Available online at www.sciencedirect.com
SCIENCE DIRECT
Molecular Phylogenetics and Evolution 39 (2006) 358–372
MOLECULAR PHYLOGENETICS AND EVOLUTION
www.elsevier.com/locate/ympev

Analysis of family-level relationships in bees (Hymenoptera: Apiformes) using 28S and two previously unexplored nuclear genes: CAD and RNA polymerase II

Bryan N. Danforth *, Jennifer Fang [1], Sedonia Sipes [2]

Department of Entomology, 3119 Comstock Hall, Cornell University, Ithaca, NY 14853, USA

Received 6 July 2005; revised 13 September 2005; accepted 20 September 2005
Available online 18 January 2006

Abstract

We analyzed a combined data set of two protein-coding nuclear genes (CAD and RNA polymerase II) and a nuclear ribosomal gene (28S D2–D4 region) for 68 bee species and 11 wasp outgroups. Our taxon sampling included all seven extant bee families, 17 of 20 subfamilies, and diverse tribes. Wasp outgroups included the two families most closely related to bees: Crabronidae and Sphecidae. We analyzed the combined and single gene data sets using parsimony and Bayesian methods, which yielded largely congruent results. Our results provide reasonably strong support for family and subfamily-level relationships among bees. Our data set strongly supports the sister-group relationship of the Colletidae and Stenotritidae, and places Halictidae as sister to this clade combined. Our analyses place the Melittidae and the long-tongued (LT) bee clade (Apidae + Megachilidae) near the base of the tree with Colletidae (and Stenotritidae) in a fairly highly derived position. This topology ("Melittidae-LT basal") was obtained in previous morphological studies under certain methods of character coding. A more widely accepted tree topology that places Colletidae (and/or Stenotritidae) as sister to all other bees ("Colletidae basal") is not supported by our data. The "Melittidae-LT basal" hypothesis may better explain patterns in the bee fossil record as well as historical biogeography of certain bee groups. Our results provide new insights into higher-level bee phylogeny and indicate that CAD, RNA polymerase II, and 28S are useful data sets for resolving Cretaceous-age divergences in bees and other Hymenoptera.

Keywords: Bee phylogeny; Bee evolution; Molecular evolution; Molecular systematics

1. Introduction

Bees comprise a monophyletic group of over 16,000 species (Michener, 2000). They arose sometime in the Cretaceous (Grimaldi, 1999) and have had a long and intimate coevolutionary relationship with flowering plants. Today bees are the most important pollinators of angiosperm plants and many species are economically important crop pollinators (e.g., Barth, 1991).

While bees are an important group of insect pollinators, their higher-level phylogeny remains poorly understood. Currently, bees are divided into seven families: the long-tongued (LT) bees in the families Megachilidae and Apidae, and the short-tongued (ST) bees in the families Colletidae, Stenotritidae, Andrenidae, Halictidae, and Melittidae (Michener, 2000). Two previous studies (Alexander and Michener, 1995 and Roig-Alsina and Michener, 1993) analyzed morphological data matrices for resolving relationships among the bee families, subfamilies, and tribes in an attempt to establish a stable classification for the bees. Roig-Alsina and Michener (1993) focused their analysis on the LT bee clade and clearly established the

* Corresponding author. Fax: +1 607 255 0939.
E-mail addresses: bnd1@cornell.edu (B.N. Danforth), jfang@email.arizona.edu (J. Fang), ssipes@plant.siu.edu (S. Sipes).
[1] Present address: Department of Physiology, P.O. Box 245051, University of Arizona, Tucson, AZ 85724, USA.
[2] Present address: Department of Plant Biology, Life Science II 420, Mailcode: 6509, Southern Illinois University Carbondale, Carbondale, IL 62901-6509, USA.

1055-7903/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2005.09.022

monophyly of the two included families (Apidae and Megachilidae). Alexander and Michener (1995) provide the most recent attempt to resolve family-level phylogenies for all bees using a broad sample of bee families and subfamilies, appropriate spheciform outgroups, and modern cladistic methods. These authors were able to establish the monophyly of many families (Halictidae, Andrenidae, Megachilidae, and Apidae) but were unable to resolve clearly the relationships among the families. Monophyly of some families (Melittidae and Colletidae) was also not robustly supported.

Among the most problematic character systems to code for Alexander and Michener's (1995) analysis was the morphology of the glossa. The family Colletidae has traditionally been considered the most basal (or primitive) family of bees because they posses a glossa with a bifid (forked) apex, much like the glossa of a spheciform wasp (Malyshev, 1968, p. 299; Michener, 1944, p. 230, 1974, 1979, p. 23). Whether the bifid glossa of the colletid bees is considered a primitive or derived trait is not clear. Colletid bees use their glossa to apply a thick, cellophane-like coating to the cell and burrow walls (Batra, 1980; Espelie et al., 1992; Hefetz et al., 1979), which is not the case in spheciform wasps. There are also several colletid genera (*Hemirhiza*, *Meroglossa*, and *Paleorhiza*) in which the females have a bifid glossa and the males have an acutely pointed glossa, much like that of all other bees. The bifid glossa of Colletidae could therefore be interpreted as a synapomorphic trait for the family, rather than a plesiomorphic trait homologous to the bifid glossa in spheciform wasps. McGinley (1980) and Perkins (1912) both presented evidence in favor of this hypothesis, and this alternative interpretation of the colletid glossa has been called the Perkins–McGinley hypothesis (Alexander and Michener, 1995; Michener, 2000). Recently, Michener (2005) noted that fact that there is a genus of spheciform wasp (*Pseudoscolia*) in which females and males posses an acutely pointed glossa, much like that of a bee, further complicating the homologies of the bifid glossa in Colletidae and spheciform wasps.

Alexander and Michener (1995) coded the glossal character in two ways, which they referred to as Series I and Series II. In Series I they coded the bifid glossa of Colletidae as a plesiomorphic trait homologous with that of spheciform wasps. In series II they coded the bifid glossa as a derived trait unique to Colletidae. The two alternative codings yielded strikingly different topologies (Fig. 1). In the series I analysis (Fig. 1A), Colletidae (and Stenotritidae) tended to be basal, with LT bees (families Megachilidae and Apidae) highly derived. Alternative weighting schemes yielded differing topologies some of which showed Colletidae to be paraphyletic. Affinities of the Stenotritidae were not well established and this family appears in four different positions within the trees obtained by Alexander and Michener (1995): (1) sister group to the andrenid subfamily Oxaeinae, (2) sister group to all other bees, (3) sister group to the Colletidae, and (4) a group arising within the Colletidae. We refer to the tree topology in Fig. 1A as "Colletidae basal" in the discussion below. In the series II analysis, the

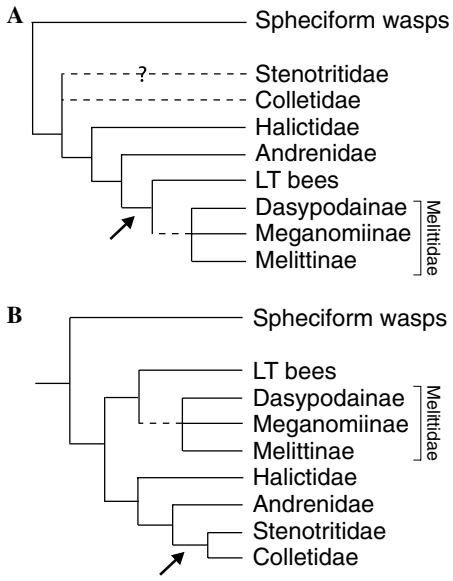

Fig. 1. Two alternative tree topologies obtained by Alexander and Michener (1995) based on Series I (A) and Series II (B) analyses. (A) The "Colletidae basal" topology and (B) the "Melittidae-LT basal" topology. Arrows indicate alternative rootings. Placement of Stenotritidae (indicated with ?) is not clear.

rooting of the tree is strikingly different (Fig. 1B). Long-tongued bees and Melittidae are basal and Colletidae and Stenotritidae are highly derived lineages nested well within the family-level phylogeny of bees. We refer to the tree topology in Fig. 1B as "Melittidae-LT basal" in the discussion below. Alternative codings of relatively few mouthpart characters resulted in markedly different interpretations of family-level relationships in bees. Alexander and Michener (1995) admitted that their data set failed to clearly resolve family-level phylogenetic relationships in bees and recommended the use of additional sources of data, such as larval morphology and molecular data (p. 419).

Slowly evolving, single-copy, and nuclear genes are a potentially valuable source of new data for resolving family-level phylogenies in bees. Previous studies have shown that these genes are capable of resolving Cretaceous-aged divergences in bees (Danforth et al., 2004), at least within one family (Halictidae). A number of promising genes have been identified for higher-level bee phylogeny including elongation factor-1α (EF-1α; Danforth, 2002; Danforth et al., 2004), long-wavelength rhodopsin (Mardulyn and Cameron, 1999), wingless (Danforth et al., 2004), phosphoenolpyruvate carboxykinase (PEPCK; Leys et al., 2002), and arginine kinase (Kawakita et al., 2003, 2004). We report here the first use of two genes, CAD and RNA polymerase II, in higher-level studies of bees. Our results indicate that these genes are unlikely to recover deep divergences in bees with high levels of support when analyzed alone, but in combination with 28S and each other these genes appear to be promising candidates for resolving basal bee divergences. Our analyses also provide new insights into the family-level phylogeny of the bees and support one of the two alternative topologies obtained by Alexander and Michener (1995).

2. Materials and methods

2.1. Genes analyzed

We generated a data set based on three nuclear genes that have previously shown promise for resolving deep divergences in insects and other arthropods. Two of these genes (CAD and RNA polymerase II) have not been used previously in bees or any Hymenoptera, whereas 28S has been used extensively (Caterino et al., 2000). DNA extractions, PCR, and sequencing protocols followed standard methods detailed in Danforth et al. (1999). PCR products were gel-purified overnight on low-melting point agarose gels and bands were extracted using the Promega Wizard PCR purification system (Promega, Madison, Wisconsin). All PCR products were sequenced in both directions, and for all three genes we had overlapping upstream and downstream PCR fragments such that much of the sequence was verified in four separate sequencing runs. Sequencing was performed using an Applied Biosystems Automated 3730 DNA Analyzer. We used Big Dye Terminator chemistry and AmpliTaq-FS DNA polymerase. Detailed protocols and additional information on all the data sets can be obtained at the first author's website: www.entomology.cornell.edu/BeePhylogeny.

