Molecular Phylogenetics and Evolution 50 (2009) 290–309

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Phylogeny of colletid bees (Hymenoptera: Colletidae) inferred from four nuclear genes

Eduardo A.B. Almeida [a,b,*], Bryan N. Danforth [a]

[a] Cornell Univ., Dept. Entomol., Ithaca, NY 14853, USA
[b] Depto. de Zoologia, Universidade Federal do Paraná, Lab. Biol. Comparada Hymenoptera, Caixa Postal 19020, CEP 81531-980, Curitiba, Paraná, Brazil

ARTICLE INFO

Article history:
Received 8 May 2008
Revised 12 August 2008
Accepted 23 September 2008
Available online 22 October 2008

Keywords:
Colletidae
Bee phylogeny
Molecular phylogeny
Molecular evolution
Gondwana
Systematics

ABSTRACT

Colletidae comprise approximately 2500 species of bees primarily distributed in the southern continents (only two colletid genera are widely distributed: *Colletes* and *Hylaeus*). Previously published studies have failed to resolve phylogenetic relationships on a worldwide basis and this has been a major barrier to the progress of research regarding systematics and evolution of colletid bees. For this study, data from four nuclear gene loci: elongation factor-1α (F2 copy), opsin, *wingless*, and 28S rRNA were analyzed for 122 species of colletid bees, representing all subfamilies and tribes currently recognized; 22 species belonging to three other bee families were used as outgroups. Bayesian, maximum likelihood, and parsimony methods were employed to investigate the phylogenetic relationships within Colletidae and resulted in highly congruent and well-resolved trees. The phylogenetic results show that Colletidae are monophyletic and that all traditionally recognized subfamilies (except Paracolletinae) are also strongly supported as monophyletic. Our phylogenetic hypothesis provides a framework within which broad questions related to the taxonomy, biogeography, morphology, evolution, and ecology of colletid bees can be addressed.

1. Introduction

Colletidae are generally considered to constitute the most ancient bee lineage, i.e., the sister-group to all other bees (e.g., Michener, 1944, 1974, 1979; O'Toole and Raw, 1991; Engel, 2001). This hypothesis is not ubiquitous, though, and its validity has been questioned (e.g., McGinley, 1980; Michener, 2007; Danforth et al., 2006a,b). The most comprehensive studies of bee phylogeny were those by Alexander and Michener (1995) and Danforth et al. (2006b). Alexander and Michener (1995) analyzed a morphological matrix of adult and larval characters based on weighted and unweighted parsimony. Their data firmly established the monophyly of many bee families, but their results were inconclusive about the overall relationships *among* families. Tree topologies varied widely with different analytical methods and nodal support was weak, especially at the base of the tree. Danforth et al. (2006b) incorporated data from five nuclear genes and analyzed the data separately and in combination with Alexander and Michener's morphological matrix. Their results strongly pointed to the root of the bee clade being positioned among the lineages of "Melittidae" (*sensu lato*). Colletidae were placed far from the root node and statistical tests suggested that a basal placement of Colletidae was significantly incongruent with the data (Danforth et al., 2006b).

1.1. Colletid monophyly

Some of Alexander and Michener's (1995) analyses did not strongly support the monophyly of Colletidae, even though there are a number of unique traits, which appear to be synapomorphies for this family. The cellophane-like cell lining (e.g., Hefetz et al., 1979; Espelie et al., 1992; reviewed by Almeida, 2008a) has been considered to be the strongest evidence for monophyly of the family because it is a unique and unreversed character found only among colletid bees. McGinley (1980) presented a list of potential synapomorphies from mouthpart morphology. The most recent evidence for the monophyly of this family was given by a molecular character: the presence of a unique intron in the F1 copy of the gene elongation factor-1α in all colletids sampled, but not in Stenotritidae or any other bee family (Brady and Danforth, 2004). Additionally, Colletidae have been consistently recovered as a strongly supported monophyletic group based on several independent molecular data sets (Danforth et al., 2006a,b).

1.2. Diversity and classification of Colletidae

The Colletidae, as considered here, are currently divided into seven subfamilies: Colletinae, Euryglossinae, Hylaeinae, Paracolleti-

* Corresponding author. Address: Depto. de Zoologia, Universidade Federal do Paraná, Lab. Biol. Comparada Hymenoptera, Caixa Postal 19020, CEP 81531-980, Curitiba, Paraná, Brazil. Fax: +55 41 3361 1763.
E-mail address: eabalmeida@gmail.com (E.A.B. Almeida).

1055-7903/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2008.09.028

nae, Scrapterinae, and Xeromelissinae. These subfamilies correspond to the five subfamilies recognized by Michener (2007), except that the three tribes of Colletinae: Colletini, Paracolletini, and Scrapterini are treated as independent subfamilies. This classification is also congruent with the classification by Melo and Gonçalves (2005), except that those authors treat the subfamilies of Colletidae as tribes.

Colletid bees are diverse, ranging from small, slender, relatively hairless bees (such as Euryglossinae) to large, robust, hairy bees (such as Diphaglossinae). Females carry pollen either externally, in a well-developed trochanteral and femoral scopa (Colletinae and Diphaglossinae) or a scopa formed by sparse hairs on the hind legs and long hairs primarily on the second abdominal sternum (Xeromelissinae); or internally in the crop (Euryglossinae, Hylaeinae, and few Paracolletinae). There is only one known group of cleptoparasites (some Hawaiian species of the *Hylaeus* (*Nesoprosopis*) Perkins [Daly and Magnacca, 2003]). Floral relationships in Colletidae range from polylectic (generalist) to oligolectic (specialist), with some taxa showing very narrow host-plant preferences (e.g., Wcislo and Cane, 1996). The highest diversity of colletid bees is observed in the temperate parts of southern South America and in Australia. Worldwide, there are 2485 available species names in Colletidae (Ascher et al., 2008).

The Colletinae are widely distributed (they do not occur in Australia, though) and comprise approximately 480 species. Colletinae are morphologically homogenous compared to the remainder of Colletidae, and their body size varies from 7 to 16 mm in length. The Diphaglossinae include the largest colletid bees (up to 24 mm in length). These Neotropical bees are, in general, fast flying, and many of them fly only early in the morning or before and around dusk. There are 128 available species names. This is the only colletid subfamily subdivided into tribes, and Michener (1986) resolved the tribal relationships as follows: (Caupolicanini, (Diphaglossini, Dissoglottini)). The Euryglossinae are a subfamily occurring mostly in the Australian Region (one species has been introduced in South Africa) and comprise small, hairless bees. There are almost 400 species of Euryglossinae. The Hylaeinae comprise more species than any other subfamily of Colletidae, with about 900 available names. Morphological and taxonomic diversity is highest in Australia where the group is thought to have originated; all species occurring outside the Australian region belong in the genus *Hylaeus*. Their wasp-like appearance and lack of external pollen-carrying structures made Hylaeinae candidates for the most primitive bees (e.g., Jander, 1976; Michener, 1979; see also discussion in Michener, 2007, 88–92). Females of Euryglossinae also lack pollen-carrying structures (i.e., a scopa) and transport pollen internally. The Paracolletinae are a morphologically diverse group of bees that ranges in size from 6 to 18 mm in length and there are over 400 described species, distributed throughout the Australian and Neotropical regions, mainly in subtropical and temperate dry biomes. The Scrapterinae are a monogeneric subfamily comprising 40 species of the endemic African genus *Scrapter*, most of which are distributed in southern Africa, especially the Cape region. One species was recently described from Kenya (Davies et al., 2005). Host-plant preferences and adult morphology are quite variable within this group of bees; body length varies from 3.5 to 14 mm in length (Davies and Brothers, 2007). Research on the taxonomy and biology of *Scrapter* (Rozen and Michener, 1968; Eardley, 1996; Davies et al., 2005; Davies and Brothers, 2007) makes it one of the best-studied groups of Colletidae. Until recently, *Scrapter* was considered as part of Paracolletinae. The Xeromelissinae are small, slender, and not very hairy, even though they have a small scopa on the hind leg and metasoma. Body length varies from 2.5 to 7.0 mm in length. There are almost 120 described species with the highest diversity in temperate regions of Chile and Argentina, but the group extends through the Neotropical Region as far north as Mexico. Packer (2008) and Almeida et al. (2008) present results of two phylogenetic studies of Xeromelissinae based on morphology and combined morphological and molecular data, respectively.

Although the composition of colletid subfamilies is largely settled, relationships among them are poorly understood. One of the few points of agreement regarding the phylogenetic affinities within Colletidae is the grouping of Euryglossinae, Hylaeinae, and Xeromelissinae. A graphical summary of four hypotheses of colletid phylogeny is presented in Fig. 1.

Alexander and Michener (1995) investigated the relationships among the main lineages of bees based on a morphological data set of adult and larval characters. Although a large number of colletid species were included in their analyses, the conclusions one can draw from their results are limited by the extremely variable positioning of the groups depending on the way data were analyzed. A summary of one of the consensus trees presented by Alexander and Michener (1995) is shown in Fig. 1d. Examination of previous hypotheses of relationships among lineages of Colletidae based either on taxonomic experience and intuition of bee researchers (Fig. 1a–c), or on explicit data analysis (Fig. 1d) reveals great uncertainty surrounding the relationships within this family.

The purpose of this study is to re-assess the phylogenetic relationships within Colletidae. This was done using a novel source of data for this group: genetic sequence data from four nuclear gene loci. A better understanding of colletid phylogeny is essential for providing the rationale of an improved classification of these bees, as well as for the reconstruction of their evolutionary and biogeographical history.

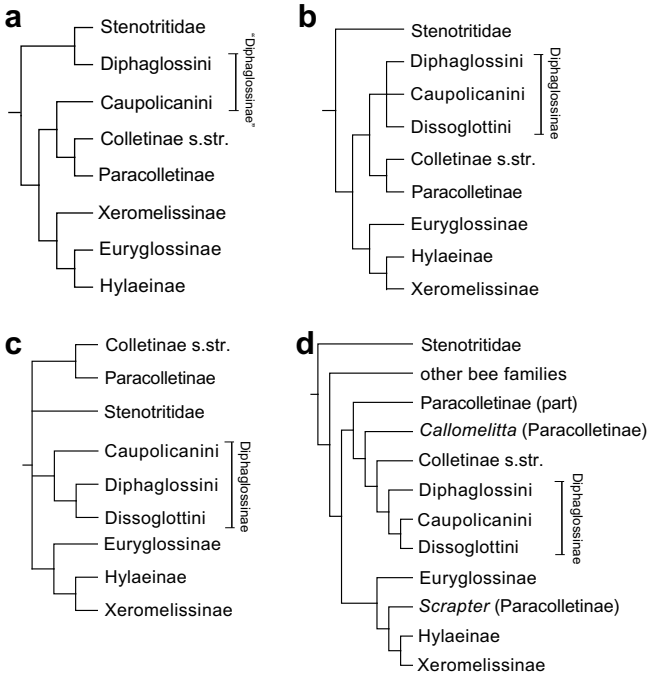

Fig. 1. Historical account of attempts to describe relationships among subfamilies and tribes of colletid bees as shown in summary trees of: (a) Michener (1944, p. 230); (b) Michener (1974, p. 23); (c) Engel (2001, p. 156); (d) Alexander and Michener (1995, p. 398), consensus of four trees resulting from the analysis of the complete data set using implied weights parsimony. The African genus *Scrapter* used to be considered part of Paracolletinae when these four hypotheses were presented. Stenotritidae is also included in the figures because, historically, it has been considered to be part of Colletidae by several authors.

2. Materials and methods

2.1. Taxon sampling

We assembled a data set that comprises a total of 144 terminal species; 122 species of Colletidae were sampled to represent all traditionally accepted subfamilies and tribes of this family. Special attention was paid to the Colletinae s.l. (*sensu* Michener, 2007; i.e., Colletinae + Paracolletinae + Scrapterinae) because it is widely believed to be para- or polyphyletic. Eighty species of Colletinae s.l. were included in the analysis (15 Colletinae s.str., 60 Paracolletinae, and 5 Scrapterinae). The remaining 42 species included representatives of the three tribes of Diphaglossinae [Caupolicanini (9 spp.), Diphaglossini (4 spp.), and Dissoglottini (2 spp.)]; Euryglossinae (6 spp.); Hylaeinae (11 spp.); and Xeromelissinae (8 spp.).

