Molecular Phylogenetics and Evolution 48 (2008) 877–891

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Evaluating nuclear protein-coding genes for phylogenetic utility in beetles

Alexander L. Wild *, David R. Maddison

Department of Entomology, University of Arizona, Tucson, AZ 85721, USA

ARTICLE INFO

Article history:
Received 11 October 2007
Revised 29 April 2008
Accepted 20 May 2008
Available online 27 May 2008

Keywords:
Gene sampling
Beetle evolution
Coleoptera
Adephaga
Phylogenetics

ABSTRACT

Although nuclear protein-coding genes have proven broadly useful for phylogenetic inference, relatively few such genes are regularly employed in studies of Coleoptera, the most diverse insect order. We increase the number of loci available for beetle systematics by developing protocols for three genes previously unused in beetles (alpha-spectrin, RNA polymerase II and topoisomerase I) and by refining protocols for five genes already in use (arginine kinase, CAD, enolase, PEPCK and *wingless*). We evaluate the phylogenetic performance of each gene in a Bayesian framework against a presumably known test phylogeny. The test phylogeny covers 31 beetle specimens and two outgroup taxa of varying age, including three of the four extant beetle suborders and a denser sampling in Adephaga and in the carabid genus *Bembidion*. All eight genes perform well for Cenozoic divergences and accurately separate closely related species within *Bembidion*, but individual genes differ markedly in accuracy over the older Mesozoic and Permian divergences. The concatenated data reconstruct the test phylogeny with high support in both Bayesian and parsimony analyses, indicating that combining data from multiple nuclear loci will be a fruitful approach for assembling the beetle tree of life.

1. Introduction

Coleoptera, with 350,000 described species, is the most species-rich metazoan clade of comparable age. Such disproportionate diversity would seem to receive a correspondingly large share of phylogenetic attention, but molecular tools for beetle phylogenetics remain limited. In a survey of 106 DNA sequence-based phylogenetic studies on various groups of beetles published since 1995 (BIOSIS previews; Thomson Scientific), 71 relied on a single locus, usually one or a handful of mitochondrial genes, and only 24 studies used nuclear protein-coding genes (A. Wild, pers. obs.). Excluding a recent study that sampled large numbers of expressed sequence tags (Hughes et al., 2006), no study has employed more than six independent loci and the mean number of loci per study is only 1.6.

Limited gene sampling leads to phylogenetic inaccuracy either through systematic errors associated with small amounts of data, or through errors stemming from incongruence between gene histories and species histories (Hudson, 1992; Page and Charleston, 1997). While some of these issues may be resolved by including more taxa (Graybeal, 1998), empirical studies (e.g., Rokas and Carroll, 2005; Edwards et al., 2007) indicate that a large number of independent loci, perhaps 20 or more, may be needed to accurately recover a fully resolved phylogeny.

The thousands of protein-coding genes in the eukaryotic nuclear genome present the richest untapped source of genetic data for phylogenetic research. These genes show a number of favorable properties for phylogenetic analysis (Wiegmann et al., 2000). They evolve more slowly and are less prone to base-composition bias than mitochondrial markers (Lin and Danforth, 2004), and they typically present fewer alignment issues than ribosomal genes (Danforth et al., 2005). On the other hand, these genes do not always contain reliable priming sites, they can be present in multiple paralogous copies, and they may contain lengthy introns that complicate amplification, alignment, and sequencing (Sanderson and Shaffer, 2002).

The aim of the present study is to increase the number of nuclear protein-coding genes available for beetle systematics. After screening 24 genes for phylogenetic potential, we selected eight of these for sequencing across 31 test taxa. We assess the phylogenetic performance of each gene against a presumably accurate test phylogeny derived from previous studies and verified with the concatenated data. We also report on additional pertinent properties of these genes, including base composition, paralogy, and the presence of introns.

2. Materials and methods

2.1. Gene and taxon sampling

Genes were selected for evaluation from a variety of sources based on utility in other insect groups and level of sequence

* Corresponding author. Address: Department of Entomology, 320 Morrill Hall, University of Illinois at Urbana-Champaign, 505 S Goodwin Avenue, Urbana, IL 61801, USA.
E-mail address: alwild@myrmecos.net (A.L. Wild).

1055-7903/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2008.05.023

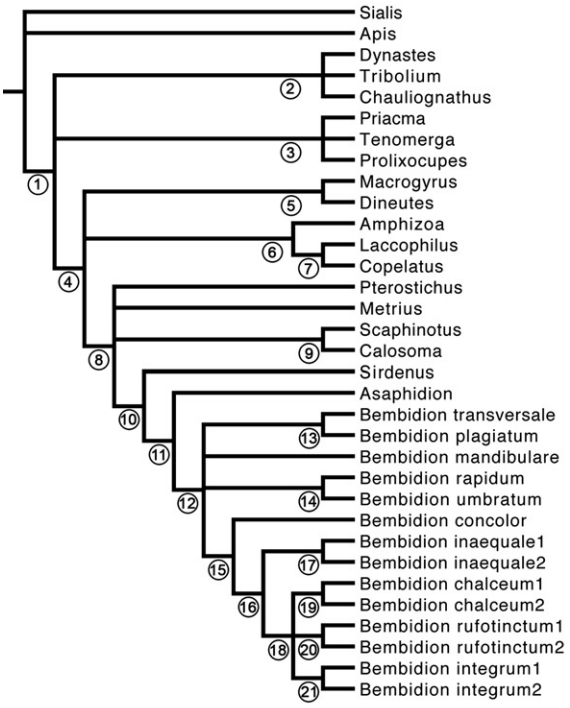
variation. When possible, each gene was located in the *Tribolium castaneum* genome sequence (available at http://www.hgsc.bcm.tmc.edu) and evaluated for presence of paralogs and introns. PCR primers for selected genes were developed in two steps. The initial step created degenerate primers for conserved regions from aligned genome sequences of *Apis mellifera*, *T. castaneum*, *Drosophila melanogaster* and other available insect sequences, and the resulting general insect primers were used on a small set of beetle taxa to generate preliminary beetle sequences. The second step used the sequences generated in the first step to refine primers for use within Coleoptera. The final, refined primers and amplification strategies are given in Table 1.

A test phylogeny for evaluating the performance of each gene was created by selecting 31 beetle taxa and two outgroups whose predicted relationships from previous work (Lawrence and Newton, 1982; Maddison et al., 1999; Beutel and Haas, 2000; Shull et al., 2001; Caterino et al., 2002) is largely uncontroversial and whose divergence dates vary from recent to near the root node for Coleoptera (Fig. 1). The test phylogeny is only partially resolved, with 21 groupings that are well-accepted among coleopterists serve as test clades for the study. Three of the four extant suborders (Adephaga, Archostemata, and Polyphaga) are represented, as are a series of adephagan beetles from several families and a denser sampling within the carabid subtribe Trechitae, especially in the genus *Bembidion*.

Relationships within *Bembidion* are confidently known based upon morphological data (Lindroth, 1963; Maddison, 1993; Maddison, in press) and data from five genes of over 200 species (D. Maddison, unpublished). Included in the *Bembidion* sample are two specimens, each from opposite ends of the geographic range of three closely related species in the subgenus *Pseudoperyphus* (*B. chalceum*, *B. integrum*, and *B. rufotinctum*; Maddison, in press),

Table 1
Primers and amplification strategies

Gene	Strategy	Primer	Sequence (5′–3′)	Source
Alpha-spectrin	Nested PCR of AS1822F/AS2053R amplicon using internal primers AS1831F/AS2053R or AS1822F/AS2030R	AS1822F	TCTCACGARCCNGCNATYCAAGC	This study
		AS1831F	CAAGAAGCTGGNGARAARCTNATGGAYGT	This study
		AS1925F	GCTTACYTNCARTTYATGTGGAARGC	This study
		AS2030R	TCTTCGATTTGDCGGAAYTGRTCYTGCAT	This study
		AS2053R	TCCTCCTCAGCRTTYTCRAACCANGA	This study
Arginine kinase	AK183F/AK939R or AK168F/AK939R for most beetles; AK270F/AK950R for some Adephaga; for difficult taxa, nested PCR of AK168F/AK939R amplicon using internal primers	AK168F	CAGGTTTGGARAAYCACGAYTCYGG	This study
		AK183F	GATTCTGGAGTCGGNATYTAYGCNCCYGAYGC	This study
		AK270F	GGYTTCAAGAAGACYGACAA	This study
		AK592F2	GCNAAYGCNTGCCGYTTCTGGCC	This study
		AK660R	TGATCYTCYTCRTTGCACCA	This study
		AK939R	GCCNCCYTCRGCYTCRGTGTGYTC	This study
		AK950R	TTGTTRGARATGTCRTAGATGCC	This study
CAD	CD439F/CD688R; CD667F/CD851R; CD821F/CD1098R2; for difficult taxa, nested PCR of CD439F/CD1098R2 amplicon using internal primers	CD439F	TTCAGTGTACARTTYCAYCCHGARCAYAC	This study
		CD630F	TCTCTTGGAGGTTTRGGNTCDGGDTTYGC	This study
		CD667F	GGATGGAAGGAAGTDGARTAYGARGT	This study
		CD806F3	TTAYTGYGTTGTNAARATWCCNMGNTGGGA	This study
		CD821F	AGCACGAAAATHGGNAGYTCNATGAARAG	This study
		CD668R	ACGACTTCATAYTCNACYTCYTTCCA	This study
		CD688R	TGTATACCTAGAGGATCDACRTTYTCCATRTTRCA	This study
		CD828R	GCCATTACYTCNCCNACACTYTTCAT	This study
		CD851R	GGATCGAAGCCATTHACATTYTCRTCHACCAT	This study
		CD1098R2	GCTATGTTGTTNGGNAGYTGDCCNCCCAT	This study
Enolase	EN37F/EN731R; for difficult taxa, nested PCR of EN28F2/EN731R amplicon using internal primers	EN28F2	CAAATCTTCGACTCTCGTGGNAAYCC	This study
		EN37F	GACTCTCGTGGNAAYCCNACNGTNGAGGT	This study
		EN730R	AACTCTGANGCNGCNACRTCCATRCC	This study
		EN731R	CTTGTAGAACTCNGANGCNGCNACRTCCAT	This study
PEPCK	PK282F/PK485R; PK330F/PK485R; for difficult taxa, nested PCR of PK282F/PK501R amplicon using internal primers PK282F/PK485R or PK282F/PK485R2	PK282F	GAAGGATGGCTBGCNGARCAYATG	This study
		PK311F	GCTTGTGGTAAGACGAAYYTNGCYATGATG	This study
		PK330F	GAGTGCGTTGGDGAYGACATHGCNTGGATG	This study
		PK485R	GCAGCVGTNGCYTCRCTYCTCAT	This study
		PK485R2	GCAGCAGTNGCTTCRCTHCGCAT	This study
		PK501R	GGACGCATRGCRAANGGRTCGTGCAT	This study
RNA pol II	PL527F/PL625R; PL527F/PL758R; PL709F/PL982R; PL859F/PL1097R; for difficult taxa, nested PCR of PL527F/PL1097R amplicon using internal primers	PL527F	AAYAAACCVGTYATGGGTATTGTRCA	This study
		PL709F	GTCATAGAGGTAATCCARAARGCNCAYAAYATGGA	This study
		PL859F	CGTCTGATCAAGGCTATGGARTCNGTNATGGT	This study
		PL625R	CCCATGACTAGYTCNCCRTGYTCNACCAT	This study
		PL758R	ACGACCATAGCCTTBAGRTTRTTRTAYTC	This study
		PL982R	AARATYTTYTGYACRTTCCARATCAT	This study
		PL1097R	CCAGCGAAGTGGAAVGTRTTNAGBGTCATYTG	This study
Topoisomerase	TP643F/TP932R; TP675F/TP932R; for difficult taxa, nested PCR of TP643F/TP932R amplicon using internal primers	TP643F	GACGATTGGAARTCNAARGARATG	This study
		TP675F	GAGGACCAAGCNGAYACNGTDGGTTGTTG	This study
		TP919R	GTCTCTTTGCGTYTTRTTRTADATYTTYTC	This study
		TP932R	GGWCCDGCATCDATDGCCCA	This study
Wingless	Wg550F/WgAbrZ; for difficult taxa, nested PCR of Wg550F/WgAbrZ amplicon using Wg578F/WgAbR	Wg550F	ATGCGTCAGGARTGYAARTGYCAYGGYATGTC	This study
		Wg578F	TGCACNGTGAARACYTGCTGGATG	Ward and Downie (2005)
		WgAbRZ	CACTTNACYTCRCARCACCARTG	This study
		WgAbR	ACYTCGCAGCACCARTGGAA	Abouheif and Wray (2002)
28S	NLF184-21/LS1041R; for difficult taxa, add 10% DMSO	NLF184-21	ACCCGCTGAAYTTAAGCATAT	Van der Auwera et al. (1994)
		LS1041R	TACGGACRTCCATCAGGGTTTCCCCTGACTTC	D. Maddison (in press)