2.1.1. Conserved ATPase domain

CAD is a tightly linked group of nuclear, protein-coding genes involved in pyrimidine biosynthesis. Full-length CAD sequences are available for *Drosophila melanogaster* (GenBank Accession No. AE003503) and *Anopheles gambiae* (GenBank Accession No. EAA06526). In *Drosophila* and other insects, CAD consists of four distinct loci that catalyze the first steps in the de novo pyrimidine biosynthetic pathway: carbamoylphosphate synthetase (CPS), dihydroorotase (DHO), aspartate transcarbamylase (ATC), and glutamate aminotransferase (GAT) (Freund and Jarry, 1987; Moulton and Wiegmann, 2004). In insects and vertebrates the largest of the CAD domains is the carbamoylphosphate synthase (CPS) domain (Kim et al., 1992). CPS consists of ≈4 kb of coding sequence, and provided robust support for higher-level, Cretaceous-age divergences in Diptera (Moulton, 2003; Moulton and Wiegmann, 2004).

From GenBank, we downloaded a complete CAD sequence for *D. melanogaster*, and partial CAD coding sequences for *Heterostomus* sp. (AY280682) and *Hilarimorpha* sp. (AY280683). Using the recently released sequence of the honey bee genome, we located the CAD homolog by performing BLAST searches with fly CAD sequences. We assembled a fragment of CAD that corresponds to the CPS domain used by Moulton and Wiegmann (2004) for flies. Using this alignment we developed primers for amplifying and sequencing an approximately 3000 bp fragment of the CPS domain of CAD in bees (see below; Table 1).

We initially developed two sets of primers from the alignment of *Apis*, *Drosophila*, *Heterostomus*, and *Hilarimorpha*, in some cases using primer sites identified by Moulton and Wiegmann (2004). The upstream primer pair, ApCADfor1/Ap835rev1 (94 °C 1 min, 52 °C 1 min, and 72 °C 1 min, 35 cycles), amplified a region of about 600 bp spanning two exons and a single intron. The downstream primer pair, Ap787for2/Ap1098rev2 (94 °C 1 min, 52 °C 1 min, and 72 °C 1 min 30 s, 35 cycles) amplified a product of over 1000 bp which overlapped the intron and 200 bp of

Table 1
Primers used for CAD, RNA polymerase II, and 28S

Forward primers	Reverse primers
CAD	
ApCADfor1: 5′-GGW TAT CCC GTD ATG GCB MGW GC-3′ [23 mer, $T_m = 62.1$ °C]	Ap835rev1: 5′-GCA THA CYT CHC CCA CRC TYT TC-3′ [23 mer, $T_m = 59.8$ °C]
Ap787for2: 5′-TGC TTY GAR CCD AGY CTH GAT TAY TG-3′ [26 mer, $T_m = 60.0$ °C]	Ap1098rev2: 5′-ATA TTR TTK GGC ARY TGD CCK CCC-3′ [24 mer, $T_m = 61.1$ °C]
ApCADfor2mod: 5′-GAT GGG AYC TNR GNA ART TYC-3′ [21 mer, $T_m = 53.1$ °C]	ApCADrev1mod: 5′-GCC ATY RCY TCB CCY ACR CTY TTC AT-3′ [26 mer, $T_m = 62.2$ °C]
ApCADfor3: 5′-CTC HGT KGA RTT YGA TTG GTG YGC-3′ [24 mer, $T_m = 60.9$ °C]	ApCADrev4a: 5′-GGC CAY TGN GCN GCC ACY GTG TCT ATY TGY TTN ACC-3′ [36 mer, $T_m = 68.6$ °C]
ApCADfor4: 5′-TGG AAR GAR GTB GAR TAC GAR GTG GTY CG-3′ [29 mer, $T_m = 63.1$ °C]	
RNA polymerase II	
polfor2: 5′-TGG GAY GSY AAA ATG CCK CAA CC-3′ [23 mer, $T_m = 61.6$ °C]	polrev2: 5′-TTY ACA GCA GTA TCR ATR AGA CCT TC-3′ [26 mer, $T_m = 55.5$ °C]
polfor2a: 5′-AAY AAR CCV GTY ATG GGT ATT GTR CA-3′ [26 mer, $T_m = 58.0$ °C]	polrev2a: 5′-AGR TAN GAR TTC TCR ACG AAT CCT CT-3′ [26 mer, $T_m = 57.0$ °C]
polfor3: 5′-CAR GTT ATY GCT TGT GTS GCY CAA C-3′ [25 mer, $T_m = 59.6$ °C]	polrev3: 5′-GAA ARA TCT TYT GYA CGT TGG ADA TC-3′ [26 mer, $T_m = 54.3$ °C]
28S (D2–D4)	
D2-3665F: 5′-AGA GAG AGT TCA AGA GTA CGT G-3′ [22 mer, $T_m = 54.7$ °C]	D3-4283R: 5′-TAG TTC ACC ATC TTT CGG GTC CC-3′ [23 mer, $T_m = 59.9$ °C]
D3-4048F (28SD4for): 5′-CCC GTC TTG AAA CAC GGA CCA AGG-3′ [24 mer, $T_m = 63.5$ °C]	D5-4749R (28SD4rev): 5′-GTT ACA CAC TCC TTA GCG GA-3′ [20 mer, $T_m = 55.0$ °C]

exon of the upstream fragment. The downstream fragment also contained a second intron near the 3′ end of the PCR product. Because of the relatively high rate of sequence variation within the CAD gene, we were able to sequence only about 20% of the bee species we tried.

From an alignment of the sequences we had generated, we designed modified internal primers using our original primer sites. With primer pairs ApCADfor1/ApCADrev1mod (94 °C 1 min, 55 °C 1 min, and 72 °C 1 min, 35 cycles) and ApCADfor2mod/Ap1098rev2 (94 °C 1 min, 54 °C 1 min, and 72 °C 1.5 min, 35 cycles), we were able to sequence another 25% of the bees we tested.

Again using an alignment of the sequences we had generated, we designed a new pair of internal primers, ApCADfor4/ApCADrev4a (94 °C 1 min, 52 °C 1 min, and 72 °C 1.5 min), which spanned a smaller region of our dataset. The ApCADfor4/ApCADrev4a fragment spanned about 1200 bp of exon and the first intron. With the primers ApCADfor4/ApCADrev4a, we were able to amplify the majority of our dataset with either a single PCR or by combining an upstream and a downstream fragment using ApCADfor4/ApCADrev1mod (94 °C 1 min, 58 °C, 1 min, and 72 °C 1 min, 35 cycles) and ApCADfor2mod/ApCADrev4a (94 °C 1 min, 52 °C 1 min, and 72 °C 1.5 min, 35 cycles). These last three primer combinations would be the most efficient method for generating a complementary data set in other bees and wasps.

Interestingly, CAD contains several small introns (usually 50–60 bp) that vary in presence and absence among taxa. Intron presence/absence could be coded as a character when introns vary in position among taxa (e.g., Brady and Danforth, 2004). Fig. 2 provides a map of the CAD gene in bees with intron positions and primer sites identified.

2.1.2. RNA polymerase II

Pol II refers to the protein-coding gene which codes for the two largest subunits of the RNA polymerase II enzyme. RNA polymerase II is involved in the synthesis of pre-mRNA (Liu et al., 1999; Shultz and Regier, 2000). Pol II appears to be a single-copy gene in bees. Pol II is easy to amplify and align and contains several highly conserved regions which simplifies primer design (Liu et al., 1999), and has been used extensively in studies of higher-level arthropod and myriapod phylogeny (Regier and Shultz, 1997, 2001a,b; Regier et al., 2004a,b, 2005; Shultz and Regier, 2000, 2001).

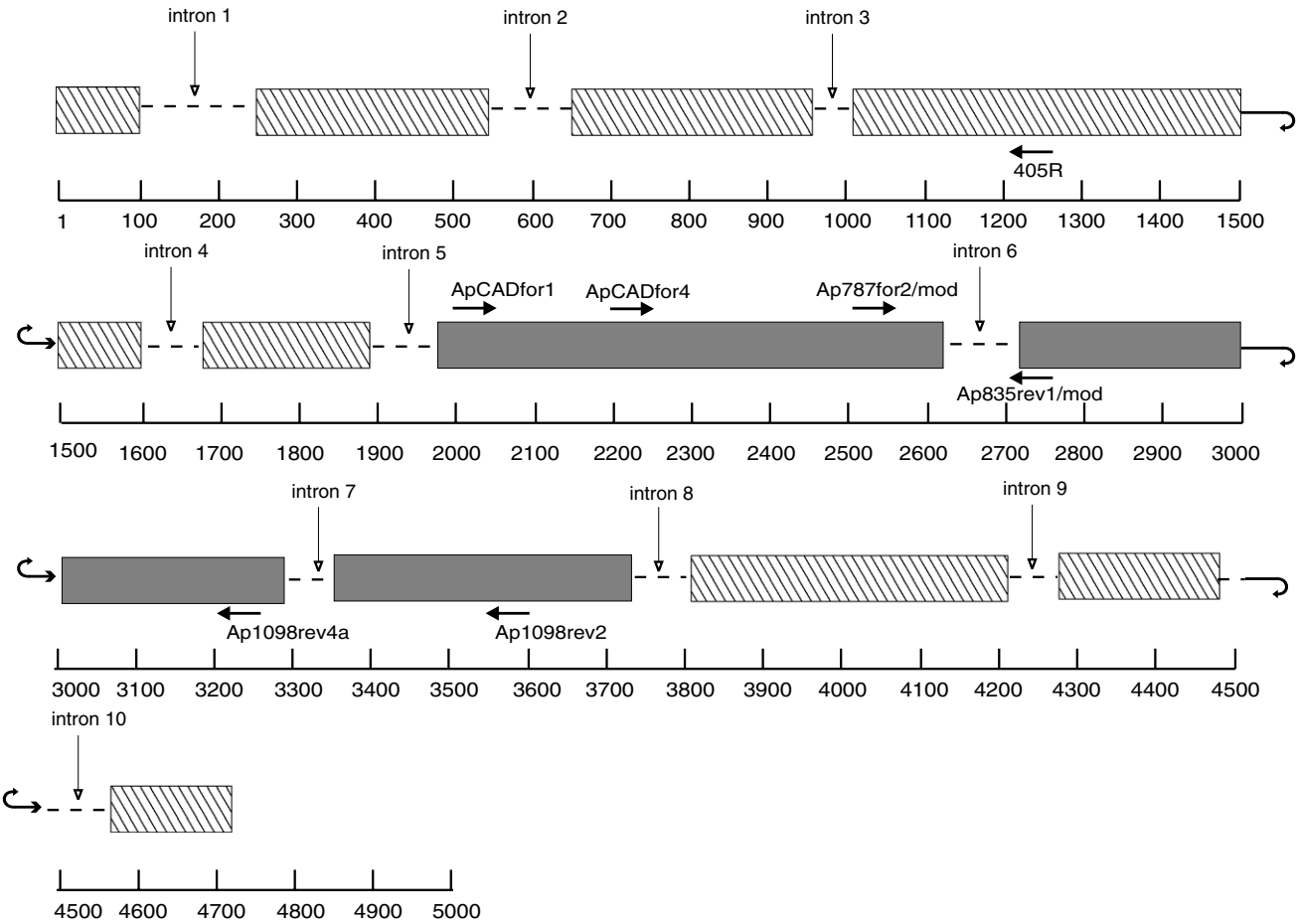

Fig. 2. Map of the CPS domain of the CAD gene in *A. mellifera* showing the locations of introns and exons. The region analyzed in the present study includes exons 6–8 (shaded bars) and the two included introns (introns 6–7).