Twenty-two representatives of three bee families were used as outgroups for this analysis: Andrenidae (8 spp.), Halictidae (10 spp.), and Stenotritidae (4 spp.). These represent the main lineages of bees closely related to Colletidae according to the results of Danforth et al. (2006a,b). This choice of outgroups does not imply an agreement with a given placement of Colletidae relative to other bee lineages. The analysis of a very similar molecular data set with taxon sampling that intended to resolve the relationships among bee families resulted in the following arrangement for the closest groups to Colletidae: (Andrenidae (Halictidae (Colletidae, Stenotritidae))) (Danforth et al., 2006b). This means that the relationships among bee families based on a data set of this sort are well-established. The purpose of outgroups is to establish the position of the ingroup root node, and Andrenidae, Halictidae, and Stenotritidae should provide sufficient information for rooting Colletidae. Even if Colletidae are indeed the earliest lineage of bees, the outgroup choice still remains unproblematic. In Nixon and Carpenter's (1993, 423) words: "[o]utgroup(s) need not be "primitive" relative to the ingroup".

Specimens used for sequencing were primarily preserved in 95% EtOH but recently collected pinned specimens and frozen specimens were also used. Pinned specimens older than 3–5 years were not suitable for DNA extractions but those collected more recently provided good quality, high-molecular weight DNA for PCR. Outgroup and ingroup taxa included in this study, locality data, specimen voucher numbers, and GenBank Accession Nos. are listed in Table 1. The detailed generic classification adopted in this study is presented in Appendix 1 (Supplementary materials). Voucher specimens are housed in the Cornell University Insect Collection or at the institutions that provided specimens for DNA extractions. Identification of various taxa included in the analyses was facilitated by comparison to material deposited in entomological collections and assistance of other bee taxonomists, especially J.S. Ascher (American Museum of Natural History, USA), T. Houston (Western Australian Museum, Australia), M. Kuhlmann (University of Muenster, Germany), G.A.R. Melo (Universidade Federal do Paraná, Brazil), and L. Packer (York University, Canada).

2.2. Data

Molecular data were collected from four nuclear loci that have been providing robust results for insect phylogenetic studies. Three of them are nuclear protein-coding genes: EF-1α (elongation factor-1α, F2 copy), opsin (long wavelength, green rhodopsin), and *wingless* (*wg*), and the fourth is a nuclear ribosomal RNA locus: the D1–D5 expansion regions and related core elements of the large subunit 28S rRNA (28S rRNA). Maps of the four gene loci and the relative positions of the primers used in this study are shown in Fig. 2, and primer information for each gene can be found in Table 2. Each of these gene loci is referred to as a partition hereafter; introns and exons of EF-1α are treated as two separate partitions. Five partitions are therefore recognized: exons and introns of EF-1α, opsin (exons), *wg* (exons), and 28S rRNA.

Slowly evolving, nuclear genes are commonly used for phylogenetic analysis in many groups of insects. They have been demonstrated to recover Cretaceous-age divergences (e.g., Danforth et al., 1999, 2004, 2006a,b; Wiegmann et al., 2000). Danforth et al. (2004, 310–311) reviewed the phylogenetic utility of EF-1α, opsin, and *wg*, and commented on the biological functions of their gene products. Among these genes, EF-1α has been the most widely used nuclear protein-coding gene for insect phylogenetics. Despite alignment challenges of some ribosomal DNA regions, these genes cannot be distinguished from protein-coding genes regarding phylogenetic utility and nucleotide substitution patterns (Danforth et al., 2005). Among the various ribosomal gene loci, 18S rRNA tends to have the slowest rate of substitution (Hillis and Dixon, 1991), while 28S rRNA has more variation and, therefore, more phylogenetic signal.

The three protein-coding genes used here comprised the data set used by Danforth et al. (2004) to assess phylogenetic relationships within Halictidae, except that the exon of *wg* sampled for the present study is approximately 250 bp longer on the upstream (i.e., 5′ end, Sipes, personal communication). The data set of Danforth et al. (2004) was robust enough to resolve relationships among halictid bees. Halictidae and the clade formed by Colletidae and Stenotritidae are presumably roughly the same age, both having originated between mid- and late-Cretaceous, are closely related, and have approximately the same species diversity. This provides the basis for the initial assumption that this molecular data set should have enough phylogenetic signal to resolve relationships within either family.

2.3. DNA extraction

Genomic DNA was extracted using phenol–chloroform protocols (Doyle and Doyle, 1990, adapted by Danforth, 1999), but without use of liquid nitrogen and RNase. Tissue was taken from the thoracic musculature and/or legs depending on the rarity and size of available specimens. Samples were (1) macerated in individual 1.5 ml Eppendorf tubes with 2× CTAB extraction buffer and 100 mg Proteinase K; (2) incubated for 2–4 h at 55 °C; (3) extracted with 24:1 chloroform–isoamyl alcohol, (4) extracted again with 25:24:1 phenol–chloroform–isoamyl alcohol; and (5) 24:1 chloroform–isoamyl alcohol.

The phenol–chloroform–isoamyl alcohol stage was performed in Phase-Lock Gel® 2.0 ml Eppendorf tubes to facilitate the separation of phenol from the remainder and thus increase the final DNA yield. DNA was (1) precipitated with 2.5 volumes of ice-cold 100% ethanol and 0.1 volume 3 M sodium acetate; (2) washed with 80% ethanol; and (3) resuspended in 50 ml Tris–EDTA (pH 7.6) buffer.

2.4. PCR and sequencing

PCR amplifications of the genes listed above were done for 35 cycles under the following conditions: an initial denaturation at 94 °C for 60 s, followed by 35 cycles of denaturation at 94 °C for 60–90 s, annealing at 48–58 °C, 60–90 s, and extension at 72 °C, 60–90 s. Specific conditions for each locus amplified are listed in Table 2. Prior to sequencing, most PCR products were gel-purified in low melting point agarose gels (FMC, Rockland, Maine) overnight at 4 °C. DNA was recovered from gel slices using the Promega Wizard PCR Preps DNA Purification kit. Gel purification was unnecessary for PCR products that produced a single product: both fragments of 28S rRNA and the upstream (1100 bp) fragment of EF-1α. Automatic DNA sequencing was performed using the Applied Biosystems Automated 3730 DNA Analyzer employing Big Dye Termi-

Table 1
List of species included in this study, taxonomic information, locality data, and GenBank Accession nos.

Species [voucher code]	Classification	Collecting data	EF-1α	Opsin	*Wingless*	28S rRNA
Colletes bicolor Smith, 1879 [EA0082]	Colletidae: Colletinae	Argentina: Tucumán. Amaicha del Valle. 24.x.2004	DQ884650	DQ884549	DQ884802	DQ768539
Colletes compactus Cresson, 1868 [EA0014]	Colletidae: Colletinae	USA: New York. Tompkins Co., Ithaca. 29.ix.1998	DQ884642	DQ884542	DQ884794	DQ768531
Colletes distinctus Cresson, 1868 [EA0013]β	Colletidae: Colletinae	USA: California. SC: Berkeley Co., Honey Hill. 27.ii.2000	DQ884641	—	DQ884793	DQ768530
Colletes floralis Eversmann, 1852 [EA0076]	Colletidae: Colletinae	Mongolia: Arkhangay Aimag. 10.vii. 2004	DQ884649	DQ884548	DQ884801	DQ768538
Colletes furfuraceus Holmberg, 1886 [EA0108]	Colletidae: Colletinae	Argentina: Prov. Tucumán. Amaicha del Valle. 25–26.x.2004	DQ884651	DQ884550	DQ884803	DQ768540
Colletes gilvus Vachal, 1909 [EA0073]	Colletidae: Colletinae	Chile: Region I. Murmuntani (near Putre). Iv.2004	DQ884648	DQ884547	DQ884800	DQ768537
Colletes inaequalis Say, 1837 [Coin450]	Colletidae: Colletinae	USA: NY. Tompkins Co., Ithaca	AY585123	DQ115542	DQ884804	AY654484
Colletes pascoensis Cockerell, 1898 [EA0015]	Colletidae: Colletinae	USA: California. Contra Costa Co., El Cerrito. 14.iv.2001	DQ884643	—	DQ884795	DQ768532
Colletes seminitidus Spinola, 1851 [EA0045]	Colletidae: Colletinae	Chile: Region IV. PanAm N Los Hornos. 9.x.2002	DQ884647	DQ884546	DQ884799	DQ768536
Colletes simulans Cresson, 1868 [EA0016]	Colletidae: Colletinae	USA: New York. Tompkins Co., Ithaca. 21.viii.1997	DQ884644	DQ884543	DQ884796	DQ768533
Colletes skinneri Viereck, 1903 [Cosk632]	Colletidae: Colletinae	USA: AZ. Cochise Co., Chiricahua Monument. 14.ix.1999	AY230130	AY227912	DQ884805	AY654485
Colletes thoracicus Smith, 1853 [EA0017]	Colletidae: Colletinae	USA: Florida. Alachua Co., Gainsville. 15.iii.2002	DQ884645	DQ884544	DQ884797	DQ768534
Hemicotelles ruizii (Herbst, 1923) [EA0032]	Colletidae: Colletinae	Chile: Region IV. El Equi, Las Placetas. 14.x.2001	DQ884638	DQ884539	DQ884790	DQ768527
Rhynchocolletes mixtus (Toro and Cabezas, 1977) [EA0033]	Colletidae: Colletinae	Chile: Region IV. Pejerreyes. 19.x.2001	DQ884639	DQ884540	DQ884791	DQ768528
Xanthocotelles sicheli (Vachal, 1909) [EA0056]	Colletidae: Colletinae	Chile: Region VII. Curicó, Laguna de Teno. 1–5.ii.2003	DQ884640	DQ884541	DQ884792	DQ768529
Caupolicana bicolor Friese, 1899 [EA0041]	Colletidae: Diphaglossinae	Chile: Region II. Huasco, 4 km N Domeyko. 10.xi.2000	DQ884576	DQ884484	DQ884718	DQ768465
Caupolicana quadrifasciata Friese, 1898 [EA0040]	Colletidae: Diphaglossinae	Chile: Region IV. Fray Jorge Pq. Ntl. 11.x.2001	DQ884575	DQ884483	DQ884717	DQ768464
Caupolicana vestita (Smith, 1879) [Cpve848]	Colletidae: Diphaglossinae	Chile: Region I. Arica Playa las Machas.	AY585124	DQ115543	DQ884726	DQ872758
Caupolicana yarrowi (Cresson, 1875) [Cpya654]	Colletidae: Diphaglossinae	USA: NM. Hidalgo Co., 20 mi S. Animas. 24.ix.1999	—	DQ115544	DQ884727	AY654487
Ptiloglossa sp. [EA0018]	Colletidae: Diphaglossinae	Costa Rica: Prov. Guanacaste, Carmona. 15.i.2003	DQ884573	DQ884481	DQ884715	DQ768462
Ptiloglossa tarsata (Friese, 1900) [EA0085]	Colletidae: Diphaglossinae	Argentina: Salta. General Güemes. 11.xi.2004	DQ884581	DQ884488	DQ884723	DQ768470
Ptiloglossa thoracica (Fox, 1895) [EA0078]	Colletidae: Diphaglossinae	Mexico: Jalisco. Res. Chamela. 06.ix.2004	DQ884579	—	DQ884721	DQ768468
Willinkapis chalybaea (Friese, 1906) [EA0118]	Colletidae: Diphaglossinae	Argentina: Catamarca. 40 km N Andalgalá. 15.ii.2003	DQ884582	DQ884489	DQ884724	DQ768471
Zikanapis clypeata (Smith, 1879) [EA0119]	Colletidae: Diphaglossinae	Mexico: Jalisco. Res.Biosfera (Manantlan). 11.ix.2004	DQ884583	DQ884490	DQ884725	DQ768472
Cadeguala albopilosa (Spinola, 1851) [EA0036]	Colletidae: Diphaglossinae	Chile: Region VIII. Cordillera, El Manzano. xii.2001	DQ884574	DQ884482	DQ884716	DQ768463
Cadeguala occidentalis (Haliday, 1836) [EA0047]	Colletidae: Diphaglossinae	Chile: Region V. Colliguay. 19.x.2002	DQ884577	DQ884485	DQ884719	DQ768466
Cadegualina andina (Friese, 1925) [EA0048]	Colletidae: Diphaglossinae	Colombia: Boyacá. El Níspero. 12.xii.2001-19.i.2002	DQ884578	DQ884486	DQ884720	DQ768467
Diphaglossa gayi Spinola, 1851 [Diga850]	Colletidae: Diphaglossinae	Chile: Region X. Aquas Calientes.	AY585125	DQ115545	DQ884728	DQ872759
Mydrosoma aterrimum (Friese, 1925) [EA0156]	Colletidae: Diphaglossinae	Bolivia: La Paz, Prov. Coroico, Cerro Uchumachi. 05.iv.2004	EF032902	EF032903	EF032905	EF028342
Mydrosoma fallax (Moure, 1953) [EA0081]	Colletidae: Diphaglossinae	Argentina: Salta. General Güemes. 11.xi.2004	DQ884580	DQ884487	DQ884722	DQ768469
Callohesma calliopsella (Cockerell, 1910) [Euca688]	Colletidae: Euryglossinae	Australia: Victoria. Yan yaen. 20.xi.1999	AY585126	DQ115550	DQ884809	DQ872768
Euhesma aff. *crabronica* (Cockerell, 1914) [EA0155]	Colletidae: Euryglossinae	Australia: WA; Eurardy Stat. 09.x.2005	DQ884654	—	DQ884808	DQ768543
Euhesma platyrhina (Cockerell, 1915) [EA0148]	Colletidae: Euryglossinae	Australia: WA; Kalbarri Ntl.Prk. 08.x.2005	DQ884652	—	DQ884806	DQ768541
Euhesma sp. [EA0149]	Colletidae: Euryglossinae	Australia: WA. 15 km ESE Southern Cross. 23.ix.2005	DQ884653	—	DQ884807	DQ768542
Euryglossina globuliceps (Cockerell, 1918) [Eugl692]	Colletidae: Euryglossinae	Australia: Victoria. Colquhuon State Forest. 26.xi.1999	AY585127	DQ115551	DQ884810	DQ872769
Xanthesma furcifera (Cockerell, 1913) [Xnfu709]	Colletidae: Euryglossinae	Australia: Victoria. Patchewollock. 10.xii.1999.	AY585140	DQ115552	DQ884811	DQ872770
Amphylaeus (Agogenohylaeus) obscuriceps (Friese, 1924) [KM264]	Colletidae: Hylaeinae	Australia: Queensland. Kawana Waters. 10.xii.2002	DQ884687	DQ884564	DQ884857	DQ768597
Hylaeus (Euprosopis) disjunctus (Cockerell, 1905) [KM252]	Colletidae: Hylaeinae	Australia: Queensland. Somerset Dam. 14.xii.2002	DQ884677	DQ884563	DQ884848	DQ768586