Number	Clade name	Age (m.y.)	Source
1	Coleoptera	≥ 270-300	Ponomarenko 1995, Grimaldi & Engel 2005, Hunt et al 2007
2	Polyphaga (excl. Scirtoidea)	240	Hunt et al 2007
3	Cupedidae	≥ 220-236	Ponomarenko 1995, Tan et al 2006, Grimaldi & Engel 2005
4	Adephaga	237	Hunt et al 2007
5	Enhydrini (Gyrinidae)	≥ 56-59	Nel, 1989
6	*Amphizoa* + Dytiscidae	150	Hunt et al 2007
7	Dytiscidae	150	Hunt et al 2007
8	Carabidae	215	Hunt et al 2007
9	Carabitae	≥ 48	Ponomarenko & Kirejtshuk 2005
10	Trechitae	≥ 35-48	Lindroth 1974, Ponomarenko & Kirejtshuk 2005
11	Bembiidina	≥ 35-48	Lindroth 1974, Ponomarenko & Kirejtshuk 2005
12	*Bembidion*	≥ 35-48	Lindroth 1974, Ponomarenko & Kirejtshuk 2005
13	*Ocydromus*	?	
14	*Notaphus*	?	
15	*B. concolor* + *B. inaequale* + *Pseudoperyphus*	≥ 7-13	Maddison 1993
16	*B. inaequale* + *Pseudoperyphus*	≥ 7-13	Maddison 1993
17	*B. inaequale*	?	
18	*Pseudoperyphus*	?	
19	*B. chalceum*	?	
20	*B. rufotinctum*	?	
21	*B. integrum*	?	

Fig. 1. Test phylogeny for evaluating the phylogenetic performance of individual genes. The tree is based on well-accepted taxonomic relationships (references cited in text) and confirmed with Bayesian and parsimony analyses of amino acid and nucleotide data from nine nuclear genes. Where known, clade ages are provided (See above-mentioned references for further information).

as well as two specimens of *Bembidion inaequale*, which belongs to the closely related subgenus *Bracteon*. Multiple specimens of closely related species are included to test the efficacy of these genes to distinguish species. Taxon sampling and locality information is provided in Table 2, and GenBank accession numbers are provided in Table 3. A few non-*Pseudoperyphus* taxa were represented by composites of different specimens, as some sequences—particularly for the gene *wingless*—had previously been assembled for other projects. Composite taxa are indicated in Table 2.

2.2. Molecular lab protocol

Specimens were collected into 100% ethanol and/or freeze-killed at −80 °C. All specimens were stored at −20 °C in 100% ethanol, and the ethanol was periodically replaced. Genomic DNA was extracted using the Qiagen DNeasy extraction kit, eluted into TAE buffer and stored at −20 °C. The voucher of *Macrogyrus oblongus* is deposited at Brigham Young University. All other vouchers with "BT" prefixes (Table 2) are stored in the Museum of Comparative Zoology, Harvard University; those with "DRM" prefixes are stored in the collection of DRM, to be eventually deposited in a public collection.

Polymerase Chain Reaction (PCR) was conducted on genomic DNA using the manufacturer's protocol for Takara Ex Taq (http://bio.takara.co.jp). Annealing temperatures varied from 50 to 60 °C, depending on primer melting temperature, and extension times varied from 50 to 90 s, depending on the target fragment length. Some of the more difficult taxa were amplified using a nested or hemi-nested strategy that used internal primers on an initial, longer amplicon (e.g., primers PK330/PK485R used on the amplicon from a PK282/PK485R reaction). More specific protocols and recommendations for each gene are available from the authors. PCR products were purified with a Millipore filter and sequenced in both forward and reverse directions using an Applied Biosystems 3730XL DNA Analyzer.

2.3. Phylogenetic analyses

Alignments were created using Clustal W 1.83 (Thompson and Higgins, 1994) implemented through Mesquite 1.12 (Maddison and Maddison, 2006) and adjusted manually to remove alignment artifacts. Highly variable regions in 28S and introns in the protein-coding genes were excluded from alignments prior to phylogenetic analysis.

Each gene was analyzed in four phylogenetic contexts: Bayesian analysis of nucleotide sequence data, parsimony analysis of nucleotide sequence data, Bayesian analysis of amino acid sequence data, and parsimony analysis of amino acid sequence data. Each analysis was rooted using *A. mellifera* as the outgroup.

Bayesian MCMC searches of the nucleotide and the amino acid data were performed using MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003). The nucleotide data were partitioned by codon position and the substitution model for each partition was selected using the Akaike Information Criterion implemented in MrModelTest (Nylander, 2004). The amino acid data were analyzed in a single partition per gene with the substitution model selected during the analysis by estimating the model as a separate parameter. Nucleotide and amino acid substitution models selected in the analyses are given in Table 6.

Searches were carried out in two concurrent sets of four chains with trees sampled every 1000 generations for nucleotide data and every 100 generations for amino acid data. Each search was run between 5×10^6 and 1×10^7 generations for the nucleotide data and 3×10^5 and 8×10^5 generations for the amino acid data. These searches were long enough to accumulate at least 8000 trees for nucleotide data and 3000 trees for amino acid data following the removal of the "burn-in" period. We determined the burn-in conservatively, as 1/2 the number of generations taken by the search to drop below an average standard deviation of split frequencies of 0.01 of the two concurrent runs. All parameter estimates were evaluated for stability after the burn-in using the program Tracer 1.3 (Rambaut and Drummond, 2004).

Table 2
Taxon sampling and locality information

Taxon	Locality	Extraction code
Sialis sp. (outgroup)	Canada, Nova Scotia, Glace Bay, 46.2°N 60.0°W	BT0019
Apis mellifera Linn. (outgroup)	Genome project	
Tribolium castaneum Herbst	Genome project	
**Tribolium castaneum* Herbst	USA, AZ, Tucson, UA lab colony	BT0052
Chauliognathus opacus LeConte	USA, AZ, 11k S Sierra Vista, 31.4°N 110.3°W	BT0054
Dynastes granti Horn	USA, AZ, 21k WNW Portal, 32.0°N 109.4°W	BT0056
**Dynastes granti* Horn	USA, AZ, Tucson, 32.3°N 110.9°W	DRM0017
Priacma serrata LeConte	USA, ID, 39 mi NE Lowell, 46.4°N 115.1°W	BT0004
**Priacma serrata* LeConte	USA, MT, Hyalite Canyon, 45.6°N 111.0°W	DRM0340
Tenomerga cinerea Say	USA, MA, Jamaica Plain, 42.3°N 71.1°W	DRM0334
Prolixocupes lobiceps LeConte	USA, CA, 2k N Yucca Valley, 34.1°N 116.4°W	BT0021
Calosoma scrutator Fabricius	USA, AZ, Pena Blanca, 31.2°N 111.1°W	DRM2249
Scaphinotus petersi Roeschke	USA, AZ, Mt. Lemmon, 32.4°N 110.8°W	DRM0878
**Scaphinotus petersi* Roeschke	USA, AZ, Mt. Lemmon, 32.4°N 110.8°W	DRM0849
Metrius contractus Eschscholtz	USA, OR, Corvallis, 44.5°N 123.3°W	BT0048
**Metrius contractus* Eschscholtz	USA, CA, Marin Co., Lagunitas Creek	DRM0138
Pterostichus melanarius Illiger	USA, WI, Madison, 43.1°N 89.4°W	BT0062
**Pterostichus melanarius* Illiger	Canada, Alberta, Edmonton, 53.544°N 113.542°W	DRM0357
Amphizoa lecontei Matthews	USA, ID, Boise Nat'l Forest, 44.1°N 115.6°W	BT0064
**Amphizoa lecontei* Matthews	USA, ID, Boise Nat'l Forest, 44.1°N 115.6°W	DRM0790
Laccophilus pictus Laporte	USA, AZ, 24k WNW Nogales, 31.4°N 111.2°W	BT0072
**Laccophilus pictus* Laporte	USA, AZ	DRM0319
Copelatus distinctus Aubé	USA, AZ, 20k ESE Green Valley, 31.8°N 110.8°W	BT0089
**Copelatus chevrolatei renovatus* Guignot	USA, AZ, Buenos Aires National Wildlife Refuge	DRM0783
Dineutes sublineatus Chevrolat	USA, AZ, 24k WNW Nogales, 31.4°N 111.2°W	BT0069
Macrogyrus oblongus Boisduval	Australia, QLD, Mt Mee State Forest, 27.1°S 152.7°E	CO132
Sirdenus grayii Wollaston	Spain, Albacete, Salinas de Pinilla, 38.8°N 2.6°W	DRM1777
Asaphidion yukonense Wickham	Canada, Alberta, Rock Lake, 53.5°N 118.2°W	DRM1897
Bembidion umbratum LeConte	Canada, Alberta, Fort MacLeod, 49.7°N 113.4°W	DRM2166
Bembidion rapidum LeConte	Canada, Ontario, Burlington, 43.3°N 79.8°W	DRM1754
Bembidion mandibulare Solier	Chile, Chiloé, Cucao, 42.6°S 74.1°W	DRM2203
Bembidion transversale Dejean	USA, WY, Laramie, 41.3°N 105.6°W	DRM2157
Bembidion plagiatum Zimmermann	USA, NC, Spruce Pine, 35.9°N 82.1°W	DRM1745
Bembidion concolor Kirby	Canada, Nova Scotia, Moose R. Hwy 2, 45.4°N 64.2°W	DRM1470
**Bembidion concolor* Kirby	Canada, Yukon T., AK Hwy Mile 867.3, 60.6°N 134.7°W	DRM2264
Bembidion inaequale Say	USA, AL, W.B. Bankhead Nat'l For., 34.3°N 87.5°W	DRM0820
Bembidion inaequale Say	USA, OR, Siletz R. E Kernville, 44.9°N 123.9°W	DRM2316
Bembidion chalceum Dejean	Canada, British Columbia, Hope, 49.4°N 121.4°W	DRM1602
Bembidion chalceum Dejean	USA, NC, Huntdale 36.0°N 82.3°W	DRM1539
Bembidion integrum Casey	USA, IA, Des Moines R., Stratford, 42.3°N 93.9°W	DRM1624
Bembidion integrum Casey	USA, MS, Port Gibson, 32.0°N 91.0°W	DRM1630
Bembidion rufotinctum Chaudoir	USA, NC, Burke Co., Linville R., 35.9°N 81.9°W	DRM1556
Bembidion rufotinctum Chaudoir	USA, VT, Quechee Gorge, 43.6°N 72.4°W	DRM1544

Taxa marked with "*" indicate additional samples used in a minority of PCRs to help fill the data matrix, resulting in composite taxa.

Equal-weights parsimony analyses of both nucleotide and amino acid data for each gene were conducted using the heuristic search option in PAUP* 4.0b10 (Swofford, 2002) with *Apis* as the outgroup. Parsimony searches employed TBR branch-swapping and 1000 random stepwise taxon addition sequence replicates. To reduce excessive search times on amino acid data, we used the option to save no more than 25 trees of length greater than or equal to one or two steps longer than the treelength of trees found in a preliminary search. Separate 1000 replicate bootstrap analyses were also conducted for each gene using TBR branch-swapping and four random taxon addition sequence replicates per bootstrap replicate.