Using the complete coding sequence of RNA Polymerase II (Pol II) from *D. melanogaster* (NM_078569) and a partial coding sequence from *Periplaneta americana* (U90040), we generated a consensus sequence and performed a BLAST search against the recently released *Apis* genome to obtain a complete Pol II sequence for *Apis mellifera*. From the alignment of *Apis*, *Drosophila*, and *Periplaneta*, we developed a pair of primers, polfor2/polrev2 (94 °C 1 min, 52 °C 1 min, and 72 °C 1 min, 35 cycles), which spanned an ≈800 bp, intron-less region of the large subunit of the Pol II gene.

Using polfor2/polrev2, we were able to amplify and sequence approximately one-third of our dataset. From these sequences, we designed a second set of primers, polfor2a/polrev2a (94 °C 1 min, 52 °C 1 min, and 72 °C 1 min, 35 cycles). The fragment amplified by polfor2a/polrev2a is shifted slightly downstream from the fragment generated by polfor2/polrev2, but overlaps it by over 700 bp. Polfor2a/polrev2a amplified our dataset with a much higher rate of success and we would recommend these primers in future studies (Table 1).

2.1.3. 28S D2–D4 region

The 28S D2–D3 region has been used previously in bees (Cameron and Mardulyn, 2001). The Cameron and Mardulyn (2001) data set spans approximately 690 bp and was generated with primers D2-3665F (Bel28S) and D3-4283R (Mar28Srev) (94 °C 1 min, 65 °C 1 min, and 72 °C 1 min 35 cycles). We used these same primers for the data set we report here in addition to primers D3-4048F and D5-4749R (94 °C 1 min, 52 °C 1 min, and 72 °C 1 min, 35 cycles) to amplify an ≈700 bp fragment of the D3–D4 region that overlaps the D2–D3 fragment by almost 300 bp (Table 1). The resulting data set spans ≈1200 bp of the D2–D4 region of the gene. The alignment has some highly conserved regions, but several expansion regions (loops) make portions of the alignment ambiguous. Regions of ambiguous alignment were excluded from the analysis.

2.2. Phylogenetic methods

We included a total of 82 individuals (11 wasp outgroups and 71 bee ingroups) representing all seven families of bees (sensu Michener, 2000) and 17 of the 20 subfamilies (we were unable to amplify CAD for Fideliinae, Xylocopinae, and Nomioidinae) (Table 2). Our taxon sampling was driven partially by our ability to amplify CAD in the taxa we tested. Not all species of bees we tested could be amplified for CAD and this limited the taxa we could add to the final data set. In a small number of cases species were represented by two individuals collected from the same or separate localities (*Zacosmia maculata*, *Thyreus delumbatus*, *Protandrena verbesinae*, and *Rophites algirus*; Table 2). Outgroups included representatives of two of the four spheciform families, Crabronidae and Sphecidae (Melo, 1999; Table 2). Voucher specimens are deposited in the Cornell University Insect Collection. GenBank accession numbers and specimen voucher codes are listed in Table 2. Complete locality data and our combined data set in Nexus format is available as Supplementary information from the Elsevier website (www.sciencedirect.com).

Alignments for all genes were generated in the Lasergene DNA Star software package using Clustal W. Long CAD introns in some species (e.g., Diphaglossinae; Table 2) required manual alignments or exclusion of introns. Alignments for the 28S D2–D4 region were adjusted by eye and some unalignable regions were excluded from the analysis. Reading frames and intron/exon boundaries were determined by comparison with sequences obtained for the honey bee, *A. mellifera*.

2.2.1. Parsimony methods

We performed maximum parsimony (MP) analyses using PAUP* v. 4.0b10 (Swofford, 2002). Initially we performed equal weights parsimony analyses on each of the three data sets and then combined the data sets into a single analysis. We tested for data set congruence using the incongruence length difference test (ILD test; Farris et al., 1995) implemented in PAUP*. Branch support for the individual data sets as well as the combined data set were estimated using bootstrap analysis (Felsenstein, 1985). For parsimony searches we performed 500 random sequence additions. For calculating bootstrap proportions we performed 500 replicates with 10 random sequence additions per replicate. Bremer support (Bremer, 1988) and partitioned Bremer support (Baker and DeSalle, 1997; Baker et al., 1998, 2001) were also calculated for the total data set using TreeRot v.2 (Sorenson, 1999). We calculated data decisiveness (DD) for the separate and combined data sets as outlined in Goloboff (1991).

2.2.2. Bayesian methods

For the Bayesian analyses we used MrBayes v. 3.0 (Huelsenbeck and Ronquist, 2001; http://morphbank.ebc.uu.se/mrbayes3/). We analyzed the combined data set using a range of models including the Kimura 2-parameter model (K2P), the Hasegawa–Kishino–Yano model (HKY), and the general time reversible (GTR) model. Various among-site rate variation models were used to account for rate variation among genes and codon positions. In the simplest method (site-specific rates models; SSR), we assigned each codon position within the two protein-coding genes (CAD and RNA polymerase II) to its own rate and applied a single rate to the 28S data. Site-specific rates models are a reasonable choice for protein-coding genes, where codon positions differ substantially in rate and can be identified a priori. In order to allow for more rate variation within the 28S data set, we also performed an analysis (SSR + G[28S]) in which we applied separate rates to each codon position and a gamma distribution to the 28S data. Finally, we also performed analyses in which we fitted a separate gamma distribution and invariant sites model to each gene (I + G). For all analyses we treated the separate genes (and codon positions) as "unlinked" so that separate parameter

Table 2
Wasp and bee taxa included in this study with the GenBank accession numbers for each gene, intron lengths for the two CAD introns and the collector