(*continued on next page*)

Table 1 (*continued*)

Species [voucher code]	Classification	Collecting data	EF-1α	Opsin	*Wingless*	28S rRNA
Hylaeus (Euprosopis) elegans (Smith, 1853) [Hyel697]	Colletidae: Hylaeinae	Australia: South Australia. 10 km E Kimba. 5.i.1999	AY585129	DQ115547	DQ884839	DQ872778
Hylaeus (Gnathoprosopis) amiculus (Smith, 1879) [Hyam698]	Colletidae: Hylaeinae	Australia: South Australia. 10 km E Kimba. 5.i.1999	AY585128	DQ115546	DQ884838	DQ872777
Hylaeus (Macrohylaeus) alcyoneus (Erichson, 1842) [EA0129]	Colletidae: Hylaeinae	Australia: WA; Badgingarra Ntl.Prk. 11.x.2005	DQ884668	DQ884562	DQ884837	DQ768577
Hylaeus (Pseudhylaeus) aff. *simplus* Houston, 1993 [EA0125]	Colletidae: Hylaeinae	Australia: WA; Boorabbin Ntl.Prk. 26.ix.2005	DQ884664	DQ884561	DQ884833	DQ768573
Hylaeus (Rhodohylaeus) proximus (Smith, 1879) [Hypr699]	Colletidae: Hylaeinae	Australia: South Australia. 10 km E Kimba. 5.i.1999	AY585130	DQ115548	EF032906	DQ872779
Hyleoides concinna (Fabricius, 1775) [KM268]	Colletidae: Hylaeinae	Australia: Queensland. South of Eukey. 18.xii.2002	DQ884691	DQ884560	DQ884861	DQ768601
Meroglossa itamuca (Cockerell, 1910) [KM271]	Colletidae: Hylaeinae	Australia: Queensland. Noosa North Shore. 10.i.2003	DQ884694	DQ884565	DQ884863	DQ768604
Palaeorhiza (Heterorhiza) sp. [KM275]	Colletidae: Hylaeinae	Australia: Queensland. Kuranda. O4.i.2003	DQ884697	DQ884566	DQ884867	DQ768608
Palaeorhiza (Palaeorhiza) sp. [KM276]	Colletidae: Hylaeinae	Australia: Queensland. Kawana Waters. 10.xii.2002	DQ884698	DQ884567	DQ884868	DQ768609
Andrenopsis sp. [EA0094]	Colletidae: Paracolletinae	Australia: WA; ~18.4 km NE Menzies. 24.ix.2005	DQ884624	DQ884529	DQ884772	DQ768513
Anthoglossa sp. [EA0115]	Colletidae: Paracolletinae	Australia: WA; Boorabbin Ntl.Prk.25.ix.2005	DQ884585	DQ884492	DQ884730	DQ768474
Anthoglossa cfr. *robustus* (Cockerell, 1929) [EA0116]	Colletidae: Paracolletinae	Australia: WA; Kalbarri Ntl.Prk. 08.x.2005	DQ884586	DQ884493	DQ884731	DQ768475
Baeocolletes minimus (Michener, 1965) [EA0095]	Colletidae: Paracolletinae	Australia: WA; ~72.6 km NE Menzies. 27.ix.2005	DQ884608	DQ884514	DQ884756	DQ768497
Baeocolletes sp. [EA0096]	Colletidae: Paracolletinae	Australia: WA; Eurardy Station. 09.x.2005	DQ884609	—	DQ884757	DQ768498
Belopria nitidior Moure, 1956 [EA0027]	Colletidae: Paracolletinae	Brazil: Paraná. 15 km S Bocaiúva do Sul. 12.ix.2002	DQ884593	DQ884500	DQ884739	DQ768482
Brachyglossula communis Trucco-Aleman, 1999 [EA0083]	Colletidae: Paracolletinae	Argentina: Catamarca. El Rodeo. 20.xi.2004	DQ884600	DQ884506	DQ884748	DQ768489
Callomelitta antipodes (Smith, 1853) [Cman687]	Colletidae: Paracolletinae	Australia: NSW. Guyra, 74 km E. 5.xii.1999	AY585122	DQ115563	EF032907	DQ872767
Cephalocolletes isabelae Urban, 1995 [EA0012]	Colletidae: Paracolletinae	Brazil: Santa Catarina. Laguna. 31.xii.2001.	DQ884589	DQ884496	DQ884735	DQ768478
Cephalocolletes laticeps (Friese, 1906) [EA0086]	Colletidae: Paracolletinae	Argentina: San Juan. 14 km W Media Agua. 27.xi.2004	DQ884602	DQ884508	DQ884750	DQ768491
Chilicolletes delahozii (Toro, 1973) [Lesp568]	Colletidae: Paracolletinae	Chile: Region IV. Llano de la Hignera.	AF435392	AY227914	DQ884746	DQ872762
Colletellus aff. *velutinus* (Cockerell, 1929) [EA0097]	Colletidae: Paracolletinae	Australia: WA; Eurardy Stat. 09.x.2005	DQ884625	DQ884530	DQ884773	DQ768514
Edwyniana sp. [EA0071]	Colletidae: Paracolletinae	Chile: Region IV. Guampulla SW Samo Alto. 19.x.2001	DQ884598	DQ884504	DQ884745	DQ768487
Eulonchopria punctatissima Michener, 1963 [EA0077]	Colletidae: Paracolletinae	Mexico: Jalisco. Carretera 200, km 55. 02.ix.2004	DQ884599	DQ884505	DQ884747	DQ768488
Eulonchopria simplicicrus (Michener, 1989) [EA0009]	Colletidae: Paracolletinae	Brazil: Minas Gerais. Belo Horizonte. 06.v.2002	DQ884588	DQ884495	DQ884734	DQ768477
Euryglossidia sp.1 [EA0102]	Colletidae: Paracolletinae	Australia: WA; Boorabbin Ntl.Prk. 25.ix.2005	DQ884611	DQ884516	DQ884759	DQ768500
Euryglossidia sp.2 [EA0103]	Colletidae: Paracolletinae	Australia: WA; ~6 km E Merredin 23.ix.2005	DQ884612	DQ884517	DQ884760	DQ768501
Euryglossidia sp.3 [EA0104]	Colletidae: Paracolletinae	Australia: WA; Eurardy Station. 09.x.2005	DQ884613	DQ884518	DQ884761	DQ768502
Euryglossidia sp.4 [EA0151]	Colletidae: Paracolletinae	Australia: WA; Badgingarra Ntl.Prk. 11.x.2005	DQ884618	DQ884523	DQ884766	DQ768507
Euryglossidia sp.5 [EA0152]	Colletidae: Paracolletinae	Australia: WA; ~5.8 km NE Menzies 24.ix.2005	DQ884619	DQ884524	DQ884767	DQ768508
Excolletes sp. [EA0098]	Colletidae: Paracolletinae	Australia: WA; Kalbarri Ntl.Prk. 08.x.2005	DQ884626	—	DQ884774	DQ768515
Glossurocolletes bilobatus (Michener, 1965) [EA0093]	Colletidae: Paracolletinae	Australia: WA; Kalbarri Ntl.Prk. 06.x.2005	DQ884623	DQ884528	DQ884771	DQ768512
Goniocolletes fimbriatinus (Cockerell, 1910) [Lefm702]	Colletidae: Paracolletinae	Australia: Victoria. 12 km E Hattah. 6.i.1999	AY585131	DQ115554	DQ884786	DQ872763
Goniocolletes perfasciatus (Cockerell, 1906) [Lepr704]	Colletidae: Paracolletinae	Australia: Victoria. 12 km E Hattah. 9.i.1999	AY585134	DQ115557	DQ884787	DQ872764
Halictanthrena malpighiacearum Ducke, 1907 [EA0088]	Colletidae: Paracolletinae	Brazil: Minas Gerais. Serra do Salitre. 11.xi.2004	DQ884604	DQ884510	DQ884752	DQ768493
Hexantheda missionica Ogloblin, 1948 [EA0124]	Colletidae: Paracolletinae	Brazil: Minas Gerais. Brumadinho. 12.i.2001	DQ884615	DQ884520	DQ884763	DQ768504
Hoplocolletes ventralis (Friese, 1924) [EA0021]	Colletidae: Paracolletinae	Brazil: Minas Gerais. Florestal. 03.xii.2001	DQ884590	DQ884497	DQ884736	DQ768479
Kylopasiphae pruinosa (Michener, 1989) [EA0084]	Colletidae: Paracolletinae	Argentina: Mendoza. 37 km SSE Uspallata. 29–30.xi.2004	DQ884601	DQ884507	DQ884749	DQ768490

Table 1 (*continued*)