The test phylogeny was evaluated against phylogenies constructed using the concatenated data. Analyses of the concatenated data were conducted in Bayesian and parsimony approaches on both nucleotide and amino acid sequences as for the individual genes, except the Bayesian nucleotide data were run for 5×10^6 generations and the amino acid data for 2.5×10^5 generations. Parsimony searches using the concatenated nucleotide data were conducted both with and without third codon positions.

The phylogenetic performance of each gene was measured as accuracy in recovering the test phylogeny. In a Bayesian framework, accuracy was scored by comparing the sum of the observed posterior probabilities for each of the test clades against the maximum possible sum of posterior probabilities. Similarly, the phylogenetic accuracy of each gene in a parsimony framework was scored by comparing the sum of bootstrap percentages for each test clade against the maximum possible sum of bootstrap percentages. As a point of comparison to the protein-coding genes, we sequenced and analyzed the commonly used D1–D3 portion of 28S rDNA.

3. Results

3.1. Gene selection

Sufficient DNA sequence was obtained across the test taxa to evaluate the phylogenetic performance of eight nuclear protein-coding gene fragments. Among these were fragments of several genes previously published in studies of beetle phylogenetics, including *wingless*, phosphoenolpyruvate carboxykinase (PEPCK), and enolase. Genes not yet employed for coleopteran phylogenetics include alpha-spectrin, RNA polymerase II, and topoisomerase I. Two genes that have been in development for beetle phylogenetics in various laboratories are CAD (or rudimentary) and arginine kinase (Jordal, 2007; Maddison, in press; M. Whiting, pers. com.). Fragment length, base composition, the proportion of invariant and parsimony-informative sites, and the presence and length of introns are given for each gene in Table 4. The concatenated data set, including

Table 3
GenBank accession numbers

Taxon	Extr.#	A.spec.	Arg.k.	CAD	Enolase	PEPCK	Pol II	Top.I.	Wg	28S
Sialis sp.	BT0019	EU677476	EU677513	EU677523	EU677548	EU677597	EU677578	EU677626	EU677653	EU677674
Apis mellifera	–	XM_623688.2	NM_001011603.1	XM_393888.2	XM_625053.2	XM_396295.3	XM_623278.1	XM_396203.3	XM_396946.3	AY703551.1
Tribolium castaneum	–	XM_968657.1	XM_966707.1		XM_962466.1	XM_961684.1	XM_968377.1	XM_966102.1	XM_962887.1	
Tribolium castaneum	BT0052			EU677538						EU677678
Chauliognathus opacus	BT0054	EU677478	EU677502	EU677536	EU677549	EU677598	EU677580	EU677628	EU677654	EU677679
Dynastes granti	BT0056	EU677479	EU677503	EU677537		EU677599				
Dynastes granti	DRM0017								EU658921	EU658919
Priacma serrata	BT0004	EU677477	EU677504	EU677524	EU677550					
Priacma serrata	DRM0340								EU677656	EU677676
Tenomerga cinerea	DRM0334		EU677505	EU677525	EU677552	EU677600	EU677579		EU677657	EU677675
Prolixocupes lobiceps	BT0021		EU677506	EU677526	EU677551	EU677601		EU677627	EU677658	EU677677
Calosoma scrutator	DRM2249		EU681831	EU677530	EU677562	EU677606		EU677634	EU677661	to upload
Scaphinotus petersi	DRM0878			EU677531	EU677559		EU677581		EU658922	EU658920
Scaphinotus petersi	DRM0849					EU677607		EU677635		
Metrius contractus	BT0048	EU677497	EU677512	EU677527	EU677577	EU677609		EU677636	EU677663	to upload
Metrius contractus	DRM0138								AF398687	AF398605
Pterostichus melanarius	BT0062	EU677496		EU677533	EU677558	EU677608	EU677582			
Pterostichus melanarius	DRM0357								AF398623	AF398707
Amphizoa lecontei	BT0064		EU677509	EU677532	EU677555		EU677585	EU677631		
Amphizoa lecontei	DRM0790								EU677662	EU677680
Laccophilus pictus	BT0072	EU677501	EU677508	EU677528	EU677553	EU677602	EU677586	EU677629	EU677659	
Laccophilus pictus	DRM0319									EU677681
Copelatus distinctus	BT0089	EU677500	EU677507	EU677529	EU677554	EU677603		EU677630		
C. chevrolatei renovatus	DRM0783								EU677660	EU658918
Dineutes sublineatus	BT0069	EU677498		EU677534	EU677557	EU677604	EU677583	EU677632		EU677683
Macrogyrus oblongus	CO132	EU677499	EU677510	EU677535	EU677556	EU677605	EU677584	EU677633	EU677664	EU677682
Sirdenus grayii	DRM1777	EU677495	EU677514	EU677539	EU677560	EU677611	EU677588	EU677637	EU677665	EU677685
Asaphidion yukonense	DRM1897	EU677494	EU677515	EU677540	EU677561	EU677610	EU677587	EU677638	EU677666	EU677686
Bembidion umbratum	DRM2166	EU677481	EU677520	EU677544	EU677569	EU677618	EU677591	EU677640	EU677671	EU677691
Bembidion rapidum	DRM1754	EU677480	EU677518	EU677543	EU677564	EU677625	EU677593	EU677642	EU677668	EU677690
Bembidion mandibulare	DRM2203	EU677493	EU677519	EU677545	EU677570	EU677621	EU677592	EU677643	EU677669	EU677689
Bembidion transversale	DRM2157	EU677491	EU677517	EU677541	EU677567	EU677617	EU677589	EU677639	EU677667	EU677688
Bembidion plagiatum	DRM1745	EU677492		EU677542	EU677568	EU677624	EU677590	EU677641	EU677670	EU677687
Bembidion concolor	DRM1470		EF648693	EF649387			EF648778		EF649472	EF648833
Bembidion concolor	DRM2264	EU677490			EU677571	EU677612		EU677644		
Bembidion inaequale	DRM0820	EU677487	EU677521	EU677547	EU677573	EU677614	EU677594	EU677647	EU677672	EU677692
Bembidion inaequale	DRM2316	EU677486	EU677522	EU677546	EU677572	EU677613	EU677595	EU677648	EU677673	EU677693
Bembidion chalceum	DRM1602	EU677484	EF648733	EF649427	EU677566	EU677623	EF648806	EU677649	EF649543	EF648882
Bembidion chalceum	DRM1539	EU677485	EF648737	EF649431	EU677574	EU677622	EF648808	EU677650	EF649548	EF648892
Bembidion integrum	DRM1624	EU677482	EF648751	EF649445	EU677575	EU677615	EF648816	EU677646	EF649609	EF649056
Bembidion integrum	DRM1630	EU677483	EF648755	EF649449	EU677576	EU677616	EU677596	EU677645	EF649633	EF649096
Bembidion rufotinctum	DRM1556	EU677488	EF648772	EF649466	EU677565	EU677620	EF648828	EU677651	EF649671	EF649100
Bembidion rufotinctum	DRM1544	EU677489	EF648776	EF649470	EU677563	EU677619	EF648831	EU677652	EF649676	EF649108

Table 4
Gene properties

Gene name	Fragment length	%AT	%Inv	%PI	#Introns (Trechites)	Intron length (Trechites)	#Introns (all taxa)	Intron length (all taxa)	#Introns per taxon	Ease of amplification	Evidence of paralogs
Alpha-spectrin	745	0.542	0.475	0.456	3	50–120	8	50–600	0–4	Hard	No
Arginine kinase	721	0.500	0.538	0.397	0		2	90–300	0–2	Moderate	No
CAD	2020	0.598	0.454	0.482	0		9	50–1200	0–2	Moderate	No
Enolase	661	0.552	0.442	0.484	0		1	50	0–1	Moderate	Yes
PEPCK	582	0.552	0.448	0.462	0		5	50–1000	0–3	Hard	No
RNA pol II	1687	0.578	0.558	0.378	3	50–650	6	50–800	0–4	Hard	Yes
Topoisomerase I	742	0.566	0.468	0.449	0		2	50–80	0–1	Hard	No
Wingless	451	0.502	0.346	0.585	0		2	60–2000	0–1	Moderate	Yes
28S	627	0.456	0.459	0.356	n	n	n	n	n	Easy	No

the additional ribosomal gene 28S, was about 90% complete, comprising sequences from at least five coding genes for each taxon. Holes in the data set were due largely to failed amplifications outside trechites, possibly a result of primer mismatch and lengthy introns. An additional 16 genes not sequenced across the test taxa were explored to varying stages of development (Table 5).

3.2. Evaluation of selected genes

3.2.1. Alpha-spectrin

This gene encodes one of two subunits of the spectrin protein, a key component of the eukaryotic cytoskeleton (Broderick and Winder, 2002). The alpha chain consists largely of sequential repeats of a 106 amino acid motif. Alpha-spectrin has not been widely used in insect systematics, but Regier (2007) has developed protocols for amplifying parts of the gene in arthropods. We explored three different regions across the test taxa and settled on a 745 bp downstream fragment that is less prone to non-target amplification than the upstream fragments.

Amplification of the target fragment is complicated by the presence of up to four introns in some taxa, some longer than 500 bp. For taxa with lengthy introns, we used internal primers to amplify and sequence the fragment in two parts. The fragment was particularly difficult to amplify within Cupedidae, where sequence was

Table 5
Genes screened but not sequenced for this study

Gene	Source	Development	Notes
Abdominal A	Ward and Downie (2005)	First round PCR	Primers failed to amplify across Polyphaga, Adephaga using standard methods
Acetyl-coA carboxylase (ACC)	Regier (2007)	Initial exploration	Excessive amino acid variation across Holometabola; difficult to find conserved priming sites
Arginyl-tRNA synthetase	GenBank; this study	Initial exploration	Excessive amino acid variation across Holometabola; difficult to find conserved priming sites
Achaete-scute homolog (ASH)	Skaer et al. (2002)	Initial exploration	Alignable region across Holometabola very short
Dopa decarboxylase (DDC)	Regier (2007)	Initial exploration	Excessive introns in *Tribolium* genome; amino acid variation looks optimal for phylogenetic performance; paralog in *Tribolium* genome
Decapentaplegic (DPP)	Danforth et al. (2004a,b)	Initial exploration	Excessive amino acid variation across Holometabola; difficult to find conserved priming sites; long introns in *Tribolium* genome
Flightless I	GenBank; this study	Initial exploration	Excessive introns in *Tribolium* genome; amino acid variation appears optimal for phylogenetic performance
Foraging (for)	Ingram et al. (2005)	Initial exploration	Numerous paralogs in *Tribolium* genome
Glycinamide ribotide transformylase (GART)	Danforth et al. (2004a,b)	Initial exploration	Excessive amino acid variation across Holometabola; difficult to find conserved priming sites
Hedgehog		Initial exploration	Insufficient amino acid variation across Holometabola
Hunchback (HB)	Baker and DeSalle (1997)	Initial exploration	Alignable region across Holometabola very short
Phosphatidylinositol 4-kinase	Regier (2007)	First round PCR	Difficulty amplifying outside *Tribolium* and *Agyrtodes* (Leoididae); promising levels of amino acid variation
Phosphogluconate dehydrogenase (PGD)	Regier (2007)	Second round PCR	Primers work well; excessive introns in most beetle groups; amino acid variation looks optimal for phylogenetic performance
Glucose-6-phosphate isomerase (PGI)	Regier (2007)	Initial exploration	Excessive introns in *Tribolium* genome
Glutamyl-prolyl-tRNA synthetase	Regier (2007)	Initial exploration	Excessive amino acid variation across Holometabola; difficult to find conserved priming sites; several introns in *Tribolium* genome
TipTop	GenBank; this study	Initial exploration	Alignable region across Holometabola very short

Table 6
Substitution models used in Bayesian MCMC analyses

Gene	Pos1 model	Pos2 model	Pos3 model	Aamodel
CAD	GTR + I + G	GTR + I + G	HKY + I + G	Jones
Wingless	GTR + I + G	GTR + I + G	GTR + G	Jones
PEPCK	GTR + I + G	SYM + I + G	GTR + I + G	wag
Enolase	GTR + I + G	GTR + I + G	GTR + I + G	wag
Aspec	GTR + G	GTR + G	HKY + I + G	Jones
TopI	GTR + I + G	GTR + I + G	SYM + G	Dayhoff
ArgK	SYM + G	GTR + I	GTR + I + G	Dayhoff
pol2	GTR + G	GTR + I + G	GTR + I + G	Jones
28S	GTR + I + G	—	—	—
Concatenated	GTR + I + G	GTR + I + G	GTR + I + G	Jones

obtained only for *Priacma*. In some taxa, the repeated helical structure of alpha-spectrin resulted in primers binding to non-target sequence within the same gene, or amplifying multiple spectrin regions simultaneously. The variable presence of introns compromises the ability to discern non-target amplification in the gel, and alpha-spectrin remains one of the more problematic genes to amplify in the test set. Alignment of coding sequence using amino acid data is straightforward and the test matrix contained no insertions or deletions.