Family	Subfamily	Species	Abbrev.	No.	GenBank CAD	GenBank 28S	GenBank Pol II	CAD intron 6	CAD intron 7	Collector
Crabronidae	Bembicinae	*Ochleroptera bipunctata*	Ocbi	815	DQ067115	AY654463	AY945147	67	113	Ascher
Crabronidae	Bembicinae	*Xerostictia* sp.	Xrst	814	DQ067116	AY654471	AY945175	76	134	Danforth
Crabronidae	Crabroninae	*Anacrabro ocellatus*	Anoc	803	DQ067117	DQ072142	AY945091	72	na	Danforth
Crabronidae	Crabroninae	*Oxybelus* sp.	Oxsp	805	DQ067118	DQ072153	AY945148	149	na	Danforth
Crabronidae	Crabroninae	*Plenoculus* sp.	Plsp	810	DQ067119	AY654465	AY945150	67	na	Danforth
Crabronidae	Crabroninae	*Tachysphex* sp.	Txsp	807	DQ067120	AY654470	AY945172	76	na	Danforth
Crabronidae	Pemphredoninae	*Stigmus* sp.	Sgsp	818	DQ067121	AY654467	AY945165	73	178	Danforth
Crabronidae	Philanthinae	*Cerceris* sp.	Cesp	825	DQ067122	AY654460	AY945100	109	67	Danforth
Crabronidae	Philanthinae	*Clypeadon* sp.	Clsp	819	DQ067123	AY654461	AY945104	76	85	Danforth
Crabronidae	Philanthinae	*Philanthus gibbosus*	Phgi	828	DQ067124	AY654464	AY945149	84	69	Ascher
Sphecidae	Sphecinae	*Sceliphron cementarium*	Spcm	836	DQ067125	AY654468	AY945166	101	na	Danforth
Andrenidae	Andreninae	*Andrena brooksi*	Ansp	643	DQ067126	AY654474	AY945092	81	na	Danforth
Andrenidae	Andreninae	*Andrena nasonii*	Anna	231	DQ067127	DQ060849	AY945177	79	83	Ascher
Andrenidae	Oxaeinae	*Protoxaea gloriosa*	Pxgl	226	DQ067128	AY654480	AY945157	65	na	Danforth
Andrenidae	Panurginae	*Calliopsis pugionis*	Capu	509	DQ067129	AY654477	AY945098	92	79	Danforth
Andrenidae	Panurginae	*Calliopsis anthidia*	Caan	598	DQ067130	AY654475	AY945096	72	na	Ascher
Andrenidae	Panurginae	*Macrotera latior*	Pdla	224	DQ067131	DQ060863	AY945197	82	na	Danforth
Andrenidae	Panurginae	*Melitturga clavicornis*	Mtcl	959	DQ067134	AY654478	AY945145	66	85	Danforth
Andrenidae	Panurginae	*Meliturgula haematopsila*	Mtha	1035	DQ067132	DQ060859	AY945194	65	70	Danforth
Andrenidae	Panurginae	*Meliturgula minima*	Mtmi	1037	DQ067133	DQ072150	AY945193	65	70	Danforth
Andrenidae	Panurginae	*Panurgus calcarata*	Pnca	514	DQ067135	AY654479	AY945152	73	73	Ascher
Andrenidae	Panurginae	*Protandrena verbesinae*	Prve	619	DQ067136	DQ060866	AY945199	95	na	Danforth
Andrenidae	Panurginae	*Protandrena verbesinae*	Prve	1093	DQ067137	DQ060865	DQ069325	95	83	Danforth
Andrenidae	Panurginae	*Pseudopanurgus fraterculus*	Psfr	644	DQ067138	DQ060867	AY945200	78	na	Danforth
Colletidae	Colletinae	*Colletes inaequalis*	Coin	450	DQ067139	AY654484	AY945107	130	na	Danforth
Colletidae	Colletinae	*Colletes skinneri*	Cosk	632	DQ067140	AY654485	AY945108	144	259	Danforth
Colletidae	Colletinae	*Scrapter heterodoxus*	Scht	903	DQ067145	AY654500	AY945162	168	na	Danforth
Colletidae	Diphaglossinae	*Caupolicana vestita*	Cpve	848	DQ067141	AY654486	AY945109	621	na	Packer
Colletidae	Diphaglossinae	*Diphaglossa gayi*	Diga	850	DQ067142	AY654488	AY945115	764	na	Packer
Colletidae	Euryglossinae	*Xanthesma furcifera*	Xnfu	709	DQ067143	AY654505	AY945173	172	na	Danforth
Colletidae	Hylaeinae	*Hylaeus amiculus*	Hyam	698	DQ067144	AY654491	AY945126	214	na	Danforth
Colletidae	Xeromelissinae	*Chilimelissa rozeni*	Chrz	857	DQ067146	AY654481	AY945102	70	72	Packer
Halictidae	Halictinae	*Sphecodes pecosensis*	Sppe	1114	DQ067147	DQ072154	AY945204	96	102	Danforth
Halictidae	Halictinae	*Sphecodes* sp.	Spsp	1055	DQ067148	DQ072155	AY945203	98	64	Danforth
Halictidae	Halictinae	*Zonalictus near kabetense*	Zoap	1051	DQ067149	DQ060870	AY945206	84	79	Danforth
Halictidae	Nomiinae	*Dieunomia heteropoda*	Noht	1113	DQ067150	DQ072151	AY945182	83	57	Danforth
Halictidae	Nomiinae	*Dieunomia nevadensis*	None	1111	DQ067151	DQ060852	AY945183	83	57	Danforth
Halictidae	Nomiinae	*Lipotriches patellifera*	Lipt	1059	DQ067152	DQ072146	DQ069326	81	64	Danforth
Halictidae	Nomiinae	*Macronomia aureozonata*	Mcau	1061	DQ067153	DQ072149	DQ069327	83	77	Danforth
Halictidae	Nomiinae	*Macronomia sanguinolenta*	Masg	1065	DQ067154	DQ072148	AY945191	54	76	Danforth
Halictidae	Nomiinae	*Nomia tetrazonata*	Note	1129	DQ067155	DQ072152	DQ069328	79	76	Danforth
Halictidae	Nomiinae	*Pseudapis obesula*	Psob	1072	DQ067156	DQ060868	DQ069329	92	57	Danforth
Halictidae	Rophitinae	*Conanthalictus conanthi*	Coco	1117	DQ067157	DQ072144	DQ069330	62	57	Danforth
Halictidae	Rophitinae	*Rophites algirus*	Roal	968	DQ067158	AY654515	AY945158	103	71	Danforth
Halictidae	Rophitinae	*Rophites algirus*	Roal	972	DQ067159	DQ072159	AY945158	103	71	Danforth
Halictidae	Rophitinae	*Systropha glabriventris*	Sysp	1079	DQ067160	DQ072156	DQ069331	98	73	Danforth
Melittidae	Dasypodainae	*Dasypoda argentata*	Daar	973	DQ067161	AY654518	AY945112	no intron	na	Danforth
Melittidae	Dasypodainae	*Dasypoda hirtipes*	Dahi	975	DQ067162	AY654519	AY945113	no intron	59	Danforth
Melittidae	Dasypodainae	*Dasypoda visnaga*	Davi	977	DQ067163	DQ060851	AY945114	59	83	Danforth
Melittidae	Dasypodainae	*Hesperapis (Capicola)* sp.	Hesp	940	DQ067165	AY654523	AY945123	74	83	Danforth
Melittidae	Dasypodainae	*Hesperapis rhodocerata*	Herh	1122	DQ067166	DQ060856	AY945186	no intron	71	Danforth
Melittidae	Dasypodainae	*Hesperapis larreae*	Hela	488	DQ067167	AY654521	AY945121	69	na	Ascher
Melittidae	Dasypodainae	*Hesperapis regularis*	Herg	469	DQ067168	AY654456	AY945122	68	83	Moeller
Melittidae	Dasypodainae	*Haplomelitta griseonigra*	Hpgr	939	DQ067164	AY654524	AY945125	56	58	Danforth
Melittidae	Meganomiinae	*Meganomia binghami*	Mgbg	1021	DQ067169	AY654528	AY945144	no intron	59	Danforth
Melittidae	Melittinae	*Macropis europaea*	Maeu	980	DQ067170	AY654525	AY945138	73	na	Danforth
Melittidae	Melittinae	*Macropis nuda*	Manu	17ja	DQ067171	AY654454	AY945139	73	na	Ascher
Melittidae	Melittinae	*Melitta dimidiata capensis*	Meca	942	DQ067172	AY654526	AY945140	74	85	Danforth
Melittidae	Melittinae	*Melitta eickworti*	Meew	508	DQ067173	AY654527	AY945141	74	na	Ascher
Melittidae	Melittinae	*Melitta leporina*	Mele	981	DQ067174	AY654529	AY945142	75	na	Danforth
Melittidae	Melittinae	*Rediviva mcgregori*	Rvmc	945	DQ067175	AY654531	AY945159	73	na	Danforth
Stenotritidae		*Stenotritus* sp.	Stsp	1015	DQ067176	AY654503	AY945167	66	na	Houston

(*continued on next page*)

Table 2 (*continued*)

Family	Subfamily	Species	Abbrev.	No.	GenBank CAD	GenBank 28S	GenBank Pol II	CAD intron 6	CAD intron 7	Collector
Apidae	Apinae	*Anthophora montana*	Anmo	633	DQ067177	AY654533	AY945090	81	89	Danforth
Apidae	Apinae	*Apis mellifera* (reference)	Apme		DQ067178	AY703551	DQ069332	86	68	
Apidae	Apinae	*Centris rhodopus*	Cnrh	615	DQ067179	AY654537	AY945106	87	64	Danforth
Apidae	Apinae	*Martinapis luteicornis*	Malu	1101	DQ067181	DQ072147	DQ069333	no intron	62	Danforth
Apidae	Apinae	*Pachymelus peringueyi*	Pmpe	985	DQ067182	AY654544	AY945151	144	na	Danforth
Apidae	Apinae	*Tetralonia cinctula*	Ttcn	1045	DQ067183	DQ072157	AY945205	no intron	62	Danforth
Apidae	Apinae	*Thyreus delumbatus*	Thdl	987	DQ067184	AY654546	AY945169	57	63	Danforth
Apidae	Apinae	*Thyreus delumbatus*	Thdl	987	DQ067184	AY654546	AY945169	57	63	Danforth
Apidae	Apinae	*Zacosmia maculata*	Zoma	650	DQ067185	AY654548	AY945176	67	66	Danforth
Apidae	Apinae	*Zacosmia maculata*	Zoma	1110	DQ067186	DQ072158	DQ069334	67	66	Danforth
Apidae	Nomadinae	*Holcopasites ruthae*	Horu	511	DQ067187	AY654540	AY945124	65	85	Danforth
Apidae	Nomadinae	*Paranomada velutina*	Pnve	652	DQ067188	AY654545	AY945154	57	61	Danforth
Apidae	Nomadinae	*Triepeolus robustus*	Trrz	635	DQ067189	AY654547	AY945170	72	67	Danforth
Megachilidae	Megachilinae	*Lithurgus apicalis*	Liap	861	DQ067191	DQ072145	DQ069335	no intron	77	McIntosh
Megachilidae	Megachilinae	*Lithurgus echinocacti*	Liec	863	DQ067195	AY654541	AY945136	no intron	na	McIntosh
Megachilidae	Megachilinae	*Coelioxys afra*	Coaf	1027	DQ067193	DQ072143	AY945179	60	92	Danforth
Megachilidae	Megachilinae	*Megachile pugnata*	Mepg	595	DQ067196	AY654543	AY945143	67	83	Danforth
Megachilidae	Megachilinae	*Chelostoma fuliginosum*	Chfu	496	DQ067192	AY654536	AY945101	87	71	Danforth
Megachilidae	Megachilinae	*Heriades crucifera*	Hrcr	1121	DQ067194	DQ060855	DQ069336	78	73	Danforth
Megachilidae	Megachilinae	*Anthidium oblongatum*	Atob	505	DQ067197	AY654534	AY945093	61	82	Danforth

estimates were obtained for each gene and codon position for all runs. For the SSR models we allowed either 6 (or 7) discrete rate categories corresponding to the three codon positions within each protein-coding gene (CAD and RNA polymerase II) and the 28S gene, when this was treated as a single rate category.

Analyses consisted of running four simultaneous chains for 1×10^6 generations. We repeated analyses to verify that different starting points did not bias the resulting tree topologies and parameter estimates. Trees were sampled at intervals of 100 generations for a total of 10,000 trees. We plotted the likelihood scores against generation time to identify the region of the analysis in which the parameter estimates were stable. We discarded the "burn-in" region (trees and parameter estimates obtained before equilibrium; generally the first 1000–2000 trees) and calculated the mean, variance, and 95% credibility intervals of the parameter estimates using MrBayes. Trees were represented as 50% majority rule consensus trees using PAUP*.

3. Results

Our data set consisted of 1347 aligned bp of CAD (619 parsimony informative sites), 890 bp of Pol II (305 parsimony informative sites), and 1037 bp of 28S (282 parsimony informative sites; Table 3). The region of CAD we analyzed corresponds to exons six, seven, and eight in the honey bee CPS domain (Fig. 2). This region included two introns (introns six and seven) which were excluded from the analysis. The first intron was not present in all taxa and the second intron was present in all taxa for which we have data (Table 2). The phylogenetic distribution of intron six is shown in Table 2 and discussed below. The region of Pol II we analyzed lacked introns entirely. For both CAD and Pol II

Table 3
Summary of parsimony results from each gene and each codon position within the two protein-coding genes

Data partitions	A + T %	base[a] comp.	Total sites	PI sites [b]	Prop. sites PI[c]	CI[d]	Length (on MP tree)	Data decisiveness	PBS/ min steps[e]
CAD (all)	51.1	<0.001	1347	619	0.46	0.1808	6669	0.436	1.025
CAD nt1	51.4	1.00	449	122	0.27	0.2463	769		
CAD nt2	63.6	1.00	449	69	0.15	0.4436	312		
CAD nt3	38.5	<0.001	449	428	0.95	0.1629	5588		
Pol II (all)	57.1	<0.001	890	305	0.34	0.1385	3968	0.423	0.860
Pol II nt1	52.4	1.00	296	31	0.10	0.1385	253		
Pol II nt2	63.8	1.00	297	1	0.00	1	5		
Pol II nt3	55.3	<0.001	297	273	0.92	0.1393	3710		
28S (all)	44.1	1.00	1037	282	0.27	0.3629	1868	0.650	0.290
Total sites	51.6	1.00	3274	1206	0.37	0.1889	12505	0.454	0.772

[a] Significant heterogeneity in base composition among species was measured using the chi-square test implemented in PAUP.
[b] Parsimony-informative (PI) sites.
[c] Proportion of total sites that are parsimony-informative.
[d] Consistency index (CI) excluding uninformative sites.
[e] Partitioned Bremer support (PBS) divided by minimum steps to standardize across data sets.

we had a relatively small proportion of missing data in some taxa at the beginning and/or the end of the data set. This was due to the use of a variety of primer pairs, not all of which coincided with the beginning and end of the complete data set. 28S sequences were complete for all taxa. We saw no evidence of multiple copies of CAD or Pol II in the sequences we obtained. However, for all species of *Perdita* analyzed an unusually large number of polymorphic sites were detected, possibly indicating several Pol II alleles within this genus.