Species [voucher code]	Classification	Collecting data	EF-1α	Opsin	*Wingless*	28S rRNA
Lamprocolletes chalybeatus (Erichson, 1851) [EA0099]	Colletidae: Paracolletinae	Australia: WA. Boorabbin Ntl.Prk. 25.ix.2005	DQ884627	—	DQ884775	DQ768516
Leioproctus conospermi Houston, 1989 [EA0110]	Colletidae: Paracolletinae	Australia: WA; Kalbarri Ntl.Prk. 08.x.2005	DQ884632	DQ884534	DQ884780	DQ768521
Leioproctus irroratus (Smith, 1853) [Leir705]	Colletidae: Paracolletinae	Australia: NSW. Hilltop. 2.xii.1999	AY585132	DQ115555	DQ884788	DQ872765
Leioproctus lanceolatus Houston, 1990 [EA0111]	Colletidae: Paracolletinae	Australia: WA; ~90 km E Leonora. 28.ix.2005	DQ884633	DQ884535	DQ884781	DQ768522
Leioproctus megachalcoides Michener, 1965 [EA0112]	Colletidae: Paracolletinae	Australia: WA Eurardy Station. 09.x.2005	DQ884634	DQ884536	DQ884782	DQ768523
Leioproctus pappus Houston, 1989 [EA0109]	Colletidae: Paracolletinae	Australia: WA; Badgingarra Ntl.Prk. 11.x.2005	DQ884631	—	DQ884779	DQ768520
Leioproctus platycephalus (Cockerell, 1912) [EA0113]	Colletidae: Paracolletinae	Australia: WA; North Tarin Rock Nat. Res. 02.x.2005	DQ884635	DQ884537	DQ884783	DQ768524
Leioproctus plumosus (Smith, 1853) [Lepl706]	Colletidae: Paracolletinae	Australia: Victoria. Torquay. 19.xi.1999	AY585133	DQ115556	DQ884789	DQ872766
Paracolletinae sp. [EA0123]	Colletidae: Paracolletinae	Australia: WA. Badgingarra Ntl.Prk. 11.x.2005	DQ884637	DQ884538	DQ884785	DQ768526
Lonchopria (Biglossidia) robertsi Michener, 1989 [EA0080]	Colletidae: Paracolletinae	Argentina: Mendoza. 6 km SSE Uspallata. 29–30.xi.2004	DQ884622	DQ884527	DQ884770	DQ768511
Lonchopria (Biglossidia) sp. [EA0070]	Colletidae: Paracolletinae	Chile: Region I. Zapahuira (near Putre). Iv.2004	DQ884621	DQ884526	DQ884769	DQ768510
Lonchopria (Lonchopria) similis (Friese, 1906) [EA0069]	Colletidae: Paracolletinae	Chile: Region II. Pqe. Nacional Llanos de Challe. 13.x.2000	DQ884620	DQ884525	DQ884768	DQ768509
Neopasiphae mirabilis Perkins, 1912 [EA0100]	Colletidae: Paracolletinae	Australia: WA. ~18.4 km NE Menzies. 27.ix.2005	DQ884628	DQ884531	DQ884776	DQ768517
Niltonia virgilii Moure, 1964 [EA0087]	Colletidae: Paracolletinae	Brazil: Santa Catarina. Maracajá. 04.x.2004	DQ884603	DQ884509	DQ884751	DQ768492
Nomiocolletes jenseni Friese, 1906 [EA0090]	Colletidae: Paracolletinae	Argentina: Mendoza. 18 km SE Potrerillos. 29–30.xi.2004	DQ884606	DQ884512	DQ884754	DQ768495
Odontocolletes aff. *asper* (Maynard, 1997) [EA0105]	Colletidae: Paracolletinae	Australia: WA. Eurardy Station. 09.x.2005	DQ884629	DQ884532	DQ884777	DQ768518
Odontocolletes pachyodontus (Cockerell, 1915) [EA0106]	Colletidae: Paracolletinae	Australia: WA; ~13 km NNE Eurardy Station. 09.x.2005	DQ884630	DQ884533	DQ884778	DQ768519
Paracolletes cfr. *crassipes* Smith, 1853 [EA0019]	Colletitidae: Paracolletinae	Australia: Queensland, Stanthorpe. 16.xii.2002	DQ884584	DQ884491	DQ884729	DQ768473
Perditomorpha laena (Vachal, 1909) [EA0024]	Colletidae: Paracolletinae	Brazil: Minas Gerais. Santana do Riacho. 20/ii/2001.	DQ884592	DQ884499	DQ884738	DQ768481
Perditomorpha leucostoma (Cockerell, 1917) [EA0091]	Colletidae: Paracolletinae	Argentina: Catamarca. Aconquija. 18.xi.2004	DQ884607	DQ884513	DQ884755	DQ768496
Perditomorpha neotropica (Friese, 1908) [EA0147]	Colletidae: Paracolletinae	Argentina: Tucumán. 19 km SE Amaicha del Valle. 17.ii.2003	DQ884616	DQ884521	DQ884764	DQ768505
Perditomorpha rufiventris (Spinola, 1851) [EA0046]	Colletidae: Paracolletinae	Chile: Region IV. El Tofo. 23.x.2002	DQ884595	DQ884502	DQ884742	DQ768484
Perditomorpha stilborhina (Moure, 1954) [EA0107]	Colletidae: Paracolletinae	Argentina: Prov. Tucumán. Amaicha del Valle. 26.x.2004	DQ884614	DQ884519	DQ884762	DQ768503
Perditomorpha sp. [EA0068]	Colletidae: Paracolletinae	Argentina: Chubut. 8 km N Sarmiento Hwy.24. 26–27.xi.2003	DQ884597	DQ884503	DQ884744	DQ768486
Phenacolletes mimus Cockerell, 1905 [EA0101]	Colletidae: Paracolletinae	Australia: WA; ~12 km SSE Dongara. 10.x.2005	DQ884610	DQ884515	DQ884758	DQ768499
Protomorpha aff. *alloeopus* (Maynard, 1991) [EA0114]	Colletidae: Paracolletinae	Australia: WA; Kalbarri Ntl.Prk. 08.x.2005	DQ884636	—	DQ884784	DQ768525
Reedapis bathycyanea (Toro, 1973) [Lesp851]	Colletidae: Paracolletinae	Chile: Region II. Santa Juana, E of Vallenar	AY585141	DQ115553	DQ884741	DQ872761
Spinolapis caerulescens (Spinola, 1851) [EA0034]	Colletidae: Paracolletinae	Chile: Region IV. Parque Nacional Talinay 10.x.2001	DQ884594	DQ884501	DQ884740	DQ768483
Spinolapis sp. [EA0089]	Colletidae: Paracolletinae	Argentina: Mendoza. 37 km SSE Uspallata. 29–30.xi.2004	DQ884605	DQ884511	DQ884753	DQ768494
Tetraglossula anthracina (Michener, 1989) [EA0023]	Colletidae: Paracolletinae	Brazil: Minas Gerais. Santana do Riacho. 15.iv.2001	DQ884591	DQ884498	DQ884737	DQ768480
Trichocolletes (Trichocolletes) aff. *venustus* (Smith, 1862) [EA0117]	Colletidae: Paracolletinae	Australia: WA; Boorabbin Ntl.Prk. 25.ix.2005	DQ884587	DQ884494	DQ884732	DQ768476
Trichocolletes (Trichocolletes) sp. [Trsp708]	Colletidae: Paracolletinae	Australia: NSW. 53 km S Oberon. 30.xi.1999	AY585139	DQ115562	DQ884733	DQ872760
Scrapter algoensis (Friese, 1925) [Scal899]	Colletidae: Scrapterinae	South Africa: NCP. 90 km ENE Sprinbok. 10.ix.2001	EF032901	EF032904	DQ884812	DQ872771
Scrapter erubescens (Friese, 1925) [Scer901]	Colletidae: Scrapterinae	South Africa: WCP. Pakhuis pass. Sept. 8.ix.2001	AY585135	DQ115558	DQ884813	DQ872772
Scrapter heterodoxus (Cockerell, 1921) [Scht903]	Colletidae: Scrapterinae	South Africa: WCP. 31 km S Clanwillian. 7.ix.2001	AY585136	DQ115559	DQ884814	DQ872773
Scrapter niger Lepeletier and Serville, 1825 [Scng905]	Colletidae: Scrapterinae	South Africa: WCP. 21 km N Hermanus. 28.ix.2001	AY585137	DQ115560	DQ884815	DQ872774
Scrapter ruficornis (Cockerell, 1916) [Scrc937]	Colletidae: Scrapterinae	South Africa: WCP. Kunje Farm, near Citrusdal. 23.ix.2001	AY585138	DQ115561	DQ884816	DQ872775

(*continued on next page*)

Table 1 (*continued*)

Species [voucher code]	Classification	Collecting data	EF-1α	Opsin	*Wingless*	28S rRNA
Chilicola (Anoediscelis) herbsti (Friese, 1906) [EA0140]	Colletidae: Xeromelissinae	Chile: Region IV. Liman, Chañar. 04.ix.2004	DQ884663	DQ884559	DQ884831	DQ768572
Chilicola (Oediscelis) vicugna Toro and Moldenke, 1979 [EA0136]	Colletidae: Xeromelissinae	Chile: Region IV. Elqui, Pangue. 11–30.ix.2004	DQ884661	DQ884557	DQ884828	DQ768569
Chilicola (Pseudiscelis) rostrata (Friese, 1906) [EA0137]	Colletidae: Xeromelissinae	Argentina: Tucumán. 19 km SE Amaicha del Valle. 17.ii.2003	DQ884662	DQ884558	DQ884829	DQ768570
Geodiscelis longiceps Packer, 2005 [EA0049]	Colletidae: Xeromelissinae	Argentina: Tucumán. 19 km SE Amaicha del Valle. 17.ii.2003	DQ884655	DQ884551	DQ884817	DQ768544
Xenochilicola mamigna Toro and Moldenke, 1979 [EA0055]	Colletidae: Xeromelissinae	Chile: Region II. Aguas Blancas (SSE San Pedro Atacama)	DQ884660	DQ884556	DQ884827	DQ768558
Xeromelissa australis (Toro and Moldenke, 1979) [EA0051]	Colletidae: Xeromelissinae	Chile: Region II. Panamerican Hwy., km 1005, NE Chanaral.	DQ884656	DQ884552	DQ884818	DQ768545
Xeromelissa irwini (Toro and Moldenke, 1979) [EA0053]	Colletidae: Xeromelissinae	Chile: Region I. 83.5 km ESE Pozo Almonte. 8–20.iv.2004	DQ884658	DQ884554	DQ884821	DQ768547
Xeromelissa nortina (Toro and Moldenke, 1979) [EA0052]	Colletidae: Xeromelissinae	Argentina: Santa Cruz. 20 km E Los Antiguos. 17.xi.2003	DQ884657	DQ884553	DQ884820	DQ768546
Xeromelissa rozeni (Toro and Moldenke, 1979) [Chrz857]	Colletidae: Xeromelissinae	Argentina: Santa Cruz. 25 km E Los Antiguos 22.xi.2003	AY585120	DQ115549	DQ884826	DQ872776
Xeromelissa sp. [EA0138]	Colletidae: Xeromelissinae	Chile: Region I. 83.5 km ESE Pozo Almonte. 20.iv.2004	DQ884659	DQ884555	DQ884824	DQ768554
Alocandrena porteri Michener, 1986 [Alpo49]	Andrenidae: Andreninae	Peru: Lima Dept. St. Bartholome. 21.x.1997	AY585099	DQ113659	—	AY654473
Andrena brooksi Larkin, 2004 [Ansp643]	Andrenidae: Andreninae	USA: NM. Hidalgo Co., 20 mi S Animas. 17.ix.1999	AY230129	AF344618	AY222551	AY654474
Orphana wagenknechti Rozen, 1971 [Orph56]	Andrenidae: Andreninae	Chile: Region IV. 7 km S Pisco Elqui	DQ884568	DQ884476	DQ884709	DQ872755
Protoxaea gloriosa (Fox, 1893) [Pxgl226]	Andrenidae: Oxaeinae	USA: AZ. Cochise Co., Portal	AY585106	DQ113658	—	AY654480
Nolanomelissa toroi Rozen, 2003 [Noto74]	Andrenidae: Panurginae	Chile: Region II. 4 km N Domeyko	DQ884569	DQ884477	DQ884710	DQ872756
Calliopsis (Nomadopsis) fracta (Rozen, 1952) [Cafr515]	Andrenidae: Panurginae	USA: CA. Santa Cruz Co., San Antonio junction. 28.v.1999	AY585101	AF344587	—	AY654476
Panurgus (Panurgus) calcaratus (Scopoli, 1763) [Pnca514]	Andrenidae: Panurginae	Italy: Rome. 07.vi.1998	AY585105	AF344612	—	AY654479
Melitturga (Melitturga) clavicornis (Lattreille, 1806) [Mtcl959]	Andrenidae: Panurginae	France: Herault. Causse de la Selle. 17.vi.2002	AY585104	DQ116703	—	AY654478
Agapostemon tyleri (Cockerell, 1917) [Agty230]	Halictidae: Halictinae	USA: AZ. Cochise Co., Portal	AF140320	AY227940	AY222577	AY654506
Augochlorella pomoniella (Cockerell, 1915) [Aupo592]	Halictidae: Halictinae	USA: CA. Inyo Co., Big Pine. 15.vi.1999	AF435373	AY227935	AY222572	AY654507
Mexalictus arizonensis Eickwort, 1978 [Mxaz98]	Halictidae: Halictinae	USA: AZ. Santa Cruz Co.	AF140322	AY227959	AY222595	—
Halictus (Halictus) rubicundus (Christ, 1791) [Haru32]	Halictidae: Halictinae	USA: MT. Missoula Co., Missoula	AF140335	DQ116674	AY222592	AY654510
Dieunomia (Epinomia) nevadensis (Cresson, 1874) [None207]	Halictidae: Nomiinae	USA: AZ. Cochise Co., 1 mi E. Apache, 22.ix.1999	AF435396	AY227931	AY222568	AY654512
Conanthalictus wilmattae Cockerell, 1936 [Cowi351]	Halictidae: Rophitinae	USA: CA. Riverside Co., 10 mi S Palm Desert. 15.iii.1997	AF435378	AY227934	AY222553	AY654511
Dufourea mulleri (Cockerell, 1898) [Dumu233]	Halictidae: Rophitinae	USA: Michigan	AF435383	AY227918	AY222555	AY654509
Penapis penai Michener, 1965 [Pnpe572]	Halictidae: Rophitinae	Chile: Region II. N Vallenar	AF435401	AY227921	AY222558	AY654513
Rophites (Rophites) algirus Perez, 1895 [Roal968]	Halictidae: Rophitinae	France: Var. 5 km S Entrecasteaux. 14.vi.2002	AY585144	DQ116675	—	AY654515
Systropha curvicornis (Scopoli, 1770) [Sycu350]	Halictidae: Rophitinae	Austria: Vienna	AF435411	AY227925	AY222562	AY654516
Ctenocolletes nigricans Houston, 1985 [EA0120]	Stenotritidae	Australia: WA; ~18.4 km NNE Eurardy Station. 05.x.2005	DQ884570	DQ884478	DQ884711	DQ768459
Ctenocolletes rufescens Houston, 1983 [EA0121]	Stenotritidae	Australia: WA. Boorabbin Ntl.Prk. 25.ix.2005	DQ884571	DQ884479	DQ884712	DQ768460
Ctenocolletes smaragdinus (Smith, 1868) [EA0122]	Stenotritidae	Australia: WA. Boorabbin Ntl.Prk. 25.ix.2005	DQ884572	DQ884480	DQ884713	DQ768461
Stenotritus sp. [Stsp1015]	Stenotritidae	Australia: WA. 23 km SW Coorow. 17.xi.1997	DQ141115	DQ115564	DQ884714	DQ872757