Alpha-spectrin nucleotide and amino acid data recover basal relationships more reliably than most of the tested gene fragments (Figs. 12 and 13), resolving all deep clades except for Carabidae (Fig. 3). In contrast, alpha-spectrin performs weakly within Trechitae.

3.2.2. Arginine kinase

This gene encodes a phosphotransferase that aids regulation of metabolism in systems with fluctuating energy needs (Kucharski and Maleszka, 1998; Tanaka et al., 2007). Arginine kinase is relatively conserved and has been published previously for phylogenetic inference in Hymenoptera (Kawakita et al., 2003; Banks and Whitfield, 2006). Most beetles lack the downstream intron present in Hymenoptera and other insect groups (Danforth et al., 2004a,b), allowing for the amplification of a contiguous 721 bp downstream exon fragment.

Amplification of arginine kinase is straightforward among the test taxa. A few polyphagans and myxophagans outside of the test set possess introns in differing positions, but introns in the fragment appear to be rare among beetles and when present are normally less than 300 bp in length (A. Wild, unpublished data). Alignment of coding sequence using amino acid data is straightforward and the test matrix contained no insertions or deletions.

Arginine kinase reconstructs deeper divergences slightly more accurately than shallower divergences within the test set (Figs. 4, 12 and 13). The Bayesian reconstruction using nucleotide data matched the test topology over all deep clades except for *Amphizoa* + Dytiscidae (Fig. 4). There is no evidence at present for paralogs of arginine kinase within beetles.

3.2.3. CAD

CAD, also known as *rudimentary*, has been used previously for phylogenetic inference in Diptera (Moulton and Wiegmann, 2004) and in Hymenoptera (Danforth et al., 2006). This gene region encodes several enzymes that catalyze pyrimidine biosynthesis. As with previous insect phylogeny studies, we target the longest of these, the carbamoylphosphate synthetase (CPS) locus. The fragment evaluated here is a 2020 bp segment near the midpoint of the gene, a region free of introns in a majority of beetles screened.

Successful amplification using single round of PCR is sporadic across beetles using the primers developed here, and more consistent results were obtained by amplifying the target fragment in two steps. The first step uses external primers to create a long template amplicon that may or may not be easily visualized on a gel, and the second step uses internal primers for nested or hemi-nested amplifications of the amplicon to generate three slightly overlapping internal fragments. With the exception of a variable region in a roughly 200 bp section at the 5′ end of the fragment, the coding regions align unambiguously and insertions and deletions never span more than two to three amino acid sites.

CAD performs more reliably for shallower than deeper divergences (Figs. 12 and 13), failing to reconstruct a monophyletic Polyphaga and *Amphizoa* + Dytiscidae in the Bayesian analyses of both nucleotide (Fig. 5) and amino acid data. CAD was the highest

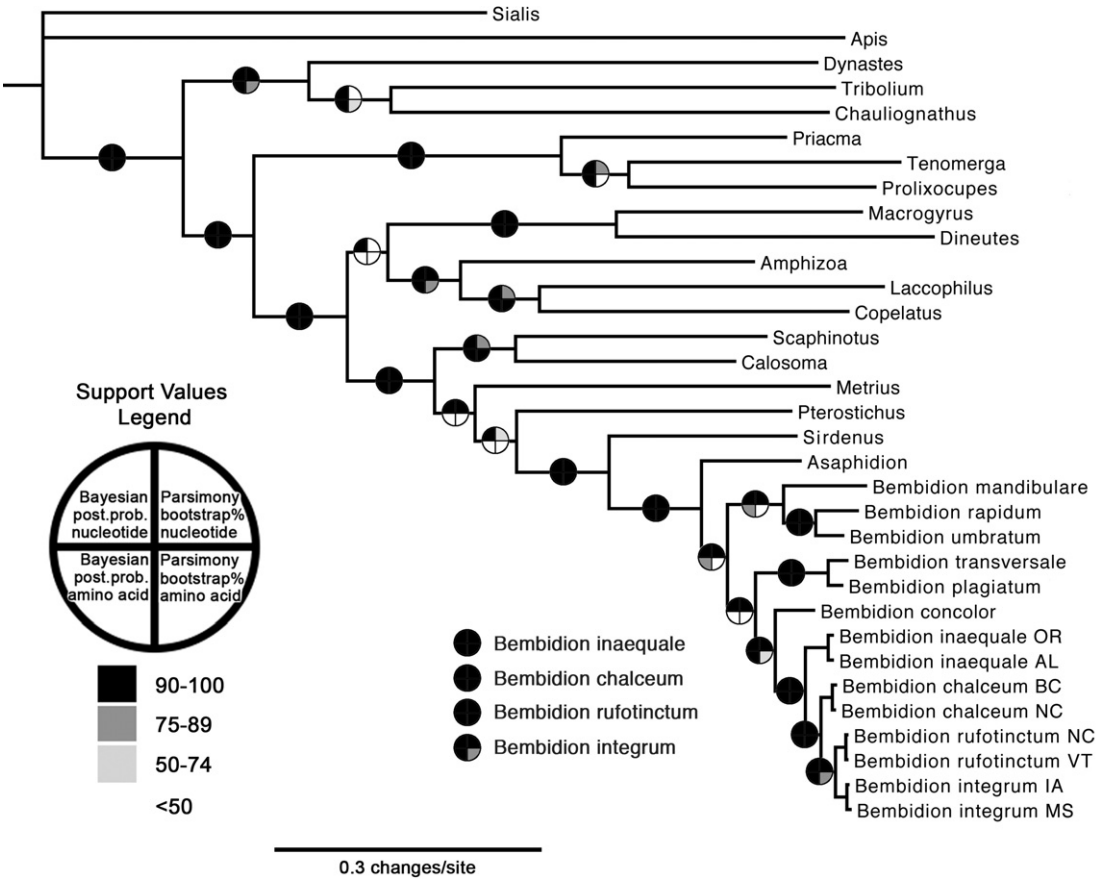


Fig. 2. Phylogram depicting the relationships of the test taxa based on a Bayesian analysis of the concatenated nine-gene, 8236 bp nucleotide data under a four-partition GTR + I + G model. Support is indicated for Bayesian and parsimony analyses of both nucleotide and amino acid data. The support values legend applies to this and all subsequent tree figures.

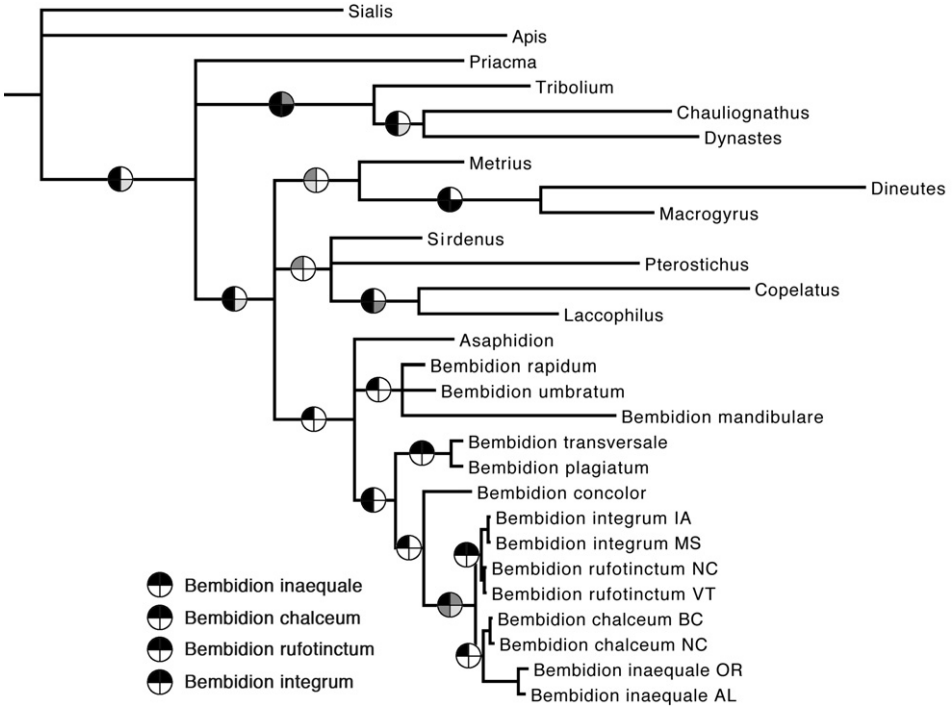


Fig. 3. Phylogram depicting the relationships of the test taxa based on a Bayesian analysis of a 745 bp exonic fragment of alpha-spectrin, analyzed as nucleotides, and partitioned by codon.

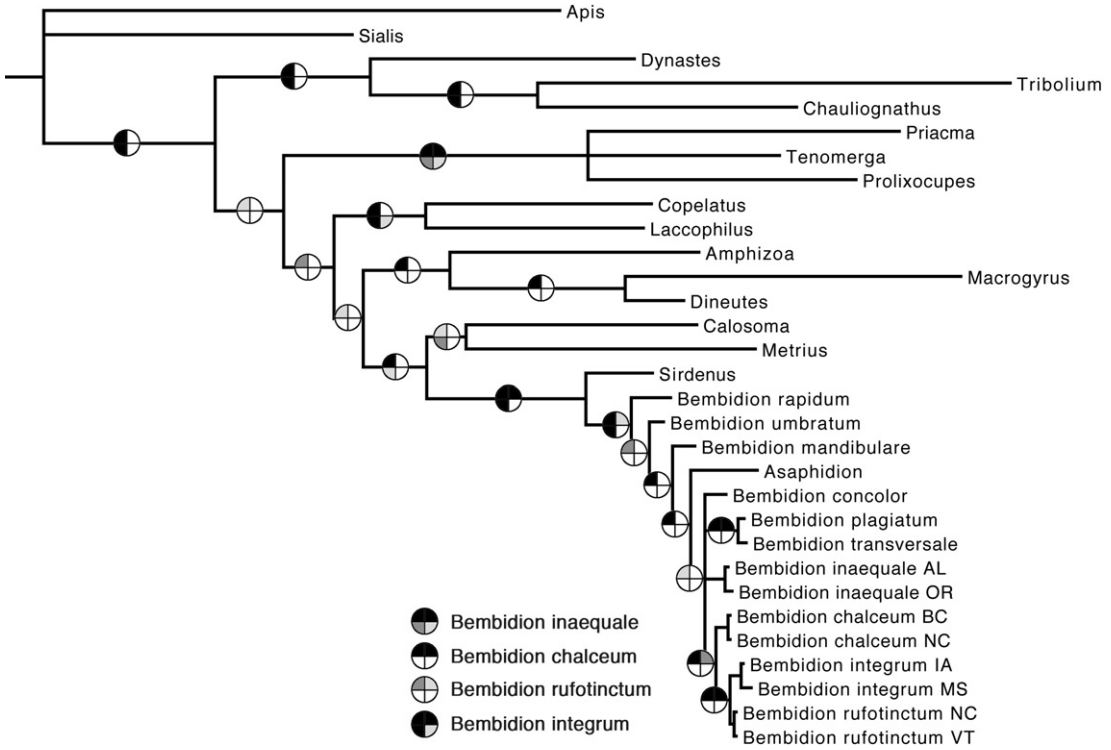


Fig. 4. Phylogram depicting the relationships of the test taxa based on a Bayesian analysis of a 721 bp exonic fragment of arginine kinase, analyzed as nucleotides, and partitioned by codon.