Among the genes, base composition varied from 38.5% A/T (CAD nt3) to 63.8% A/T (Pol II nt2). CAD showed a slight G/C bias in third codon positions, while Pol II showed a slight A/T bias (Table 3). Overall, CAD and Pol II had nearly 50% A/T and 28S showed a slight G/C bias. Significant heterogeneity existed among taxa for both the CAD and Pol II data sets (in overall base composition as well as nt3 base composition; Table 3).

3.1. Phylogenetic results

Analyses of individual genes yielded variable levels of phylogenetic resolution. For CAD alone, equal weights analysis recovered 12 families and/or subfamilies that were recovered in the combined analysis: Rophitinae, Nomiinae, Halictinae, Halictidae, Colletidae, Andreninae, Panurginae, Andrenidae, Melittinae, Dasypodainae, Megachilidae, and Apidae. In addition, CAD recovered supra-familial groupings such as Colletidae + Strenotritidae, Halictidae + (Colletidae + Strenotritidae), and Andrenidae + (Halictidae + (Colletidae + Strenotritidae)). RNA polymerase II, when analyzed alone by equal weights parsimony, recovered only one subfamily: Dasypodainae. 28S analyzed alone recovered eight families and/or subfamilies, including Rophitinae, Nomiinae, Halictinae, Halictidae, Colletidae, Panurginae (but not Andrenidae), Dasypodainae, Megachilidae, and several supra-familial groups, such as Colletidae + Strenotritidae, and Halictidae + (Colletidae + Strenotritidae). No data set recovered nodes with greater than 50% bootstrap support that were not well supported based on the combined analysis. Overall, CAD alone yielded far more nodes supported by bootstrap values greater than 50% than any other gene, and CAD appears to be the most informative data set of the three we analyzed.

The ILD test detected significant incongruence ($p < 0.01$) among the three genes. Nevertheless, we combined the data sets into a single analysis because there appeared to be no significant topological incongruence among them (as judged by the bootstrap analyses of each data set separately). The ILD test is highly sensitive and can give inflated levels of significance when data partitions differ in size (Dowton and Austin, 2002). We do not consider a significant p value reason for not combining data sets.

When the combined data set was analyzed by equal weights parsimony we obtained four trees of equal length (12505 steps). The strict consensus of these four trees recovered monophyly of the bees and all bee families except Apidae and Melittidae (Fig. 3). Melittidae appears as a polyphyletic group (Fig. 3). The basal branch of the bees appears to be the melittid subfamily Dasypodainae. Bootstrap and Bremer analyses indicate that monophyly of several families is well supported by our data, including Megachilidae (96% bootstrap support), Andrenidae (80% bootstrap support), Colletidae (99% bootstrap support), and Halictidae (100% bootstrap support). Relationships among the ST bee families (excluding Melittidae) were also well supported. Stenotritidae is unambiguously sister to Colletidae (99% bootstrap support), Halictidae forms the sister group to Stenotritidae + Colletidae (88% bootstrap support), and Andrenidae forms the sister group to these three families (74% bootstrap support). Overall, the parsimony results support a highly derived position for the Colletidae and a basal position for the Melittidae and LT bees.

One anomalous result of the parsimony analysis was the placement of Megachilidae within Apidae (see below). This was not well supported based on the bootstrap analysis and is possibly an artifact of limited taxon sampling within the LT bees (particularly Apidae) as a whole. Apidae was a well-supported group in the Alexander and Michener (1995) and the Roig-Alsina and Michener (1993) studies.

We mapped the distribution of intron six on the parsimony tree (Fig. 3). Intron six in CAD was generally less than 100 bp in length, however, some taxa (e.g., Diphaglossinae) possessed introns of more than 600 bp in length (Table 2). The intron is primitively present in the bees and is lost several times within three families (Melittidae, Apidae, and Megachilidae). Within closely related genera of Melittidae (e.g., *Hesperapis* and *Dasypoda*) the character appears to be lost repeatedly. As in the previous study of CAD (Moulton and Wiegmann, 2004) and an analysis of intron evolution in the *white* gene (Krzywinski and Besansky, 2002), our results suggest that intron gains and/or loses may occur frequently in some genes.

Bayesian results were largely congruent with the parsimony results. The major differences involved the placement of Melittidae, which was recovered as a monophyletic group in all Bayesian analyses (Figs. 4 and 5). Contrary to the parsimony results, the Bayesian analyses also supported monophyly of the ST bees. The GTR + I + G model differed from all the site-specific rates models in supporting monophyly of the Apidae (Fig. 5).

If one looks across the Bayesian and parsimony support values (Table 4), most higher-level groups are well supported by both methods. However, ST bee monophyly and monophyly of the melittid subfamily Melittinae were only weakly supported. The major differences between the parsimony and Bayesian results is that Bayesian analyses consistently supported melittid monophyly (with 100% posterior probability) while our parsimony analyses indicate that this family is paraphyletic. Morphological support for melittid monophyly is weak (see below).

Our results strongly support the sister group relationship between Stenotritidae and Colletidae, effectively rejecting the hypothesis that Stenotritidae arises from within the

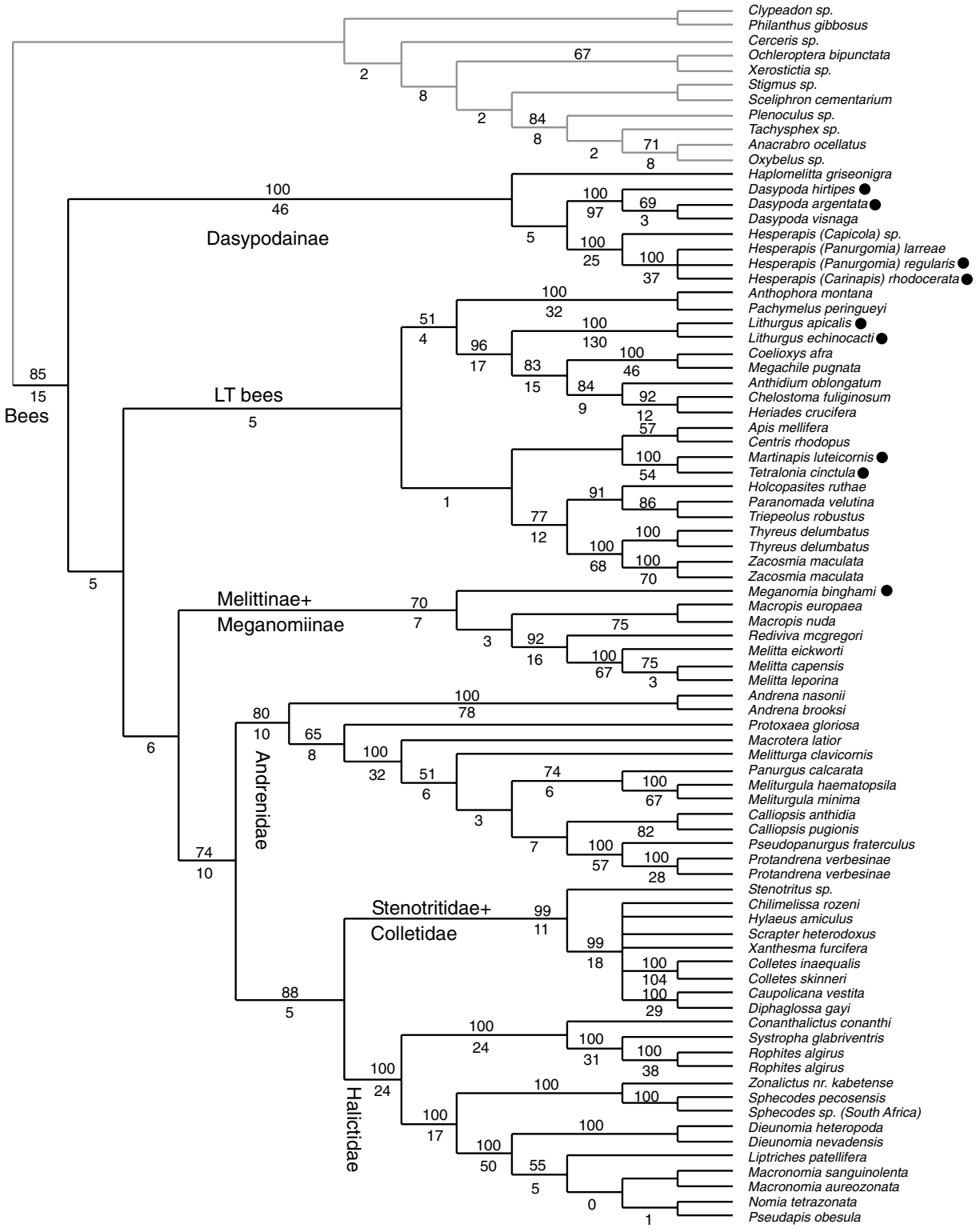

Fig. 3. Strict consensus of four equally parsimonious trees obtained with equal weights analysis of all three genes combined. Bootstrap values are above the branches. Bremer support values are below the branches. Major lineages are labeled. Closed circles indicate taxa lacking intron six.

Colletidae (Engel, 2001). The presence of a unique intron in the F1 copy of EF-1α also supports monophyly of Colletidae and exclusion of Stenotritidae from this family (Brady and Danforth, 2004). Overall our results support the "Melittidae-LT basal" hypothesis of Alexander and Michener (1995).

