids Res. 22:4922–4931.

WUYTS, J., P. DE RIJK, Y. VAN DE PEER, G. PISON, P. ROUSSEEUW, AND R. DE WACHTER. 2000. Comparative analysis of more than 3000 sequences reveals the existence of two pseudoknots in area V4 of eukaryotic small subunit ribosomal RNA. Nucleic Acids Res. 28:4698–4708.

Received 16 November 1999; accepted 12 May 2000
Associate Editor: C. Simon

Downloaded from http://sysbio.oxfordjournals.org/ at Zoological institute RAS on March 20, 2015