Fig. 5. Phylogram depicting the relationships of the test taxa based on a Bayesian analysis of a 2020 bp exonic fragment of CAD, analyzed as nucleotides, and partitioned by codon.

performing gene fragment among those evaluated here, apparently a result of its strong performance among recent divergences and of the high support values across test clades. There is no evidence at present for paralogs of CAD within beetles.

3.2.4. Enolase

Enolase catalyzes the penultimate step in glycolysis. Systematists have previously employed enolase in the phylogeny of weevils (Farrell et al., 2001; Jordal et al., 2006; Jordal, 2007), but the gene

has not been used extensively for phylogenetic inference in insects. We target a 661 bp fragment largely overlapping but slightly upstream of the region employed by Farrell et al. (2001).

Farrell et al. (2001) reported a second copy of enolase that can be separated in weevils on the basis of an intron. A larger set of beetle taxa than that reported on here show at least three different copies of the gene, some of them comparable in length to the target fragment (A. Wild, unpublished data). One of these copies can be avoided by using EN730R instead of EN731R as the reverse primer, but the near-ubiquitous presence of paralogs remains problematic. The fragment targeted here showed no introns in the test taxa, but enolase gathered from a larger sample of Myxophaga and Polyphaga showed introns in varying locations (A. Wild, unpublished data).

Within Trechitae, nucleotide data recovered the test topology in both Bayesian and parsimony searches (Fig. 6), suggesting that we amplified a single gene copy from that group. However, the fragment performed sporadically at deeper levels, failing to recover the monophyly of Coleoptera, Adephaga, Carabidae, and Dytiscidae. The poor performance of enolase at deeper levels (Figs. 12 and 13) may stem from unintended amplification of paralogs in distantly related taxa.

3.2.5. PEPCK

PEPCK is an enzyme that catalyzes gluconeogenesis and is important in maintaining glucose homeostasis (Hanson and Reshef, 1997). The PEPCK gene has been used previously in the phylogenetics of Lepidoptera (Friedlander et al., 1996), Diptera (Moulton, 2000, 2003), and carabid beetles (Sota and Vogler, 2001). We target a 582 bp fragment largely overlapping that employed by Sota and Vogler (2001) but beginning about 100 bp upstream.

Amplification of PEPCK is unproblematic within trechites, but the frequent and unpredictable presence of lengthy introns in other beetle groups compromises the more general utility of PEPCK. A majority of non-adephagan taxa in the test set have an intron at the 5′ end of the target fragment that can be avoided using the internal forward primers PK311F and PK330F instead of PK282F. Alignment of coding sequence using amino acid data is straightforward, and the test matrix contained no indels longer than a single amino acid residue.

PEPCK performs consistently across the range of divergences tested here (Figs. 7, 12 and 13), although the nucleotide and the amino acid data tested here failed to recover a monophyletic Adephaga, Carabidae, and *Bembidion*. There is no evidence at present for paralogs of PEPCK within beetles.

3.2.6. RNA polymerase II

RNA polymerases catalyze the transcription of RNA from a DNA template (Nikolov and Burley, 1997). In eukaryotes, RNA polymerases are present in three forms, and here we target a 1687 bp fragment from the large subunit of the most studied form, RNA polymerase II. This highly-conserved gene has been used in the systematics of arthropods (Schultz and Regier, 2000) and bees (Danforth et al., 2006). Our target fragment encompasses the region used by Danforth et al. (2006) and extends roughly 900 bases downstream.

Of the eight genes evaluated here, RNA pol II was the most difficult to amplify reliably across the beetles, and nine of the 33 test taxa failed to yield complete sequence after several attempts. Many amplification and sequencing problems may be the result of introns up to 800 bp in length at varying sites within the target fragment. Successful amplification using single round of PCR is sporadic across beetles using the primers developed here, and more consistent results were obtained by amplifying the target fragment in two steps. The first step uses external primers to create a long template amplicon that may or may not be easily visualized on a gel, and the second step uses internal primers for nested or hemi-nested amplifications of the amplicon to generate three slightly overlapping internal fragments. Alignment is straightforward as the amino acid sequence is highly conserved.

RNA pol II performs well within Carabidae, recovering the test phylogeny with high confidence (Figs. 8, 12 and 13). However, at

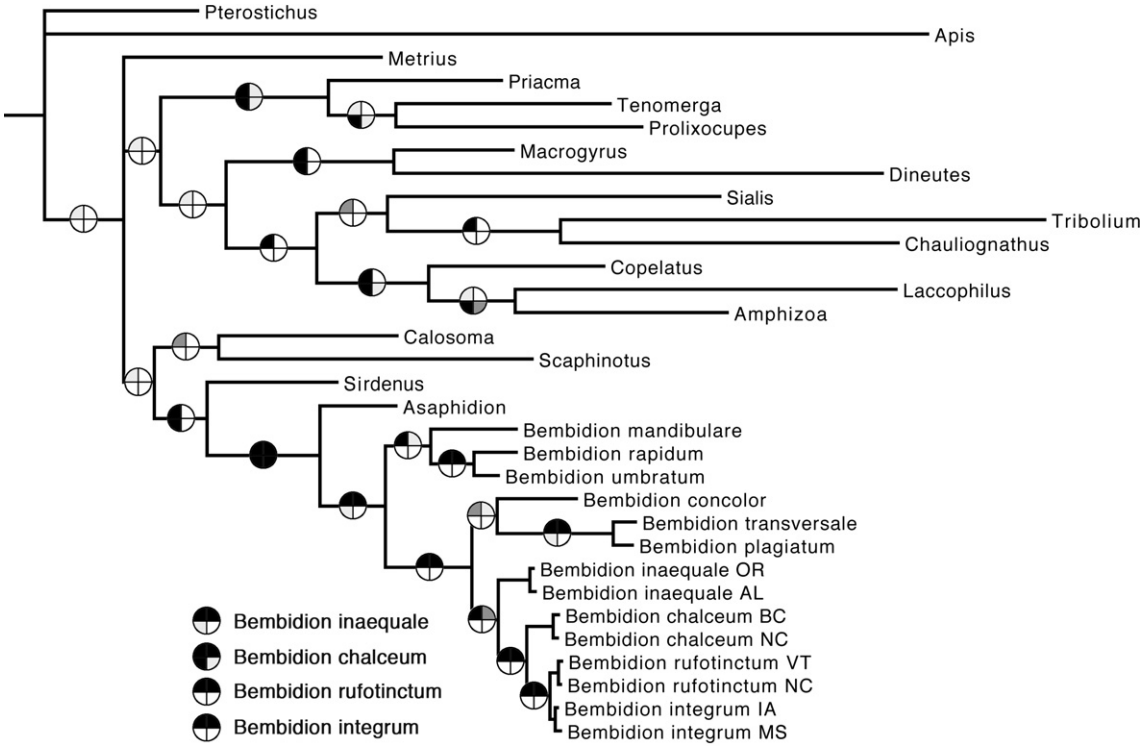


Fig. 6. Phylogram depicting the relationships of the test taxa based on a Bayesian analysis of a 661 bp exonic fragment of enolase, analyzed as nucleotides, and partitioned by codon.

deeper levels the gene fails to recover a monophyletic Coleoptera, Adephaga, Polyphaga, and Gyrinidae. One of the test taxa, *Dynastes*, repeatedly produced a paralog of comparable length, and we have noticed paralogs in several phytophagan beetles (A. Wild, unpublished data; B. Jordal, pers. com.).

3.2.7. Topoisomerase I

Topoisomerases catalyze reactions that relax the tightly-wound helical structure of the DNA molecule, allowing replication and transcription to proceed (Champoux, 2001). Topoisomerase I is one of at least three types of eukaryotic topoisomerase, and it has been used recently in metazoan phylogeny (Delsuc et al., 2006). We targeted a 742 bp fragment at the 3′ end of the gene that is largely free of introns in the test set.

Amplifications within Coleoptera using TP675F/TP932R occasionally produce a second, approximately 2 kb band on the gel. This double amplification can be avoided with a short, 35 s extension for the first few PCR cycles, favoring the shorter product. Most taxa in the test set lack introns in the target segment, but *Prolixocupes* and *Chauliognathus* each have separate, 50–80 bp introns near the 5′ end. Alignment is straightforward, and the coding sequence contains no insertions or deletions in the target fragment.

Topoisomerase I performed well within trechites, recovering the test phylogeny (Fig. 9), but performed unusually poorly for deeper divergences, failing to recover a monophyletic Coleoptera, Adephaga, and Dytiscidae (Figs. 9, 12 and 13). Aside from the occasional double PCR product, there is no evidence of paralogy among the test taxa.

3.2.8. Wingless

The *wingless* protein, a member of the conserved wnt family, is a signaling molecule that establishes segment polarity during embryonic development (Siegfried and Perrimon, 1994.). This gene has been extensively used for phylogenetic inference in various groups of insects (Brower and DeSalle, 1998; Danforth et al., 2004a,b; Brady et al., 2006; Cryan et al., 2004; Sota and Vogler, 2003). We targeted a roughly 450 bp downstream fragment that is commonly used for phylogenetic reconstruction.

About 20% of amplifications using Wg550F/WgAbRZ produce a paralog slightly shorter than the target sequence, a phenomenon that has been noted for other insect groups (Danforth et al., 2004a,b). This paralog blasts to wnt-6 in the *Tribolium* and *Apis* genomes and can often be avoided by using Wg578F as an alternate forward primer. The central 100 bp region of *wingless* is highly variable, containing numerous amino acid insertions and deletions, and is challenging to align across beetles. The upstream and downstream regions present few alignment problems. *Wingless* performs well across the test taxa (Figs. 10, 12 and 13), although the nucleotide and amino acid data in the test set failed to consistently recover a monophyletic Polyphaga or *Bembidion*.

3.3. Additional genes

Genes were rejected for further development for numerous reasons, but the most common were a lack of conserved priming sites across Coleoptera and the presence of problematic paralogs or long introns in the available *Tribolium* genome sequence. Comments on the individual rejected genes are provided in Table 5. Many of these genes may be useful for studies of more recent divergences where clade-specific primers can be designed, or for mRNA-based studies that preclude intron-related complications. Indeed, the level of amino acid variation in several of them, including ACC, flightless I, GART, DDC, and DPP, is similar to that of the high-performing genes CAD, *wingless*, and PEPCK.