3.2. Comparison among genes

We assessed the characteristics and performance of our genes using both parsimony and Bayesian methods. Data decisiveness was equal in the two protein-coding

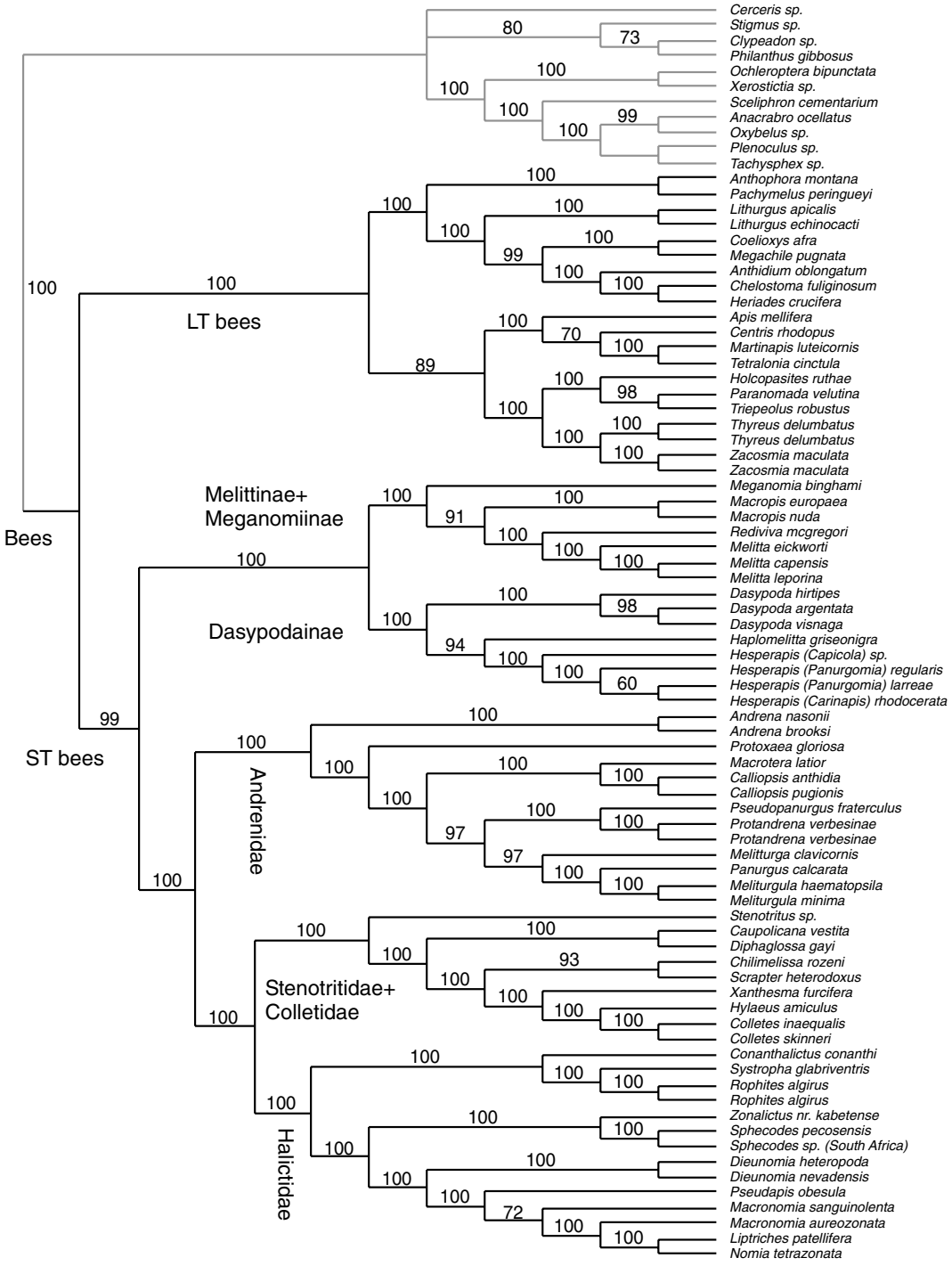

Fig. 4. Majority rule consensus of trees 2000–10,000 in the Bayesian analysis using the GTR + SSR model with 7 rate categories corresponding to the three codon positions of each protein-coding gene and the 28S gene. Values above the branches are posterior probabilities. Average Ln likelihood = −55565.49.

genes (0.436 for CAD and 0.423 for Pol II) and slightly higher for the ribosomal gene (0.650 for 28S; Table 3). Partitioned Bremer support standardized by the minimum number of steps indicated that CAD and Pol II are providing the most support (1.025 and 0.860, respectively), and the 28S data set is providing substantially less support (0.290; Table 3). CAD appears to be the most useful data set of the three we analyzed here based on consistency index, number of nodes recovered, number of nodes recovered with >50% bootstrap support, and partitioned Bremer support.

Examination of the number and proportion of parsimony informative sites across genes and across codon positions indicates that CAD shows significantly more nt1 and nt2 variation than Pol II (Table 3). For both genes, roughly 95% of all third codon positions are parsimony

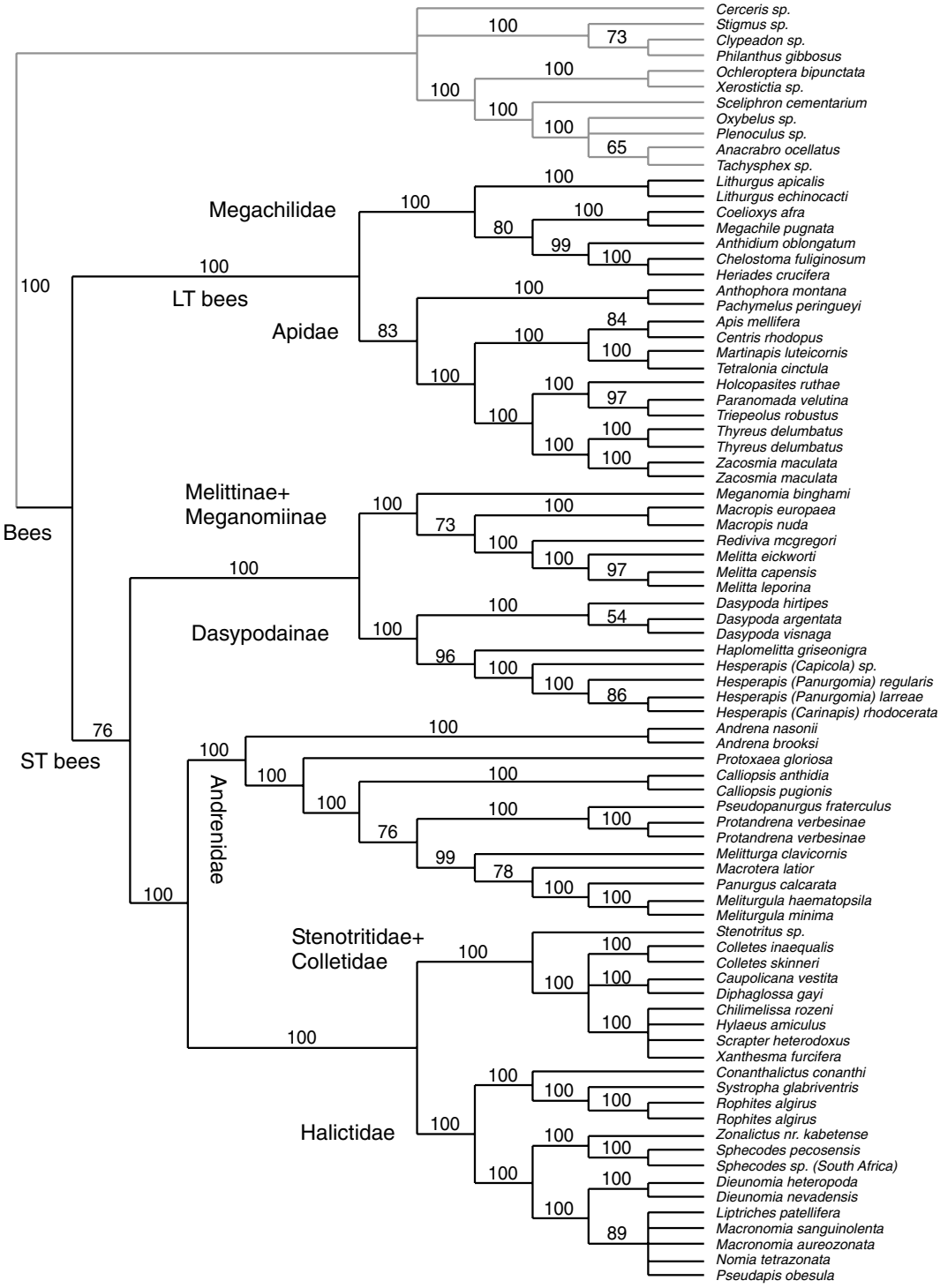

Fig. 5. Majority rule consensus of trees 2000–10,000 in the Bayesian analysis using the GTR + I + G model with a separate model applied to each of the three genes. Values above the branches are posterior probabilities. Average Ln likelihood = −534959.06.

informative, whereas Pol II shows little nt1 and virtually no nt2 variation. This is also reflected in the relative rates obtained from the GTR + SSR Bayesian analysis (Fig. 6). First and second nucleotide positions in Pol II evolve more slowly than first and second position sites in CAD. The differences in relative rates may partly explain the differences between CAD and Pol II in partitioned Bremer support. One would expect that genes with more nt1 and nt2 variation should be less prone to saturation than genes with exclusively nt3 variation (such as Pol II). Overall, CAD appears to be the more robust data set, but the fact that Pol II is easier to amplify may make Pol II a more universally applicable data set for hymenopteran phylogenetic analyses.

Table 4
Summary of support measures for different lineages of bees

	Bootstrap	K2P + SSR	HKY + SSR	GTR + SSR	GTR + I + G	GTR + SSR + G[28S]
Bee monophyly	82	100	99	100	100	100
LT bee monophyly	na	100	99	100	100	100
ST bee monophyly*	na	100	84	99	76	75
Megachilidae	96	100	100	100	100	100
Apidae	na	na	na	na	100	na
Melittidae*	na	100	100	100	100	100
Melittinae*	70	61	56	91	73	99
Melittinae + Meganomiinae	na	100	100	100	100	100
Dasypodainae	100	100	100	100	100	100
Andrenidae	80	100	100	100	100	100
Oxaeinae + Panurginae	65	100	100	100	100	100
Panurginae	100	100	100	100	100	100
Halictidae	100	100	100	100	100	100
Rophitinae	100	100	100	100	100	100
Halictinae	100	100	100	100	100	100
Nomiinae	100	100	100	100	100	100
Colletidae	99	100	100	100	100	100
Stenotridae + Colletidae	99	100	100	100	100	100
Halictidae, Steno., Colletidae	88	100	100	100	100	100
And., Halic., Steno., Colletidae	74	100	100	100	100	100
Ave. Ln likelihood		−56334.97	−56088.24	−555565.48	−53459.06	−54349.40

Groups indicated by an asterisk (*) are considered weakly supported. Average -Ln likelihood scores were based on the last 8000 generations from each analysis.