nator chemistry and AmpliTaq-FS DNA Polymerase at Cornell University Life Sciences Core Laboratories Center.

GenBank Accession Nos. (Table 1) starting with "DQ7", "DQ8", and "EF0" are novel sequences. The remaining sequences were generated by B.N. Danforth and co-workers (Danforth et al., 2004, 2006a,b). The *wingless* sequence of *Halictus quadricinctus* (Fabricius) was used to represent that of *Halictus rubicundus* (Christ).

2.5. Alignment

Alignments for the individual gene data matrices were generated using similarity calculated at the nucleotide level ("−n") with DIALIGN 2.2 (Morgenstern, 1999) and with the ClustalW (Lasergene DNA Star software package). The resulting alignments were then corrected manually for obvious alignment errors using MacClade v 4.08 OSX (Maddison and Maddison, 2005) and WinClada 1.00.08 (Nixon, 2002). Regions of 28S rRNA and introns of EF-1α where alignment was overly ambiguous were excluded from the phylogenetic analyses. Despite losing some information, the removal of problematic regions of the alignments potentially increases the actual phylogenetic signal (see discussion by Talavera and Castresana, 2007). For the protein-coding genes, honey-bee (*Apis mellifera* Linnaeus) sequences were used to establish reading frames and intron/exon boundaries. Introns of the three protein-

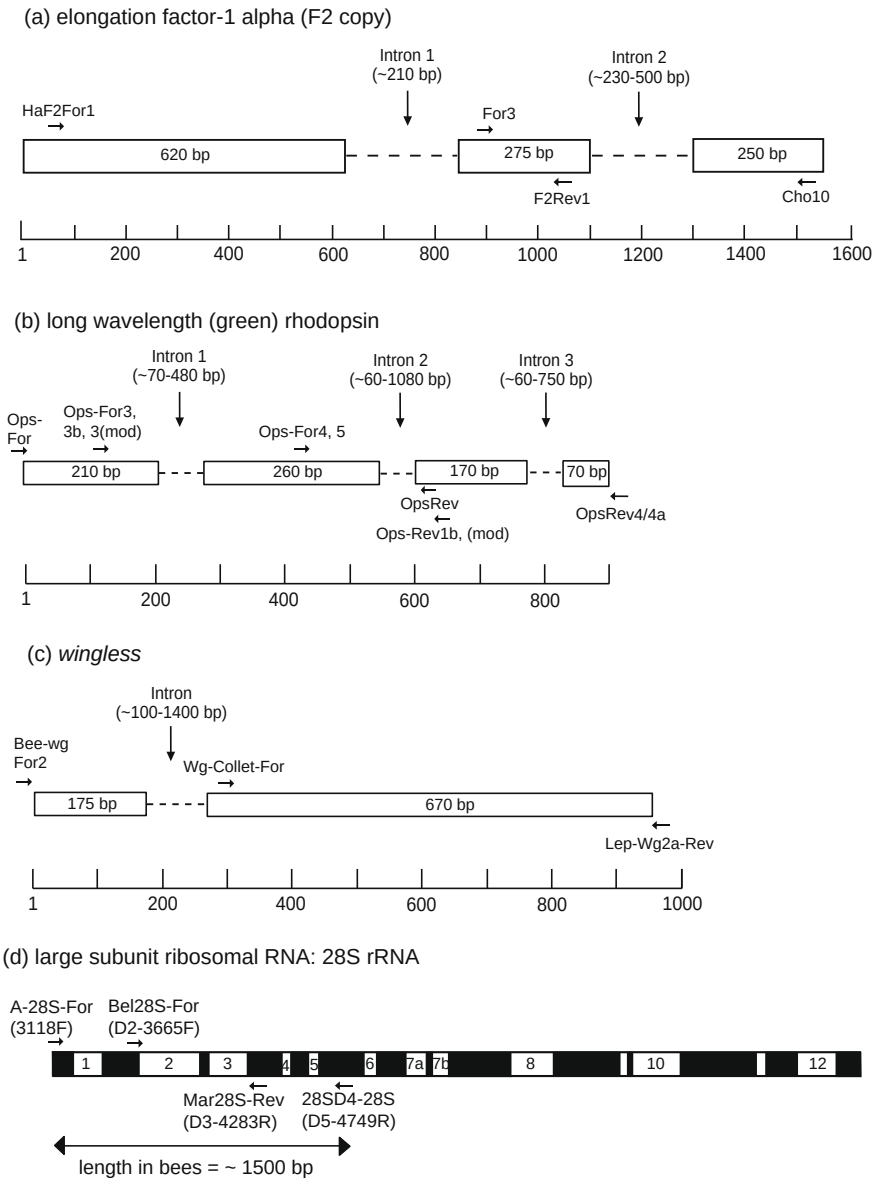

Fig. 2. Maps of the gene loci sampled for this study. Primer sites (see Table 2 for a complete list of primer sequences), introns, and exons are indicated. Introns, when variable in length, are represented on their lower limit of the range; length variation reported in the figure is observed in the sample of colletid bees included in this study; (a) elongation factor-1α, F2 copy (adapted from Danforth et al., 2004), numbering based on *Apis mellifera* coding sequence (Walldorf and Hoverman, 1990, GenBank Accession AF015267); (b) long wavelength rhodopsin (adapted from Danforth et al., 2004), numbering based on *A. mellifera* coding sequence (Chang et al., 1996; GenBank Accession U26026); (c) *wingless*, numbering based on *Drosophila melanogaster* coding sequence (Uzvölgyi et al., 1988; GenBank Accession J03650); (d) 28S rRNA (adapted from Caterino et al., 2000), nucleotide numbering and region names based on *D. melanogaster* ribosomal genes (Hancock et al., 1988; Linares et al., 1991; Tautz et al., 1988, GenBank Accession M21017).

coding genes sampled were initially aligned for all taxa. Introns of opsin and *wingless* were subsequently excluded from the analyses because length variation and sequence variability observed in these regions made them unalignable. Both EF-1α introns could only be aligned for colletid and stenotritid sequences (i.e., introns from all other outgroup terminals were discarded to allow for the remaining to be alignable). The second EF-1α intron is more variable than intron 1 (Fig. 2), so was aligned separately for two sets of species: (1) representatives of Stenotritidae, Diphaglossinae, and Paracolletinae (excluding *Callomelitta antipodes*); and (2) of Colletinae s.str., Euryglossinae, Hylaeinae, Scrapterinae, Xeromelissinae, plus *Callomelitta antipodes*. Selection of these blocks of taxa was made based on groupings found in preliminary analyses of exon data plus the 28S sequences, and on sequence similarity found among introns of the terminals. Although unusual, this procedure will not introduce phylogenetic artifacts. We are simply attempting to extract the maximum amount of information from the largest regions of the non-coding partitions of the data set.

In cases where multiple sequences were available for the same species, the sequences were merged after being resolved in the same clade in preliminary analyses and resulting partial polymorphisms were kept as such. Individual gene data sets were concatenated with WinClada.

The combined data set in Nexus-format used in phylogenetic analyses is available for download as Supplementary information.

2.6. Phylogenetic analyses

Bayesian phylogenetic inference was used to estimate the tree topology. Metropolis-coupled MCMC, as implemented in the serial version of MrBayes 3.1.2 (Altekar et al., 2004; Huelsenbeck and

Table 2
Primer sequences for EF-1α (F2 copy), opsin, *wingless*, and 28S rRNA used for PCR assays of bees.

Locus	Primer	Sequence	Position	Reference
EF-1α[a]	HaF2For1	5′-GGG YAA AGG WTC CTT CAA RTA TGC-3′	511	Danforth et al. (1999)
	For3rho	5′-GGY GAC AAY GTT GTT TTY AAY G-3′	1496	Danforth et al. (1999)
	F2-rev1	5′-A ATC AGC AGC ACC TTT AGG TGG-3′	1600	Danforth et al. (1999)
	Cho10-Rev(mod)	5′-AC RGC VAC KGT YTG HCK CAT GTC-3′	1887	Danforth et al. (1999)
Opsin[b]	Opsin-For [=LWRhFor)]	5′-AAT TGC TAT TAY GAR ACN TGG GT-3′	398	Mardulyn and Cameron (1999)
	Opsin-For3 [=LWRhFor3]	5′-AGA TAC AAC GTR ATC GTS AAR GGT-3′	512	Danforth et al. (2004)
	Opsin-For3(mod)	5′-TTC GAY AGA TAC AAC GTR ATC GTN AAR GG-3′	506	Danforth (unpublished)
	Opsin-For3b	5′-AGA TAC AAC GTR ATY GTN AAR GGT-3′	512	Almeida (unpublished)
	Opsin-For4	5′-GAG AAR AAY ATG CGB GAR CAA GC-3′	803	Danforth et al. (2004)
	Opsin-For5	5′-ATG CGN GAR CAR GCN AAR AAR ATG AA-3′	812	Danforth (unpublished)
	Opsin-Rev (=LWRhRev)	5′-ATA TGG AGT CCA NGC CAT RAA CCA-3′	946	Mardulyn and Cameron (1999)
	Opsin-Rev1b	5′-RTA YGG RGT CCA NGC CAT RAA CCA-3′	946	Almeida (unpublished)
	Opsin-Rev(mod)	5′-ATA NGG NGT CCA NGC CAT GAA CCA-3′	946	Danforth (unpublished)
	Opsin-Rev4	5′-GGT GGT GGT RCC GGA RAC GGT G-3′	1147	Danforth et al. (2004)
	Opsin-Rev4b	5′-GGT RCC GGA RAC GGT GGA DGT NGC RTC-3′	1147	Danforth (unpublished)
Wingless[c]	Bee-*wg*-For2	5′-GGC AGC ATY CAG TCS TGY TCC TGC GA-3′	445	Sipes (unpublished)
	Wg-Collet-For	5′-CAC GTG TCB TCB GRG ATG MGR SAG GA-3′	670	Almeida (unpublished)
	Lep-*Wg*2a-Rev	5′-ACT ICG CAR CAC CAR TGG AAT GTR CA-3′	1336	Brower and DeSalle (1998)
28S rRNA[d]	A-28S-For	5′-CCC CCT GAA TTT AAG CAT AT-3′	3318	Ward and Brady (2003)
	Bel28S-For (D2-3665F)	5′-AGA GAG AGT TCA AGA GTA CG TG-3′	3665	Belshaw and Quicke (1997)
	Mar28S-Rev (D3-4283R)	5′-TAG TTC ACC ATC TTT CGG GTC CC-3′	4283	Mardulyn and Whitfield (1999)
	28SD4-Rev (D5-4749R)	5′-GTT ACA CAC TCC TTA GCG GA-3′	4749	Danforth et al. (2006a)

[a] PCR conditions. HaF2For1/F2-rev1: 94 °C for 1 min, 48–52 °C for 1 min, 72 °C for 1.5 min (35 cycles); For3rho/Cho10-Rev(mod): 94 °C for1 min, 54–56 °C for 1 min, 72 °C for 1 min (35 cycles). Positions based on the 5′ end of the primer in *Apis mellifera* (Walldorf and Hoverman, 1990; GenBank Accession No. AF015267).