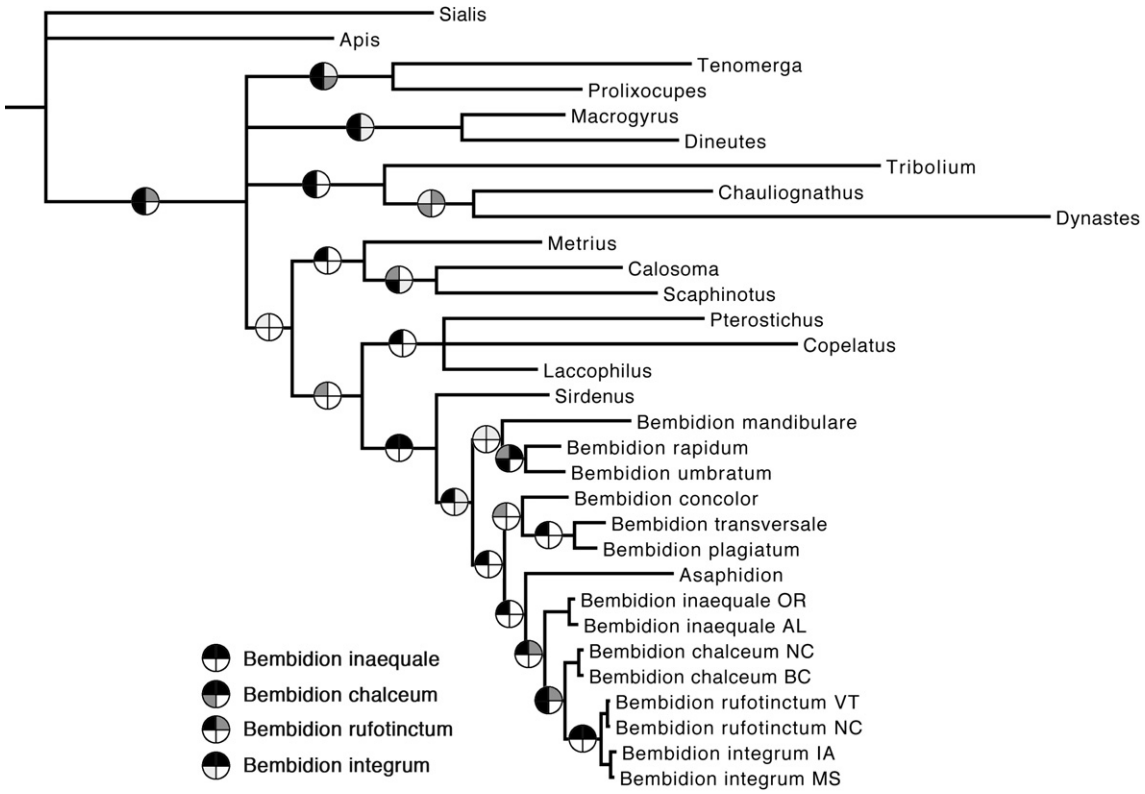


Fig. 7. Phylogram depicting the relationships of the test taxa based on a Bayesian analysis of a 582 bp exonic fragment of PEPCK, analyzed as nucleotides, and partitioned by codon.

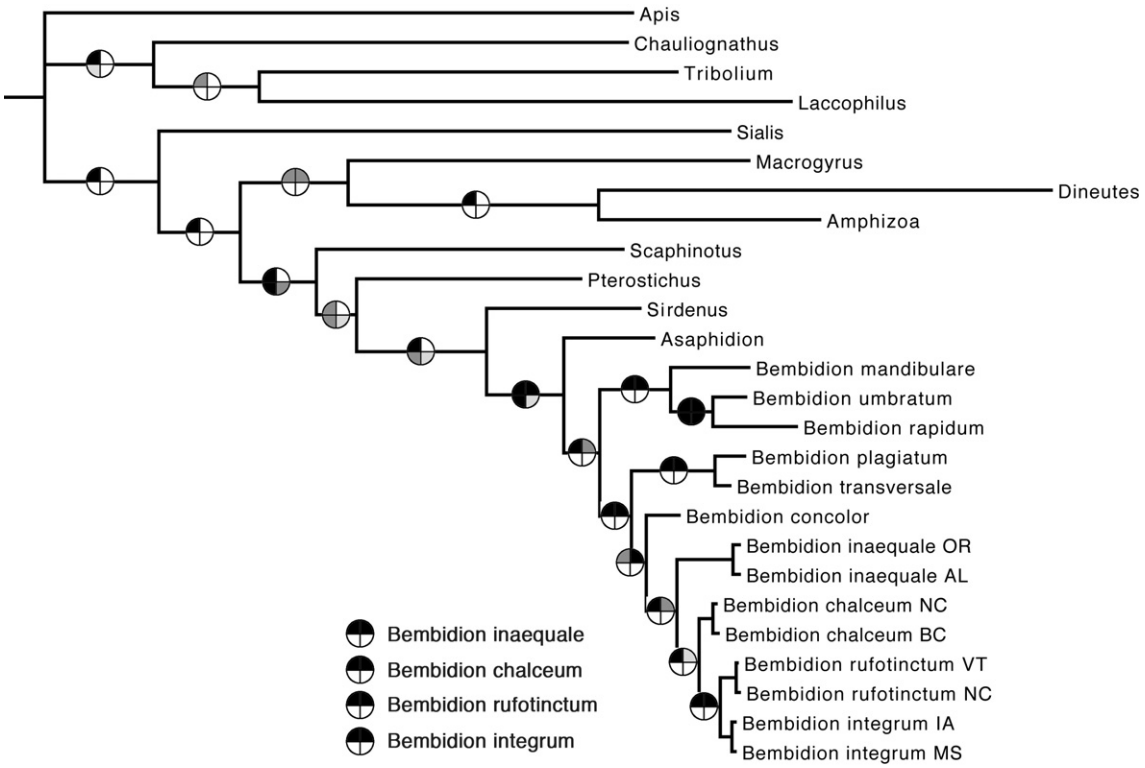


Fig. 8. Phylogram depicting the relationships of the test taxa based on a Bayesian analysis of a 1687 bp exonic fragment of RNA pol II, analyzed as nucleotides, and partitioned by codon.

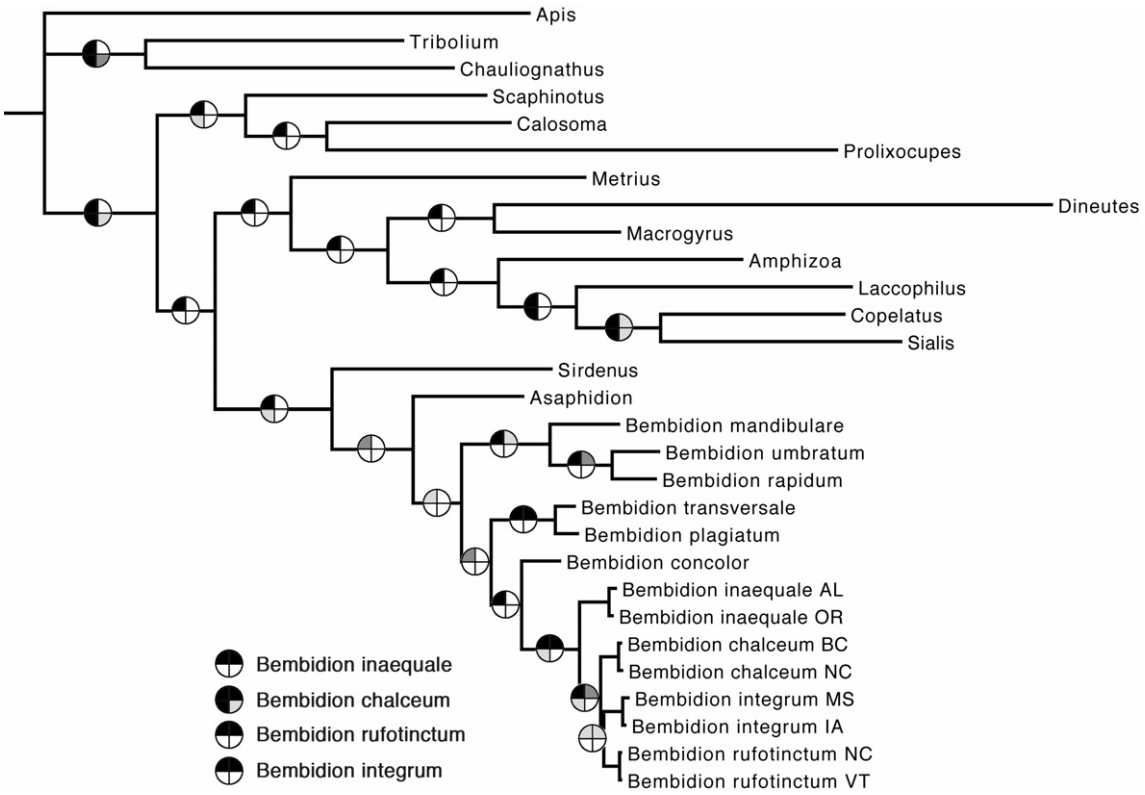


Fig. 9. Phylogram depicting the relationships of the test taxa based on a Bayesian analysis of a 742 bp exonic fragment of topoisomerase I, analyzed as nucleotides, and partitioned by codon.

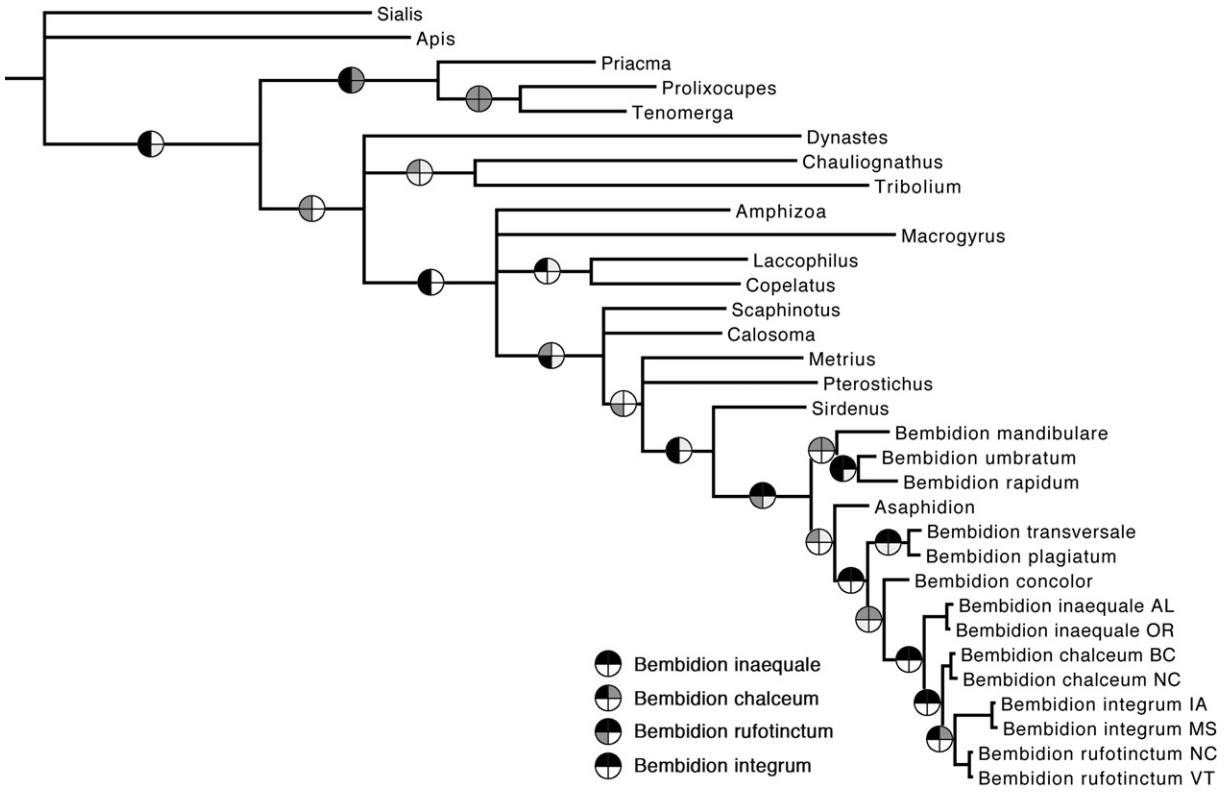


Fig. 10. Phylogram depicting the relationships of the test taxa based on a Bayesian analysis of a 451 bp exonic fragment of *wingless*, analyzed as nucleotides, and partitioned by codon.