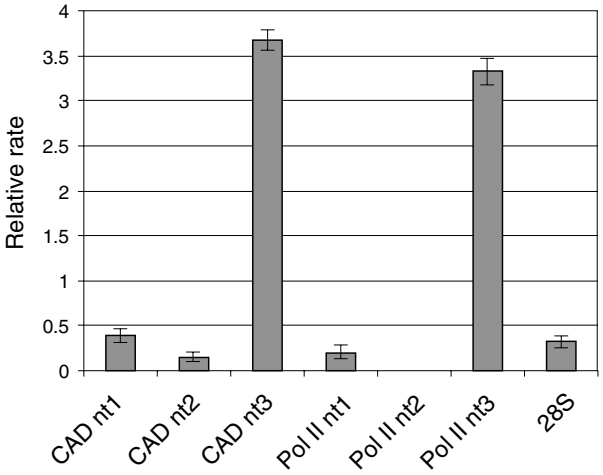

Fig. 6. Relative rates among the three codon positions for CAD and Pol II and 28S obtained from the Bayesian GTR + SSR model.

4. Discussion

4.1. Utility of CAD and RNA polymerase II

Judging by our measures of phylogenetic utility, including consistency index, number of nodes recovered, number of nodes recovered with >50% bootstrap support, data decisiveness, and partitioned Bremer support, CAD appears to be a more useful and informative gene than Pol II. CAD shows more first and second codon position variation, which may partly explain this difference in performance. CAD also shows a slightly less biased base composition overall than Pol II. 28S seems to be intermediate between these two genes. 28S recovers more nodes than Pol II but fewer nodes than CAD. 28S shows greater data decisiveness than either protein-coding gene, but contributes less to overall Bremer support than either protein-coding gene (Table 3). The results from 28S and CAD are highly congruent, with both genes supporting many of the same family, subfamily, and supra-familial groupings. In the case of the protein-coding genes, alignments were unambiguous within the coding regions which is an advantage over 28S, where alignments are problematic. We would recommend the use of CAD in phylogenetic studies of bees and other Hymenoptera in spite of the fact that this gene cannot be readily amplified in all taxa.

4.2. Insights into higher-level bee phylogeny

This is the first study to evaluate the utility of nuclear genes for resolving family-level phylogenetic questions in bees. It is clear that to obtain robust support among basal nodes in the tree will require substantially more data and we are in the process of developing more single-copy nuclear gene data sets for such genes as EF-1α, wingless, LW rhodopsin, Na/K ATPase, and others. However, our current results provide some fascinating new insights into higher-level bee phylogeny and unambiguously support the "Melittidae-LT basal" topology (Fig. 1B). The "Colletidae basal" topology is more widely accepted because of the perception that the bifid glossa of Colletidae is a plesiomorphic trait shared with the spheciform wasps. This hypothesis appears in numerous publications but is rarely supported by characters other than the overall appearance

of the glossa (Engel, 2001; Malyshev, 1968; Michener, 1944, p. 230, 1974, p. 23, 1979, p. 299). Most recently, Engel (2001) published a tree based neither on data nor an explicit method of phylogenetic analysis, which supports the "Colletidae-basal" hypothesis. He states "By undertaking a cladistic reconstruction of the Apiformes [*for which no data are presented* (italics added)] it is possible to obtain a generalized picture of the groundplan morphological and biological attributes that were likely present in the ancestor of all bees. The Colletidae are indeed the most basal family of Apiformes." (Engel, 2001, p. 155–156).

However, several authors have questioned this interpretation, including Michener (2000, p. 86): "I believe one still must say that we do not know whether the colletid glossal shape is a plesiomorphy derived from sphecoid wasps or a synapomorphy of female and most male colletids, although to me the latter seems more likely." Radchenko and Pesenko (1994) expressed a similar opinion, arguing that the biology of the common ancestor of bees (the "proto-bee") was more similar to that of a melittid or andrenid than any member of Colletidae. McGinley (1980), based on a detailed study of glossal morphology in all families of bees and spheciform wasps, stated that "examination shows the colletid glossa to possess several apparently derived features and to be only superficially similar to that of the sphecoid wasps" (McGinley, 1980, p. 546). He went on to conclude that "the most primitive bees had an acute glossa, still retained in some male colletids as well as in other families of bees. The truncate or bifid glossa appears to have been derived from the acute form, presumably in connection with the preparation by females of cell linings" (McGinley, 1980, p. 549). Most recently, Michener (2005) highlighted the morphological similarity between the pointed glossa of *Pseudoscolia* (a genus of wasps in the Crabronidae) and bees of the families Andrenidae, Melittidae, Halictidae, Apidae, and Megachilidae. In summary, the "Colletidae basal" hypothesis is not supported by any convincing morphological characters and is primarily based on the incorrect interpretation of the bifid glossal morphology as primitive. Stated another way, there is no single morphological character that unites the non-colletid bees other than the acutely pointed glossa, which is most likely a symplesiomorphy.

The one bee family whose monophyly is most in question is the Melittidae. Alexander and Michener (1995) failed to find strong support for melittid monophyly and concluded that, while the subfamilies are each monophyletic, the family is most likely paraphyletic with respect to the LT bees. Larval studies failed to support monophyly of the family (McGinley, 1981) and Michener (1981) found no adult or larval synapomorphies for the family. Roig-Alsina and Michener (1993) described an "exceedingly small and inconspicuous" sclerite between the cardo and stipes (p. 157) that they stated was a synapomorphy for the family. However, examination of the data matrix (p. 139) indicates the character is absent in some Melittidae. Alexander and Michener (1995) stated that "we found this sclerite not only hard to find but absent in some preparations, and did not include it in our study" (p. 418). In summary, no morphological studies have convincingly supported melittid monophyly. Our results are similarly ambiguous. While the parsimony results showed Dasypodainae as the basal branch of the bees, the Bayesian results unambiguously recovered melittid monophyly. Given the discrepancy between the parsimony results and the Bayesian results, and the lack of strong bootstrap support at the base of the parsimony tree, we continue to view melittid monophyly as questionable.

The "Melittidae-LT basal" hypothesis may help explain a number of things about both the bee fossil record and the biogeography of bees. One of the most puzzling aspects of the bee fossil record is the seeming overabundance of Melittidae, Apidae, and Megachilidae in the oldest deposits, such as the Eocene (Baltic) amber (Engel, 2001; Michez et al., 2006) and the Cretaceous amber from New Jersey (Engel, 2000; Michener and Grimaldi, 1988a,b). Among the bees in the Baltic amber deposits, 15 of the 18 described genera are LT bees (in the families Apidae and Megachilidae) (Engel, 2001). Melittid bees are also seemingly well represented in the Eocene both from the Baltic amber (*Eomacropis*; Engel, 2001) as well as from the French Eocene amber (*Paleomacropis*; Michez et al., 2006). *Paleomacropis* (Michez et al., 2006) is a particularly fascinating fossil because it is closely related to extant oil-collecting bees in the genus *Macropis* and clearly bears the morphological attributes of an oil collecting bee. The oldest fossil bee, *Cretotrigona prisca*, is an apid bee closely related to extant genera within the tribe Meliponini (Engel, 2000; Michener and Grimaldi, 1988a,b). In contrast, other ST bee families, such as Andrenidae and Halictidae are much less well represented in the Eocene, and representatives of Colletidae are completely absent in the fossil record up until the Miocene (Dominican) amber (Engel, 1999; Michener and Poinar, 1996). This excess of Melittidae and LT bees in the oldest fossil deposits has generally been interpreted as an artifact due to the poor fossil record of bees, and possibly a bias toward resin collecting bees, most of which are LT bees (Roig-Alsina and Michener, 1993). However, if one accepts the "Melittidae-LT basal" hypothesis, the fossil record and the phylogeny are fully congruent. Both Melittidae and LT bees are early branches of the phylogeny of the bees, and are therefore relatively old compared to some families of ST bees, such as Halictidae, Andrenidae, Colletidae, and Stenotritidae.

Biogeographic distributions of certain bee groups are also consistent with the "Melittidae-LT basal" hypothesis. Among the genera of Melittidae there are several groups which exhibit disjunct distributions, suggesting that these groups represent ancient lineages that have undergone substantial extinction. *Hesperapis*, for example, is distributed mostly in Western North America (with one eastern species restricted to the dune fields of barrier islands and coastal margins of the northern Gulf of Mexico; Cane et al., 1996) and arid regions of Southern Africa (Michener, 2000). Bees

in the megachilid tribe Fideliini occur primarily in Southern Africa (with one species in Morocco) and Chile. Engel (2002) was reluctant to consider this a group that reflects vicariance between Africa and South America because of his perception that the LT bees are nested well within the phylogeny of bees. If LT bees are an early branch of bee phylogeny, as implied by the "Melittidae-LT basal" hypothesis, the early diversification of Fideliini may indeed have occurred prior to the separation of Africa and South America ≈100 myBP. (Smith et al., 1994). If true, the "Melittidae-LT basal" topology will substantially alter our current understanding of bee phylogeny, biogeography, evolution, and early diversification.

Acknowledgments

This project was supported by collaborative National Science Foundation Research Grants in Systematic Biology to B.N.D. (DEB-0211701) and S.S. (DEB-0211837). A National Geographic Grant-in-Aid of Research (NGS Grant No. 6946-01) provided travel funds for the collection of some important specimens for this study. This grant involved a collaboration between Connal Eardley (Plant Protection Research Institute, Pretoria, and South Africa), Kenneth Walker (Museum of Victoria, Victoria, Australia), Laurence Packer (York University, York, Canada), and B.N.D. on the biology and systematics of the African halictid bees. We are grateful to the authors of the honey bee genome (http://www.hgsc.bcm.tmc.edu/projects/honeybee/), which has made it possible for us to develop new genes for higher-level bee phylogeny. We are also grateful to several collaborators who provided important specimens for this study: John Ascher, Laurence Packer, Jerome Rozen, Jr., Margrit McIntosh, Terry Houston, Robert Brooks, and David Moeller. The following people commented on earlier drafts of this paper: Eduardo Almeida, Sophie Cardinal, and Sean Brady.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2005.09.022.

References

Alexander, B.A., Michener, C.D., 1995. Phylogenetic studies of the families of short-tongued bees (Hymenoptera: Apoidea). Univ. Kans. Sci. Bull. 55 (11), 377–424.