[b] PCR conditions. Opsin-For/Opsin-Rev: 94 °C for 1 min, 50–54 °C for 1 min, 72 °C for 1.5 min (35 cycles); Opsin-For3/Opsin-Rev: 94 °C for 1 min, 52–54 °C for 1 min, 72 °C for 1 min (35 cycles); Opsin-For3(mod)/Opsin-Rev(mod): 94 °C for 1 min, 52–54 °C for 1 min, 72 °C for 1 min (35 cycles); Opsin-For3 b/Opsin-Rev1 b: 94 °C for 1 min, 50–54 °C for 1 min, 72 °C for 1 min (35 cycles); Opsin-For4/Opsin-Rev4: 94 °C for 1 min, 55–58 °C for 1 min, 72 °C for 1 min (35 cycles); Opsin-For5/Opsin Rev4b: 94 °C for 1 min, 55 °C for 1 min, 72 °C for 1 min (35 cycles). Positions based on the 5′ end of the primer in *Apis mellifera* (Chang et al., 1996; GenBank Accession No. U26026).

[c] PCR conditions. Bee-*wg*-For2/Lep-*wg*2a-Rev: 94 °C for 1 min, 54–58 °C for 1 min, 72 °C for 1.5 min (35 cycles); *Wg*-Collet-For/Lep-*wg*2a-Rev: 94 °C for 1 min, 56 °C for 1 min, 72 °C for 1 min (35 cycles). Positions based on the 5′ end of the primer in *Drosophila melanogaster* (Uzvölgyi et al., 1988; GenBank Accession No. J03650).

[d] PCR conditions. A-28S-For Mar28S-Rev: 94 °C for 1 min, 58 °C for 1 min, 72 °C for 1.5 min (35 cycles); Bel28S-For/28SD4-Rev: 94 °C for 1 min, 58 °C for 1 min, 72 °C for 1.5 min (35 cycles). Positions based on the 5′ end of the primer in *Drosophila melanogaster* (Tautz et al., 1988; Linares et al., 1991; GenBank Accession No. M21017). Numbering corresponds to the expansion regions as proposed by Hancock et al. (1988).

Ronquist, 2005), was used to estimate the posterior probability distribution. The gamma distribution of rate variation across sites was approximated by a discrete distribution with four rate categories, each category being represented by its mean rate. All chains, including coupled chains in the same run, were started from different, randomly chosen trees. Searches were run for 2×10^6 generations on two sets of 10 chains each, through the Computational Biology Service Unit at the Cornell Theory Center (http://cbsu.tc.cornell.edu/). Convergence was assessed by the standard deviation of split frequencies of the two independent MrBayes runs, by the convergence diagnostic for individual parameters employing potential scale reduction factor (Gelman and Rubin, 1992, uncorrected), and by the achievement of stationarity of the log likelihood values of the cold chain. Trees were saved every 100 generations. The initial 1000–3000 trees were discarded after examining the variation in log likelihood scores over time. Models for the five partitions of the concatenated data set (EF-1α [exons], EF-1α [introns], 28S rRNA, opsin [exons], and *wg* [exons]) were individually selected and implemented. Partitioned analyses were run with MrBayes applying models selected by different methods (see below), in order to test the effect of various levels of model complexity.

The best-fit model of evolution for each of the five partitions was statistically tested. Model selection is one of the most controversial fields of statistics (Burnham and Anderson, 2002), and this is not different in the field of phylogenetics (e.g., Posada and Crandall, 2001; Nylander et al., 2004; Posada and Buckley, 2004; Sullivan and Joyce, 2005). The four approaches most commonly used for model selection in phylogenetics are hierarchical likelihood ratio test (hLRT), Akaike information content (AIC), Bayesian information content (BIC), and decision theory (DT) (Posada and Buckley, 2004; Sullivan and Joyce, 2005). Four programs were used to shed light on the best-fit model(s) for the data: (1) DT-ModSel (Minin et al., 2003—DT model selection); (2) ModelTest 3.7 (Posada and Crandall, 1998—hLRT, AIC, BIC; α set to 0.05); (3) MrModelTest (Nylander, 2004b—hLRT, AIC); and (4) MrAIC.pl 2.2 (Nylander, 2004a—AIC and BIC). In addition to statistical tests for model selection, highly complex models were tested in MrBayes for the combined data matrix, to check the effects of higher model-realism in the phylogenetic results: GTR + I + Γ and GTR + SSR for each partition, with all parameters unlinked across partitions.

For comparison with the Bayesian analysis, the data set was subjected to maximum likelihood analysis and equal weights parsimony analyses. Maximum likelihood searches were done using Garli v0.951 (Zwickl, 2006). Garli does not allow for data set partitioning, so the complete concatenated matrix was analyzed under a GTR + I + Γ model as a means to deal with the heterogeneity within the data. When analyzing individual genes, simpler models were used when that appeared to be a reasonable alternative, but various levels of complexity were tested. The termination condition was set to 1×10^5 generations without significantly better likelihood scoring topology being found ("*genthreshfortopoterm = 100,000*"); remaining settings were left unchanged from the defaults. Non-parametric bootstrap proportions were calculated based on 1000 pseudo-replicates generated with Garli. The frequency of occurrence of each group present in the most likely tree topology was calculated using WinClada.

Parsimony analyses were conducted in PAUP* v4.0b10 (Swofford, 2002) using heuristic searches with tree bisection–reconnection (TBR), 1000 random-taxon-addition replicates holding 50 trees per replicate, and treating sequence indels as missing data. Branch support was assessed with 1000 bootstrap pseudo-replicates (Felsenstein, 1985). Each resampled matrix was searched

100 times and consensus trees found at each iteration were saved and used to calculate node support (percentage count) for each clade present in the strict consensus of the most parsimonious trees using WinClada.

2.7. Rate of substitution and parameters of molecular evolution

MrBayes was also used to explore some properties of the molecular data sets. Summarized parameters after discarding the initial trees were used as estimates of base frequencies, rates of substitution, and shape parameter (α) of the gamma distribution. Data were analyzed with GTR + Γ models in which all parameters were unlinked across partitions. This allowed base composition of each gene and within partition rate variation (α) to be estimated. In order to estimate the rates of substitution, the data set was analyzed with a site-specific rates model (GTR + SSR) with rate categories corresponding to each codon position within the protein-coding gene exons, to the intron of EF-1α, and to 28S rRNA.

2.8. Topological congruence between partition trees and the combined tree

The phylogenetic signal of each partition was assessed by comparing the Bayesian topology obtained for each gene analyzed along with the topology obtained based on an analysis of all five partitions combined. Nodes present in the tree topology within the ingroup resulting from the five partitioned analyses were compared in a pair-wise manner to the combined tree and classified as either congruent or not. A congruent node was one underpinning a monophyletic group of species found in the analysis of the combined data set or not contradicted by the results of the latter. 'Congruent topological information' (CTI) was a measure used here to approximate concordance between the partitions and the combined tree topologies. CTI was calculated as the ratio between number of congruent nodes and maximum number of resolved nodes for a given number of terminals.

2.9. Hypothesis testing

In order to compare alternative tree topologies, Kishino–Hasegawa and Shimodara–Hasegawa tests (Kishino and Hasegawa, 1989; Shimodara and Hasegawa, 1999) were performed with 1000 bootstrap replicates using PAUP*. Likelihood was estimated with the RELL (resampling of estimated log-likelihood) method (Kishino et al., 1990), as an approximation of a computationally intensive bootstrap (Felsenstein, 2004). Significance of likelihood score differences was assessed in a pair-wise manner using a one-tailed test, as suggested by Felsenstein (2004, 369). Additionally, a Bayesian approach to hypothesis testing was used for comparisons among alternative topologies. Constrained tree topologies were generated in MrBayes under a fixed model and their harmonic means were compared (Kass and Raftery, 1995; Nylander et al., 2004). From the harmonic means, $2\log_e(B_{10})$ is calculated by doubling the difference in the means of two topologies being compared and this quantity can be interpreted with the aid of the table from Kass and Raftery (1995, 277; a useful discussion about the application of Bayes factors specifically to phylogenetic questions is given by Nylander et al., 2004).

3. Results

3.1. Molecular data sets

The final combined data matrix contains 144 taxa and 5498 aligned base pairs. Among those, 1903 are parsimony-informative characters. Information of a character is a quantity defined as its maximum number of steps minus its minimum number of steps (denominator for the retention index [Farris, 1989]). The matrix's information is simply the sum of the information of each of its component characters. The information of the combined matrix was equal to 27,358 steps (calculated with WinClada). The resulting concatenated data matrix includes the exons of the three protein-coding genes (EF-1α, opsin, and *wg*), both introns of EF-1α, and most of the sequences of 28S rRNA. After the 28S sequences were aligned, regions where alignment was judged to be ambiguous were deactivated. The total of excluded 28S aligned characters was 371 steps, representing an information loss of 1928 steps. The total number of characters and of informative characters per gene, and the information content of each partition are presented in Table 3.

3.2. Model selection and comparisons among genes

A summary of the results found with the different model selection programs is presented in Table 4. The most complex model tested by those programs, GTR + I + Γ was indicated as the most appropriate one for each locus by at least one model selection strategy. Except for *wg*, all other genes had a model indicated to be superior by some of the model selection methods, which was simpler than GTR + I + Γ and implemented in MrBayes. The combined and unpartitioned data set was tested for one overall best-fit model and it was invariably GTR + I + Γ. MrBayes allows model partitioning and this is a desirable feature for allowing for a more realistic phylogenetic approach (Nylander et al., 2004). Empirical tests of evaluating the effects of using models with variable levels of complexity to the Bayesian phylogenetic analyses (partitioned and combined) did not yield significantly different results; topologies found for the data sets tended to be largely resilient to variation in model complexity, with slight changes being observed in the posterior probability values. All phylogenetic results shown are based on the simplest models considered adequate for the data by the model selection strategies tested.

The result of a characterization of the gene data is presented in Fig. 3. *Wingless* and EF-1α (both exons and introns) present a deviation from equal base composition more noticeable than in the other two genes—the former has a slight G–C bias and the latter an A–T bias (Fig. 3a). This is reflected in the models found to be the best-fit for each gene (Table 4). The simplest models shown to be adequate for opsin and 28S rRNA have equal base composition (K80 and SYM, respectively), whereas *wg* and EF-1α require more complex models (Table 4).

Relative rates of substitution of each of the gene loci are shown in Fig. 3b and an approximation of how variable these rates are within each locus is given by the α parameter of the gamma distribution (Fig. 3c). High values of α (e.g., EF-1α introns) indicate relatively homogeneous rates across sites, whereas low values of α (e.g., *wingless* exons) indicate high heterogeneity in among-site rate variation (Fig. 3c). Introns would be expected to show homogenous rates relative to protein-coding exons because there se-

Table 3
Overview of the partition and combined data sets.