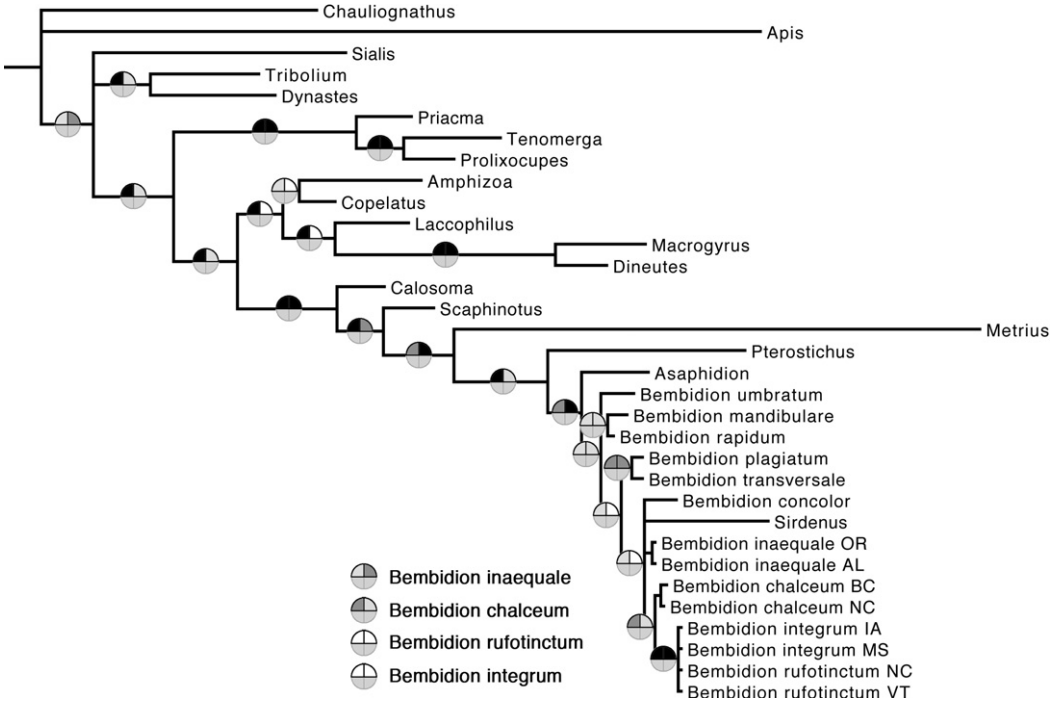
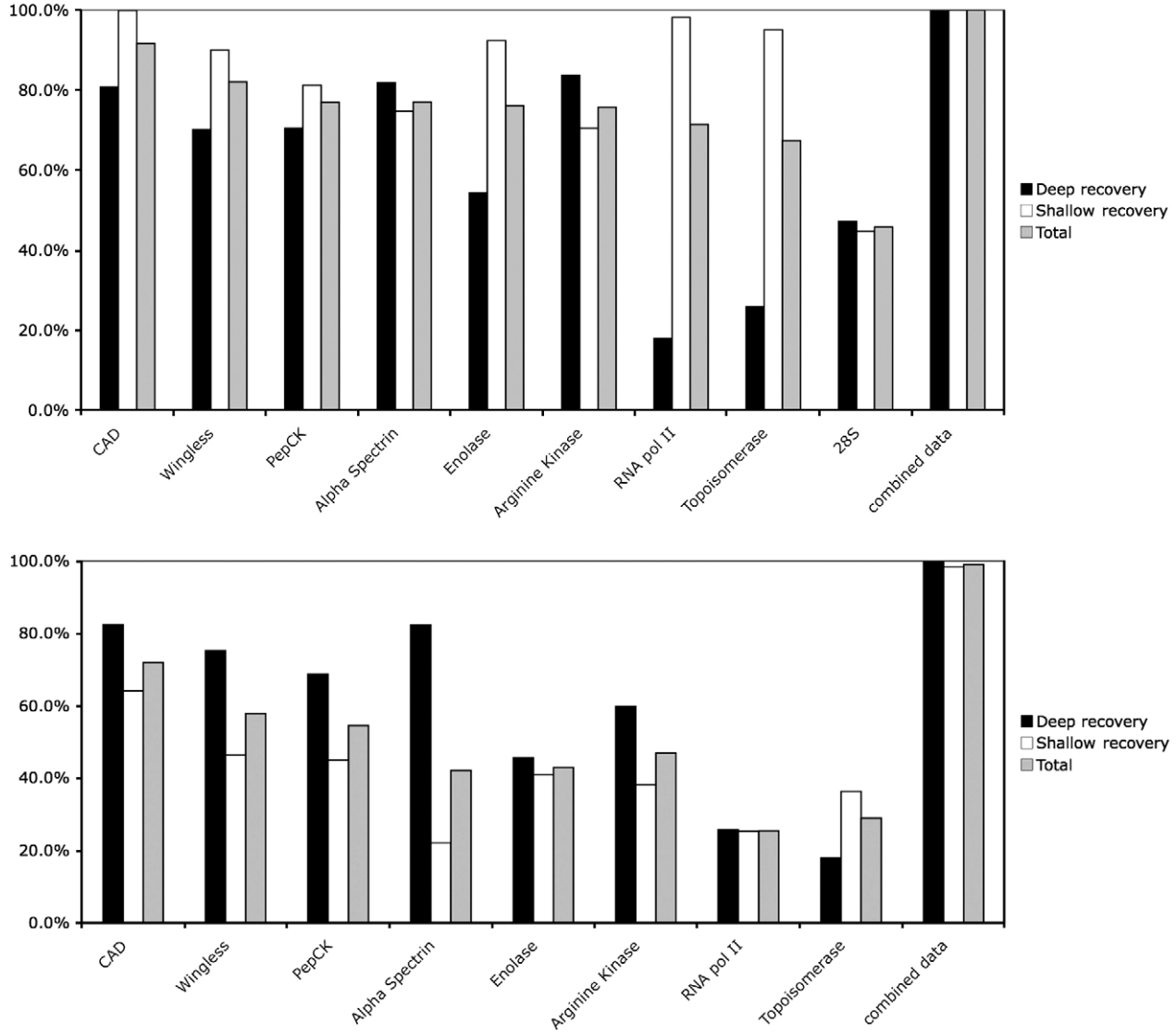


Fig. 11. Phylogram depicting the relationships of the test taxa based on a Bayesian analysis of 627 bp of the ribosomal gene 28S, in the D2–D3 region, after the removal of hypervariable regions.

3.4. Analysis of concatenated data

The test topology was recovered using the concatenated nucleotide data set with Bayesian posterior probabilities of 1.00 and parsimony bootstrap values (with third positions removed) of >75 for each of the 21 test clades (Fig. 2). Phylogenies of the concatenated data based on amino acid sequences are topologically similar to the nucleotide phylogenies and identical over the test clades. The



Figs. 12 and 13. Performance of individual gene fragments measured by the recovery of 21 test clades using Bayesian analyses of nucleotide data (Fig. 12) and amino acid data (Fig. 13). Performance is scored by summing the observed posterior probabilities across the test clades and dividing by the maximum possible posterior probability. "Deep recovery" refers to nine clades outside of Trechitae, while "Shallow recovery" refers to 12 clades within, and including, Trechitae. Gene fragments are sorted from left to right in order of descending performance.

Bayesian analysis of amino acid data differs from the nucleotide analysis only in the composition of some of the non-test clades within Adephaga and Carabidae, while the consensus of 10 equally parsimonious trees based on amino acid data differs from the nucleotide analysis only in that it fails to resolve the sister-group relationship of *Asaphidion* with *Bembidion*, a topology that appears in a subset of the equally most parsimonious trees.

3.5. Comparisons among genes

The nucleotide data from most genes recovered test clades with higher support within trechites than they did at deeper levels (Figs. 3–11, 12). Three genes, alpha-spectrin, arginine kinase and 28S, showed slightly higher accuracy at deeper levels than within trechites. In contrast, the amino acid data for most genes showed greatly reduced accuracy within trechites relative to the nucleotide data (Fig. 13), likely reflecting the lack of divergence in amino acid sequence among younger taxa.

4. Discussion

4.1. Gene performance

The protocols developed here should add several nuclear protein-coding genes to the coleopterists' toolbox. All eight of the protein-coding gene fragments performed well (>70% of maximum possible posterior probability) across the relatively recent divergences within Trechitae. This consistent performance across markers suggests that nuclear protein-coding genes may share generalized properties over shallow divergences. Such an observation is not surprising, as recent phylogenetic signal commonly emerges from synonymous and presumably neutral changes that are robust across a range of models used for phylogeny reconstruction. All else being equal, we recommend that nuclear gene sampling decisions for lower-level studies preferentially take into account ease of amplification and sequencing.

These eight genes also show promise for species delineation. Surprisingly, each protein-coding gene evaluated here recovered all four of the test *Bembidion* species as monophyletic. The only gene that did not, the ribosomal gene 28S, likely failed because the variable regions containing recent phylogenetic signal were excluded prior to analysis because of deep-level alignment issues. 28S performs well when all data are included (D. Maddison, in press).

At deep levels, genes vary strongly in their phylogenetic performance. In the present study, performance at deep levels shows a slight negative correlation with performance at shallow levels (slope = −1.4, R^2 = .36). For example, two of the most weakly performing gene fragments at deep levels, RNA pol II and Topoisomerase I, were among the strongest performers within Trechitae. The reasons behind the heterogeneous performance of genes at deep levels are complex, but may be related to a poor fit between the evolutionary models used and the historical processes behind the observed data. Gene sampling decisions in studies of older divergences will need to pay close attention to amino acid substitution patterns and to the prior track record of candidate genes.

The nine loci targeted here produced resolution and support values greater than those of the 66-locus EST study of Hughes et al. (2006) on a comparable number of beetle taxa. The superior performance of traditional PCR-based methods could involve many factors, including a more complete data matrix, a pre-selection of genes based on performance in other taxa, and a somewhat different taxon sample.

The unpredictable presence and length of introns proved to be the most troublesome obstacle for obtaining sequence data, particularly across older divergences. Introns complicate data gathering for two major reasons. First, they are more prone to length heterozygosity than coding sequence, resulting in messy and frequently unuseable sequencing reactions. Second, they can thwart amplification when longer than a few hundred base pairs, particularly when extension times are short or the Taq polymerase is not designed for long PCR. Introns can be avoided in studies where the focal taxa are young enough that intron sites are conserved, allowing for the design of specific primers to avoid problematic gene regions or to target only those introns that are useful for lower-level questions.

Between intron-related problems and primer mismatch, achieving a matrix more than 75% full for ancient Coleopteran divergences using targeted PCR-based protocols alone is unlikely. Although missing data may not be as problematic as it may seem (Wiens, 2006), systematists may wish to consider a hybrid approach where mRNA-based approaches such as RT-PCR (Regier, 2007) are used in conjunction with traditional PCR-based methods. A mixed approach will allow data to be gathered from the many of specimens whose preservation renders them unavailable for RT-PCR, supplemented by RT-PCR for critical taxa with lengthy introns for which better preserved material can be obtained.

Acknowledgments

Kip Will, Wendy Moore, and Stephen Cameron provided specimens. This research was supported by a National Science Foundation "Assembling the Tree of Life" Grant #EF-0531754 to D.R.M., and is a product of the BTOL (Beetle Tree of Life) project.

References

Abouheif, E., Wray, G.A., 2002. Evolutionary lability of the regulatory gene network underlying the wing polyphenism in ants. Science 297, 249–252.

Baker, R.H., DeSalle, R., 1997. Multiple sources of character information and the phylogeny of Hawaiian drosophilids. Syst. Biol. 46, 674–698.

Banks, J.C., Whitfield, J.B., 2006. Dissecting the ancient rapid radiation of microgastrine wasp genera using additional nuclear genes. Mol. Phylogenet. Evol. 41, 690–703.

Beutel, R.G., Haas, F., 2000. Phylogenetic relationships of the suborders of Coleoptera (Insecta). Cladistics 16, 103–141.

Brady, S.G., Schultz, T.R., Fisher, B.L., Ward, P.S., 2006. Evaluating alternative hypotheses for the early evolution and diversification of ants. Proc. Natl. Acad. Sci. USA 103, 18172–18177.

Broderick, M.J., Winder, S.J., 2002. Towards a complete atomic structure of spectrin family proteins. J. Struct. Biol. 137, 184–193.