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

Baker, R.H., Yu, X.B., DeSalle, R., 1998. Assessing relative contribution of molecular and morphological characters in simultaneous analysis trees. Mol. Phylogenet. Evol. 9, 427–436.

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

Barth, F.G., 1991. Insects and Flowers: The Biology of a Partnership. Princeton University Press, Princeton, NJ (x+408 pp.).

Batra, S.W.T., 1980. Ecology, behavior, pheromones, parasites and management of the sympatric vernal bees *Colletes inaequalis*, *C. thoracicus*, and *C. validus*. J. Kans. Entomol. Soc. 53, 509–538.

Brady, S.G., Danforth, B.N., 2004. Recent intron gain in elongation factor-1α (EF-1α) of colletid bees (Hymenoptera: Colletidae). Mol. Biol. Evol. 21 (4), 691–696.

Bremer, K., 1988. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. Evolution 42, 795–803.

Cameron, S.A., Mardulyn, P., 2001. Multiple molecular data sets suggest independent origins of highly eusocial behavior in bees (Hymenoptera: Apinae). Syst. Biol. 50 (2), 192–214.

Cane, J., Snelling, R.R., Kervin, L.J., 1996. A new monolectic coastal bee, *Hesperapis oraria* Snelling and Stage (Hymenoptera: Melittidae), with a review of desert and neotropical disjunctives in the Southeastern US. J. Kans. Entomol. Soc. 69 (Suppl. 4), 238–247.

Caterino, M.S., Cho, S., Sperling, F.A.H., 2000. The current state of insect molecular systematics: a thriving Tower of Babel. Annu. Rev. Entomol. 45, 1–54.

Danforth, B.N., 2002. Evolution of sociality in a primitively eusocial lineage of bees. Proc. Natl. Acad. Sci. USA 99 (1), 286–290.

Danforth, B.N., Brady, S.G., Sipes, S.D., Pearson, A., 2004. Single copy nuclear genes recover Cretaceous age divergences in bees. Syst. Biol. 53 (2), 309–326.

Danforth, B.N., Sauquet, H., Packer, L., 1999. Phylogeny of the bee genus *Halictus* (Hymenoptera: Halictidae) based on parsimony and likelihood analyses of nuclear EF-1α sequence data. Mol. Phylogenet. Evol. 13 (3), 605–618.

Dowton, M., Austin, A.D., 2002. Increased congruence does not necessarily indicate increased phylogenetic accuracy—The behavior of the incongruence length difference test in mixed-model analyses. Syst. Biol. 51 (1), 19–31.

Engel, M.S., 1999. A new xeromelissine bee in Tertiary amber of the Dominican Republic (Hymenoptera: Colletidae). Entomol. Scand. 30, 453–458.

Engel, M.S., 2000. A new interpretation of the oldest fossil bee (Hymenoptera: Apidae). Am. Mus. Novitates 3296, 11.

Engel, M.S., 2001. A monograph on the Baltic amber bees and evolution of the Apoidea (Hymenoptera). Bull. Am. Mus. Nat. Hist. 259, 1–192.

Engel, M.S., 2002. Phylogeny of the bee tribe Fideliini (Hymenoptera: Megachilidae), with a description of a new genus from southern Africa. Afr. Entomol. 10 (2), 305–313.

Espelie, K.E., Cane, J.H., Himmelsbach, D.S., 1992. Nest cell lining of the solitary bee *Hylaeus bisinuatus*. Experientia 48, 414–416.

Farris, J.S., Källersjö, M., Kluge, A.G., Bult, C., 1995. Testing significance of incongruence. Cladistics 10, 315–319.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.

Freund, J.N., Jarry, B.P., 1987. The rudimentary gene of *Drosophila melanogaster* encodes four enzymic functions. J. Mol. Biol. 193, 1–13.

Goloboff, P.A., 1991. Homoplasy and the choice among cladograms. Cladistics 7, 215–232.

Grimaldi, D., 1999. The co-radiations of pollinating insects and angiosperms in the Cretaceous. Ann. Mo. Bot. Gard. 86, 373–406.

Hefetz, A., Fales, H.M., Batra, S.W.T., 1979. Natural polyesters: Dufour's gland macrocyclic lactones from brood cell laminesters in *Colletes* bees. Science 204, 415–417.

Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics (Oxford) 17 (8), 754–755.

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 (1), 87–92.

Kawakita, A., Sota, T., Ito, M., Ascher, J.S., Tanaka, H., Kato, M., Roubik, D.W., 2004. Phylogeny, historical biogeography, and character evolution in bumble bees (*Bombus*: Apidae) based on simultaneous analysis of three nuclear gene sequences. Mol. Phylogenet. Evol. 31 (2), 799–804.

Kim, H., Kelly, R.E., Evans, D.R., 1992. The structural organization of the hamster multifunctional protein CAD. J. Biol. Chem. 267 (10), 7177–7184.

Krzywinski, J., Besansky, N.J., 2002. Frequent intron loss in the *white* gene: A cautionary tale for phylogeneticists. Mol. Biol. Evol. 19, 362–366.

Leys, R., Cooper, S.J.B., Schwarz, M.P., 2002. Molecular phylogeny and historical biogeography of the large carpenter bees, genus *Xylocopa* (Hymenoptera: Apidae). Biol. J. Linn. Soc. 77, 249–266.

Liu, Y.J., Whelen, S., Hall, B.D., 1999. Phylogenetic relationships among Ascomycetes: evidence from an RNA Polymerase II subunit. Mol. Biol. Evol. 16 (12), 1799–1808.

Malyshev, S.I., 1968. Genesis of the Hymenoptera and Phases of Their Evolution. Methuen, London.

Mardulyn, P., Cameron, S., 1999. The major opsin in bees (Insecta: Hymenoptera): a promising nuclear gene for higher level phylogenetics. Mol. Phylogenet. Evol. 12, 168–176.

McGinley, R.J., 1980. Glossal morphology of the Colletidae and recognition of the Stenotritidae at the family level. J. Kans. Entomol. Soc. 53, 539–552.

McGinley, R.J., 1981. Systematics of the Colletidae based on mature larvae with phenetic analysis of apoid larvae, Hymenoptera: Apoidea. Univ. Calif. Publ. Entomol. 91, i–xvi (see also pp. 1–307).

Melo, G.A.R., 1999. Phylogenetic relationships and classification of the major lineages of Apoidea (Hymenoptera), with emphasis on the carbronid wasps. Scientific Papers, University of Kansas Natural History Museum 14, 1–55.

Michener, C.D., 1944. Comparative external morphology, phylogeny, and a classification of the bees. Bull. Am. Mus. Nat. Hist. 82, 151–326.

Michener, C.D., 1974. The Social Behavior of the Bees. Harvard University Press, Cambridge MA.

Michener, C.D., 1979. Biogeography of the bees. Ann. Mo. Bot. Gard. 66, 277–347.

Michener, C.D., 1981. Classification of the bee family Melittidae with a review of species of Meganomia. Amer. Entomol. Inst. 18, 1–135.

Michener, C.D., 2000. The Bees of the World. Johns Hopkins University Press, Baltimore, MD.

Michener, C.D., 2005. *Pseudoscolia*: a spheciform wasp with a pointed glossa (Hymenoptera: Crabronidae). J. Hymen. Res. 14, 78–83.

Michener, C.D., Grimaldi, D.A., 1988a. A *Trigona* from Late Cretaceous amber of New Jersey (Hymenoptera: Apidae: Meliponinae). Am. Mus. Novitates 2917, 10pp.

Michener, C.D., Grimaldi, D.A., 1988b. The oldest fossil bee: apoid history, evolutionary stasis, and antiquity of social behavior. Proc. Natl. Acad. Sci. USA 85, 6424–6426.

Michez, D., Nel, A., Menier, J.-J., 2006. The oldest Melittidae (Hymenoptera: Apoidea) from the early Eocene of Oise (France). Zool. J. Linn. Soc. (in press).

Michener, C.D., Poinar Jr., G.O., 1996. The known bee fauna of the Dominican amber. J. Kans. Entomol. Soc. 69, 353–361.

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 (1), 363–378.

Perkins, R.C.L., 1912. Notes, with descriptions of new species, on aculeate Hymenoptera of the Australian region. Ann. Mag. Nat. Hist. 8 (9), 96–121.

Radchenko, V.G., Pesenko, Y.A., 1994. Protobee and its nests: a new hypothesis concerning the early evolution of the Apoidea. Entomol. Rev. 75, 140–162 (English translation of Entomologischeskoe Obozrenie 73, 913–933).

Regier, J.C., Shultz, J.W., 1997. Molecular phylogeny of the major arthropod groups indicates polyphyly of crustaceans and a new hypothesis for the origin of hexapods. Mol. Biol. Evol. 14 (9), 902–913.

Regier, J.C., Shultz, J.W., 2001a. A phylogenetic analysis of Myriapoda (Arthropoda) using two nuclear protein-encoding genes. Zool. J. Linn. Soc. 132 (4), 469–486.

Regier, J.C., Shultz, J.W., 2001b. Elongation factor-2: a useful gene for arthropod phylogenetics. Mol. Phylogenet. Evol. 20 (1), 136–148.

Regier, J.C., Shultz, J.W., Kambic, R.E., 2004a. Phylogeny of basal hexapod lineages and estimates of divergence times. Ann. Entomol. Soc. Am. 97 (3), 411–419.

Regier, J.C., Shultz, J.W., Kambic, R.E., Nelson, D.R., 2004b. Robust support for tardigrade clades and their ages from three protein-coding nuclear genes. Invertebr. Biol. 123 (2), 93–100.

Regier, J.C., Wilson, H.M., Shultz, J.W., 2005. Phylogenetic analysis of Myriapoda using three nuclear protein-coding genes. Mol. Phylogenet. Evol. 34 (1), 147–158.

Roig-Alsina, A., Michener, C.D., 1993. Studies of the phylogeny and classification of long-tongued bees (Hymenoptera: Apoidea). Univ. Kans. Sci. Bull. 55, 123–173.

Shultz, 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 (1447), 1011–1019.

Shultz, J.W., Regier, J.C., 2001. Phylogenetic analysis of Phalangida (Arachnida, Opiliones) using two nuclear protein-encoding genes supports monophyly of Palpatores. J. Arachnol. 29 (2), 189–200.

Smith, A.G., Smith, D.G., Funnel, B.M., 1994. Atlas of Mesozoic and Cenozoic Coastlines. Cambridge University Press, Cambridge, UK. x + 99 pp.

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

Swofford, D.L., 2002. PAUP*. Phylogenetic Analysis Using Parsimony (and other methods), v. 4.0b10. Sinauer Associates, Sunderland, MA.