	Number of characters	Informative characters	Total information	Information/number of characters
EF-1α (exons	1151	427	7966	6.92
EF-1α (introns)	1425	601	6174	4.33
Opsin (exons)	707	324	5824	8.24
wg (exons)	692	218	3862	5.58
28S rRNA	1523	333	3352	2.32
Combined	*5498*	*1903*	*27358*	*4.98*

Table 4
Models for each gene as selected by different computer programs. Models in bold and italicized are implemented in MrBayes 3.1.2.

Gene locus	Program: model selection test	Model [number of free parameters][*]
28S rRNA	DT_ModSel; modeltest: BIC	TVMef+I+Γ [4]
	MrAIC: AIC, AICc, BIC	***SYM+I+Γ*** [5]
	modeltest: hLRTs	TrN+I+Γ [5]
	modeltest: AIC; MrModelTest: hLRTs, AIC	***GTR+I+Γ*** [8]
EF-1α (exons)	DT_ModSel; modeltest: BIC; MrAIC: AIC*, AICc*, BIC*	***HKY+I+Γ*** [4]
	modeltest: hLRTs, AIC	TrN+I+Γ [5]
	MrModelTest: hLRTs, AIC	***GTR+I+Γ*** [8]
EF-1α (introns)	DT_ModSel; modeltest: AIC, BIC	***HKY+I+Γ*** [4]
	modeltest: hLRTs	TVM+I+Γ [7]
	MrModelTest: hLRTs, AIC; MrAIC: AIC, AICc, BIC	***GTR+I+Γ*** [8]
Opsin (exons)	DT_ModSel; modeltest: AIC, BIC	***K80+I+Γ*** [1]
	modeltest: hLRTs	TrN+I+Γ [5]
	MrModelTest: hLRTs, AIC; MrAIC: AIC, AICc, BIC	***GTR+I+Γ*** [8]
wingless (exons)	DT_ModSel; modeltest: AIC, BIC	TrN+I+Γ [5]
	modeltest: hLRTs; MrModelTest: hLRTs, AIC; MrAIC: AIC*, AICc*, BIC*	***GTR+I+Γ*** [8]

[*] Number of free parameters for the model of substitution alone, this number does not include parameters associated with the gamma distribution (Γ) and the proportion of invariant sites (I).

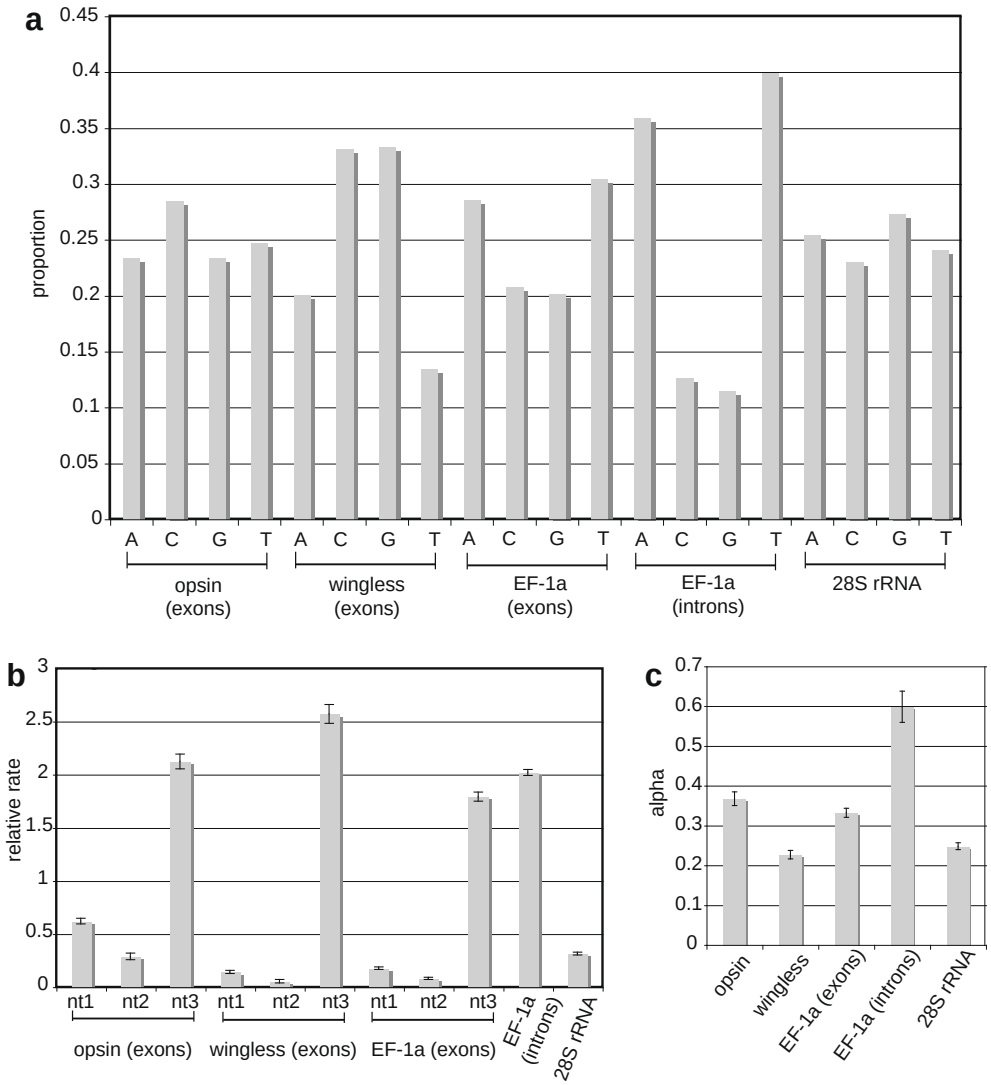

Fig. 3. Comparisons among gene loci sampled for this study based on molecular parameters estimated by MrBayes: (a) frequencies of the nucleotide bases for each gene locus; (b) relative rates of substitution; (c) α parameter of the gamma distribution of rates.

quences should be under less selection than coding regions. In relative terms, rates of substitution of first and second positions of opsin are higher and of third positions are lower than those observed in EF-1α and *wg* (Fig. 3b). The relative rate of the introns of EF-1α fits in the range of rates of third positions of the exons and that of 28S rRNA is similar to 1st positions of exons (Fig. 3b).

3.3. Phylogenetic relationships

Results of the Bayesian phylogenetic analysis are shown in Figs. 4 and 5. Overall, topologies obtained with maximum likelihood and parsimony analyses were very similar to those of the Bayesian result (Figs. S1–S3, Supplementary materials). Because of this topological congruence among the three analytical methods, the Bayesian tree is used to show node support values obtained using bootstrap proportions of parsimony and ML, as well as Bayesian posterior probabilities (Fig. 4). As expected, Bayesian posterior

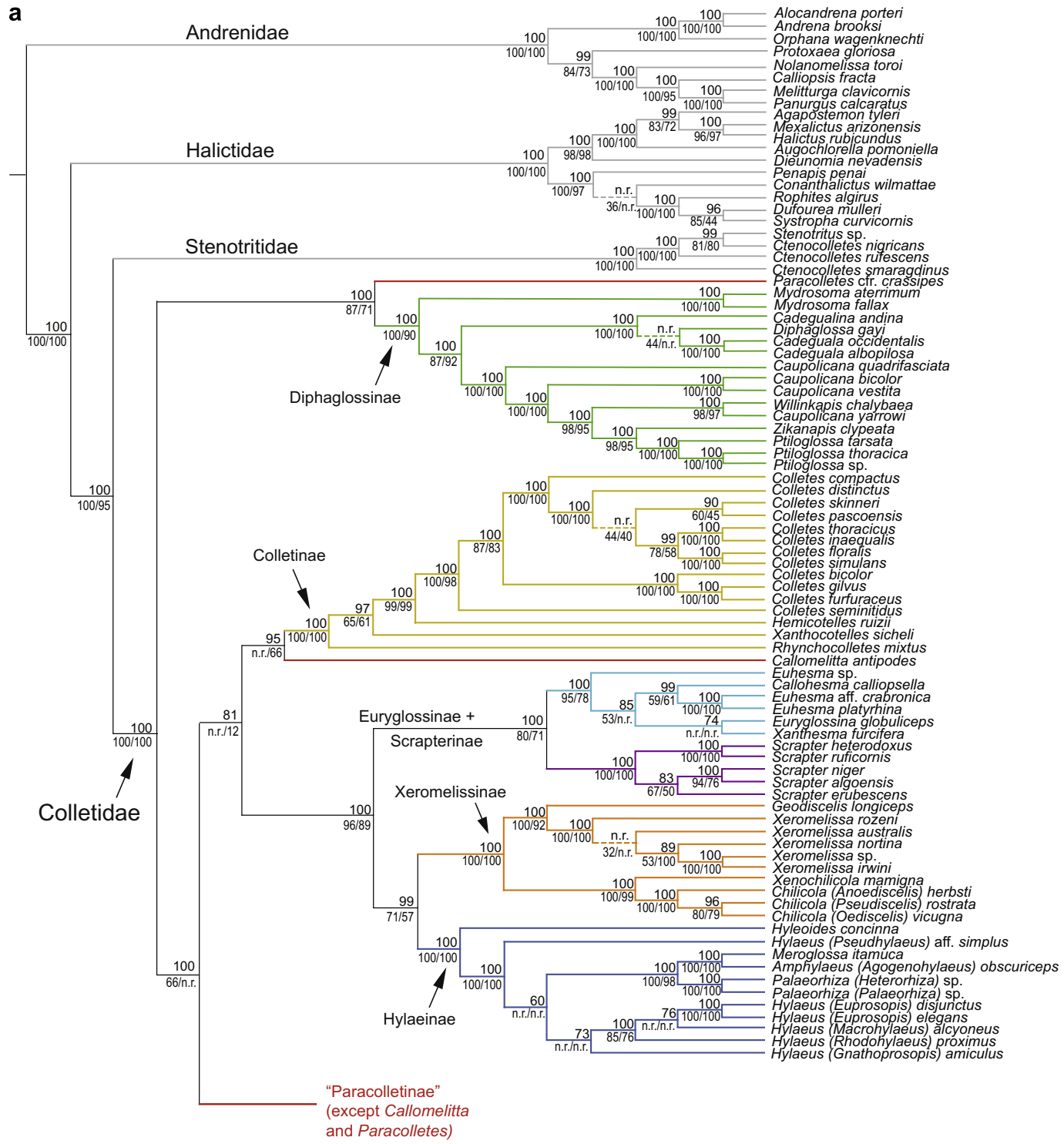

Fig. 4. Majority-rule consensus tree resulting from mixed-model Bayesian analysis of 5498 aligned bp of the exons and introns of EF-1α (model = HKY + I + Γ), exons of opsin (K80 + I + Γ), exons of *wingless* (GTR + I + Γ), and 28S rRNA (SYM + I + Γ) from 122 species of colletid bees and 22 species of other bee families. Branches represented with dashed lines were not recovered in the Bayesian consensus and are based on the maximum likelihood topology. Numbers above internodes indicate Bayesian posterior probabilities; numbers below internodes indicate non-parametric bootstrap proportions for the for the likelihood analysis (left) and parsimony analysis (right). Nodes not recovered in one of the three analyses are indicated by "n.r.". (a) Andrenidae, Halictidae, Stenotritidae, and most of the colletid subfamilies; (b) "Paracolletinae" (except *Callomelitta* and *Paracolletes*).