Brower, A.V.Z., DeSalle, R., 1998. Patterns of mitochondrial versus nuclear DNA sequence divergence among nymphalid butterflies: the utility of wingless as a source of characters for phylogenetic inference. Insect Mol. Biol. 7, 1–10.

Caterino, M.S., Shull, V.L., Hammond, P.M., Vogler, A.P., 2002. Basal relationships of Coleoptera inferred from 18S rDNA sequences. Zool. Scr. 31, 41–49.

Champoux, J.J., 2001. DNA topoisomerases: structure, function, and mechanism. Annu. Rev. Biochem. 70, 369–413.

Cryan, J.R., Wiegmann, B.M., Deitz, L.L., Dietrich, C.H., Whiting, M.F., 2004. Treehopper trees: phylogeny of Membracidae (Hemiptera: Cicadomorpha: Membracoidea) based on molecules and morphology. Syst. Entomol. 19, 1–14.

Danforth, B.N., Brady, S.G., Sipes, S.D., Pearson, A., 2004a. Single copy nuclear genes recover Cretaceous age divergences in bees. Syst. Biol. 53, 1–18.

Danforth, B.N., Fang, J., Sipes, S., Brady, S.G., Almeida, E., 2004. Phylogeny and molecular systematics of bees (Hymenoptera: Apoidea). Cornell University, Ithaca, NY. Available from: <http://www.entomology.cornell.edu/BeePhylogeny/>.

Danforth, B.N., Lin, C.P., Fang, J., 2005. How do insect nuclear ribosomal genes compare to protein-coding genes in phylogenetic utility and DNA substitution patterns? Syst. Entomol. 30, 549–562.

Danforth, B.N., Fang, J., Sipes, S., 2006. Analysis of family-level relationships in bees (Hymenoptera: Apiformes) using 28S and two previously unexplored nuclear genes: CAD and RNA polymerase II. Mol. Phylogenet. Evol. 39, 358–372.

Delsuc, F., Brinkmann, H., Chourrout, D., Philippe, H., 2006. Tunicates and not cephalochordates are the closest living relatives of vertebrates. Nature 439, 965–968.

Edwards, S.V., Liu, L., Pearl, D.K., 2007. High-resolution species trees without concatenation. Proc. Natl. Acad. Sci. USA 104, 5936–5941.

Farrell, B.D., Sequeira, A.S., O'Meara, B., Normark, B.B., Chung, J.H., Jordal, B.H., 2001. The evolution of agriculture in beetles (Curculionidae: Scolytinae and Platypodinae). Evolution 55, 2011–2027.

Friedlander, T.P., Regier, J.C., Mitter, C., Wagner, D.L., 1996. A nuclear gene for higher level phylogenetics: phosphoenolpyruvate carboxykinase tracks Mesozoic-aged divergences within Lepidoptera (Insecta). Mol. Biol. Evol. 13, 594–604.

Graybeal, A., 1998. Is it better to add taxa or characters to a difficult phylogenetic problem? Syst. Biol. 47, 9–17.

Grimaldi, D., Engel, M.S., 2005. Evolution of the Insects. Cambridge University Press (pp. xv + 755).

Hanson, R.W., Reshef, L., 1997. Regulation of phosphoenolpyruvate carboxykinase (GTP) gene. Annu. Rev. Biochem. 66, 581–611.

Hudson, R.R., 1992. Gene trees, species trees and the segregation of ancestral alleles. Genetics 131, 509–513.

Hughes, J., Longhorn, S.J., Papadopoulou, A., Theodorides, K., de Riva, A., Mejia-Chang, M., Foster, P.G., Vogler, A.P., 2006. Dense taxonomic EST sampling and its applications for molecular systematics of the Coleoptera (beetles). Mol. Biol. Evol. 23, 268–278.

Hunt, T., Bergsten, J., Levkanicova, Z., Papadopoulou, A., St John, O., Wild, R., Hammond, P.M., Ahrens, D., Balke, M., Caterino, M.S., Gómez-Zurita, J., Ribera, I., Barraclough, T.G., Bocakova, M., Bocak, L., Vogler, A.P., 2007. A comprehensive phylogeny of beetles reveals the evolutionary origins of a super-radiation. Science 318, 1913–1916.

Ingram, K.K., Oefner, P., Gordon, D.M., 2005. Task-specific expression of the foraging gene in harvester ants. Mol. Ecol. 14, 813–818.

Jordal, B.H., Emerson, B.C., Hewitt, G.M., 2006. Apparent sympatric speciation in ecologically similar herbivorous beetles facilitated by multiple colonisations of an island. Mol. Ecol. 15, 2935–2947.

Jordal, B.H., 2007. Reconstructing the phylogeny of Scolytinae and close allies: major obstacles and prospects for a solution. In: Bentz, B., Cognato, A., Raffa, K. (Eds.), Proceedings from the Third Workshop on Genetics of Bark Beetles and Associated Microorganisms. Proc. RMRS-P-45. Fort Collins, CO., US. Department of Agriculture, Forest Service, Rocky Mountain Research Station, pp. 3–9.

Lin, C.P., Danforth, B.N., 2004. How do insect nuclear and mitochondrial gene substitution patterns differ? Insights from Bayesian analyses of combined data sets. Mol. Phylogenet. Evol. 30, 686–702.

Kawakita, A., Sota, T., Ascher, J.S., Ito, M., Tanaka, H., Kato, M., 2003. Evolution and phylogenetic utility of alignment gaps within intron sequences of three nuclear genes in bumble bees (*Bombus*). Mol. Biol. Evol. 20, 87–92.

Kucharski, R., Maleszka, R., 1998. Arginine kinase is highly expressed in the compound eye of the honey bee, *Apis mellifera*. Gene 211, 343–349.

Lawrence, J.F., Newton Jr., A.F., 1982. Evolution and classification of beetles. Annu. Rev. Ecol. Syst. 13, 261–290.

Lindroth, C.H., 1963. The ground-beetles (Carabidae, excl. Cicindelinae) of Canada and Alaska. Part 3. Opuscula Entomologica Supplementum XXIV, 201–408.

Maddison, D.R., 1993. Systematics of the Holarctic beetle subgenus *Bracteon* and related *Bembidion* (Coleoptera: Carabidae). Bull. Mus. Comp. Zool. 153, 143–299.

Maddison, D.R., in press. Systematics of the North American beetle subgenus *Pseudoperyphus* (Coleoptera: Carabidae: *Bembidion*) based upon morphological, chromosomal, and molecular data. Ann. Carnegie Mus.

Maddison, D.R., Baker, M.D., Ober, K.A., 1999. Phylogeny of carabid beetles as inferred from 18S ribosomal DNA (Coleoptera: Carabidae). Syst. Entomol. 24, 103–138.

Maddison, W.P., Maddison, D.R., 2006. Mesquite: a modular system for evolutionary analysis. Version 1.12. Available from: <http://mesquiteproject.org>.

Moulton, J.K., 2000. Molecular sequence data resolves basal divergences within Simuliidae (Diptera). Syst. Entomol. 25, 95–113.

Moulton, J.K., 2003. Can the current molecular arsenal adequately track rapid divergence events within Simuliidae (Diptera)? Mol. Phylogenet. Evol. 27, 45–57.

Moulton, J.K., Wiegmann, B.M., 2004. Evolution and phylogenetic utility of CAD (rudimentary) among Mesozoic-aged Eremoneuran Diptera (Insecta). Mol. Phylogenet. Evol. 31, 363–378.

Nel, A., 1989. Les Gyrinidae fossiles de France (Coleoptera). Ann. Soc. Ent. Fr. (N.S.) 25, 321–330.

Nikolov, D.B., Burley, S.K., 1997. RNA polymerase II transcription initiation: a structural view. Proc. Natl. Acad. Sci. USA 94, 15–22.

Nylander, J.A.A., 2004. MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University. Available from: <http://www.abc.se/~nylander/>.

Page, R.D.M., Charleston, M.A., 1997. From gene to organismal phylogeny: reconciled trees and the gene tree/species tree problem. Mol. Phylogenet. Evol. 7, 231–240.

Ponomarenko, A.G., 1995. The geological history of beetles. pp. 155–171 in Pakaluk and Slipinski 1995. Biology, Phylogeny, and Classification of Coleoptera. Papers Celebrating the 80th Birthday of Roy A. Crowson. Muzeum i Instytut Zoologii PAN, Warszawa.

Ponomarenko, A.G., Kirejtshuk, A.G., 2005. Taxonomic list of fossil beetles of suborders Cupedina, Carabina and Scarabaeina (Part 1). Available from: <http://www.zin.ru/ANIMALIA/COLEOPTERA/eng/paleosy0.htm>.

Rambaut, A., Drummond, A.J., 2004. Tracer v1.3. Available from: <http://beast.bio.ed.ac.uk/Tracer>.

Regier, J.C., 2007. Protocols, Concepts, and Reagents for preparing DNA sequencing templates. Version 5/31/07. Available from: <www.umbi.umd.edu/users/jcrlab/PCR_primers.pdf>.

Rokas, A., Carroll, S.B., 2005. More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. Mol. Biol. Evol. 22, 1337–1344.

Ronquist, F., Huelsenbeck, J.P., 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Sanderson, M.J., Shaffer, H.B., 2002. Troubleshooting molecular phylogenetic analysis. Annu. Rev. Ecol. Syst. 33, 49–72.

Schultz, J.W., Regier, J.C., 2000. Phylogenetic analysis of arthropods using two nuclear protein-encoding genes supports a crustacean+hexapod clade. Proc. R. Soc. Lond. B 267, 1011–1019.

Shull, V., Vogler, A.P., Baker, M.D., Maddison, D.R., Hammond, P.M., 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. Syst. Biol. 50, 945–969.

Siegfried, E., Perrimon, N., 1994. *Drosophila* wingless: a paradigm for the function and mechanism of Wnt signaling. Bioessays 16, 395–404.

Skaer, N., Pistillo, D., Gibert, J.M., Lio, P., Wulbeck, C., Simpson, P., 2002. Gene duplication at the achaete-scute complex and morphological complexity of the peripheral nervous system in Diptera. Trends Genet. 18, 399–405.

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

Sota, T., Vogler, A.P., 2003. Reconstructing species phylogeny of the carabid beetles *Ohomopterus* using multiple nuclear DNA sequences: heterogeneous information content and the performance of simultaneous analysis. Mol. Phylogenet. Evol. 26, 139–154.

Swofford, D.L., 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4b10. Sinauer Associates, Sunderland, MA.

Tanaka, K., Ichinari, S., Iwanami, K., Yoshimatsu, S., Suzuki, R., 2007. Arginine kinase from the beetle *Cissites cephalotes* (Olivier). Molecular cloning, phylogenetic analysis and enzymatic properties. Insect Biochem. Mol. Biol. 37, 338–345.

Tan, J., Ren, D., Shih, C., 2006. First record of fossil *Priacma* (Coleoptera: Archostemata: Cupedidae) from the Jehol Biota of western Liaoning, China. Zootaxa 1326, 55–68.

Thompson, J.D., Higgins, D.G., Gibson T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting,position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680. Available from: <http://www.ebi.ac.uk/Tools/clustalw/>.

Van der Auwera, G., Chapelle, S., De Wachter, R., 1994. Structure of the large ribosomal subunit RNA of *Phytophthora megasperma*, and phylogeny of the oomycetes. FEBS Lett. 338, 133–136.

Ward, P.S., Downie, D.A., 2005. The ant subfamily Pseudomyrmecinae (Hymenoptera: Formicidae): phylogeny and evolution of big-eyed arboreal ants. Syst. Entomol. 30, 310–335.

Wiegmann, B.M., Mitter, C., Regier, J.C., Friedlander, T.P., Wagner, D.M., Nielsen, E.S., 2000. Nuclear genes resolve mesozoic-aged divergences in the insect order Lepidoptera. Mol. Phylogenet. Evol. 15, 242–259.

Wiens, J.J., 2006. Missing data and the design of phylogenetic analyses. J. Biomed. Inform. 39, 34–42.