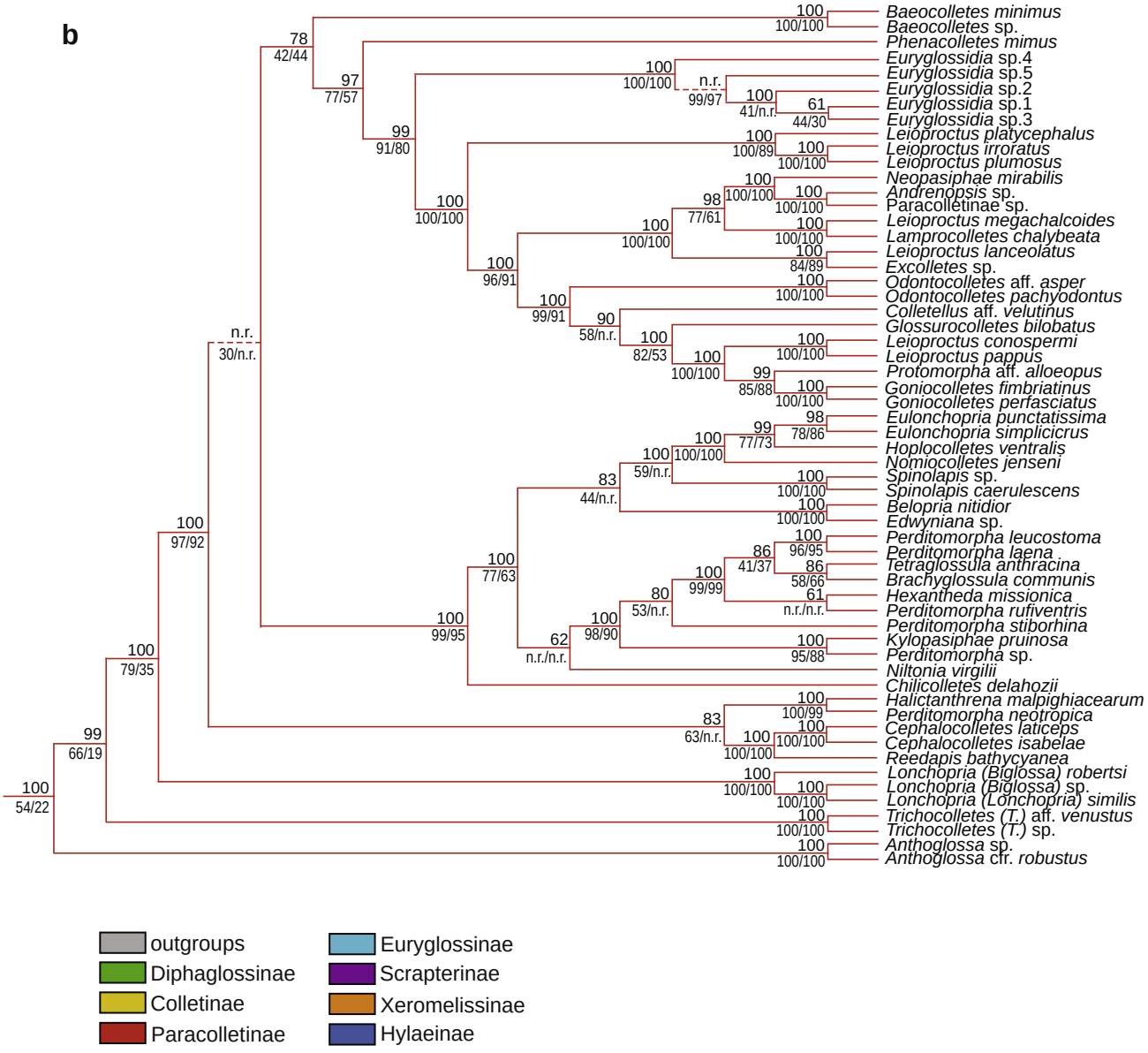

Fig. 4 *(continued)*

probabilities were higher or equal to bootstrap proportions found with ML and with parsimony.

The most unanticipated results were the placement of *Paracolletes* and *Callomelitta* as two lineages independent of the remaining Paracolletinae. To facilitate communication, the clade formed by all Paracolletinae except *Paracolletes* and *Callomelitta* will be referred to as "Paracolletinae" henceforth. For the most part, all three analytical methods found the same arrangement of subfamilies and tribes, with one noteworthy exception: ML placed *Callomelitta antipodes* as sister-group to "Paracolletinae", whereas Bayesian and parsimony analyses placed it as sister of Colletinae s.str.

Alternative phylogenetic arrangements regarding the placement of *Paracolletes* and *Callomelitta* were statistically tested. The topology favored by both parsimony and Bayesian analyses is shown in Fig. 6a; pair-wise comparisons of this tree to five alternatives (Fig. 6b–f) were done using Bayesian and likelihood-based statistics (Table 5). The only alternative topology considered to be significantly different by both statistical measures was the sister-group relationship (or inclusive relationship) of *Paracolletes* and "Paracolletinae" (Fig. 6c).

The strict consensus of the most parsimonious trees (Fig. S3) was less resolved than the Bayesian and ML topologies. Relationships between Colletinae plus *Callomelitta*, "Paracolletinae", and the clade formed by Euryglossinae, Hylaeinae, Scrapterinae, and Xeromelissinae form an unresolved trichotomy. The relationship between "Paracolletinae" and its sister-clade was strongly supported by the Bayesian analysis (100% posterior probability) but only moderately supported by ML (66% bootstrap support). All colletid subfamilies, except Paracolletinae, were very well-supported by the data as monophyletic. Furthermore, the grouping of Euryglossinae, Scrapterinae, Hylaeinae, and Xeromelissinae, as well as the sister-group relationships between [Euryglossinae + Scrapterinae], and [Hylaeinae + Xeromelissinae] all received fairly high bootstrap and Bayesian posterior probability endorsement. The levels of support for groupings within Colletidae based on the different gene partitions and the combined results can be compared in Table 6.

The sister-group relationship between Colletinae (with or without *Callomelitta*) and the clade formed by Euryglossinae, Hylaeinae, Scrapterinae, and Xeromelissinae is weakly supported (Figs. 4a, S1 and S3: 81% posterior probability; ML bootstrap = 50; parsimony bootstrap = 12). The Bayesian posterior probability value is lower than that for the majority of other clades, which are in general sup-

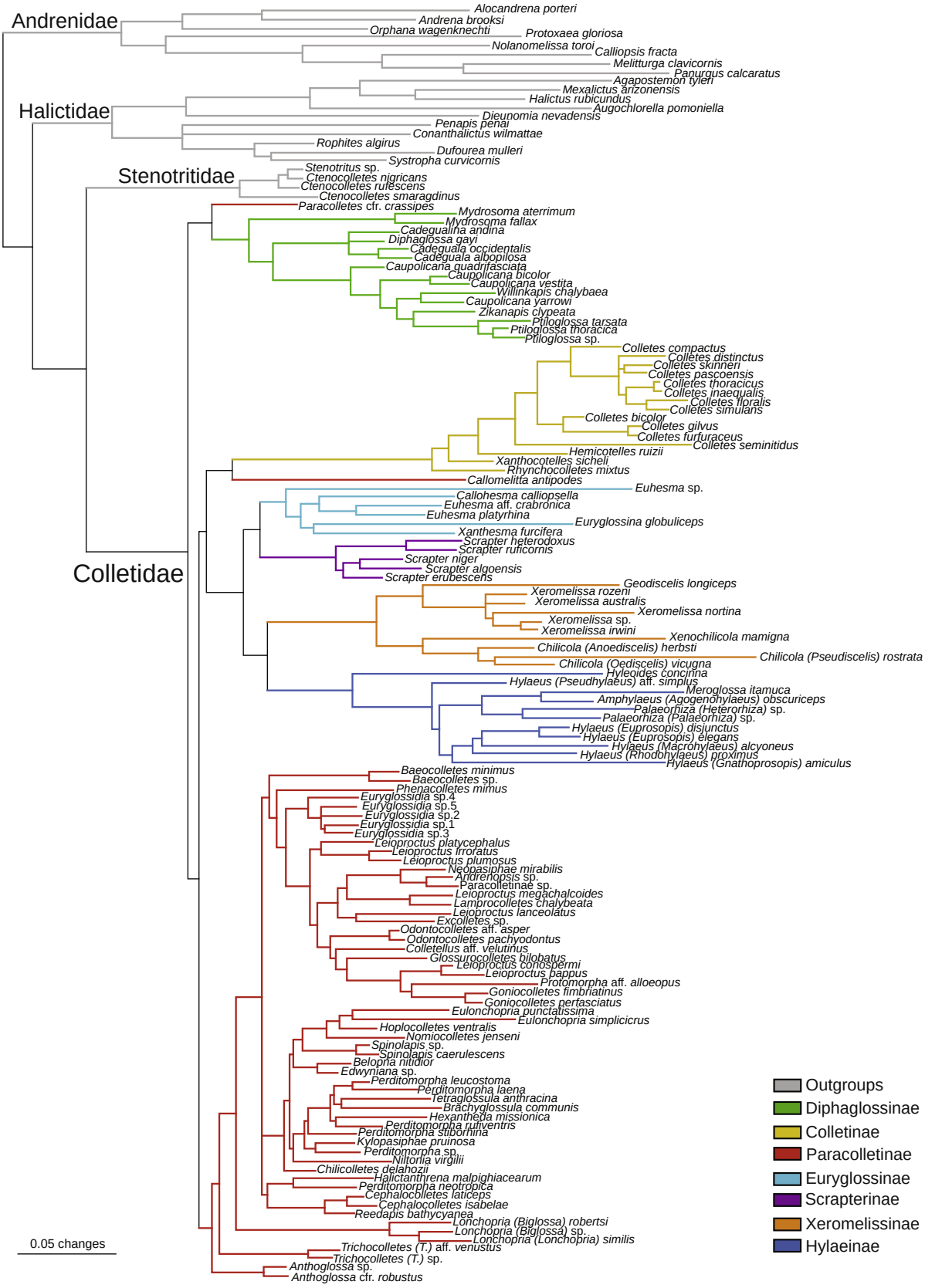

Fig. 5. Majority-rule consensus phylogram resulting from mixed-model Bayesian analysis of 5498 aligned bp of the EF-1α (exons and introns), opsin (exons), *wingless* (exons), and 28S rRNA genes from 122 species of colletid bees and 22 species of other bee families.

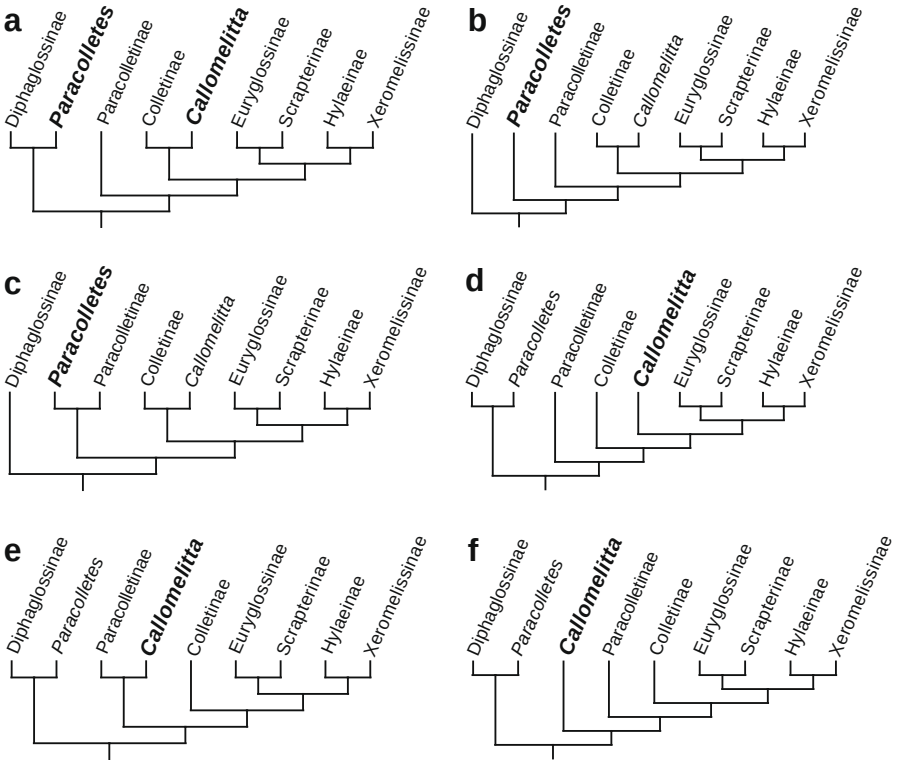

Fig. 6. Alternative tree topologies statistically compared using likelihood-based tests and Bayesian information content. The resulting topology of an unconstrained Bayesian analysis of the combined data set (a) was compared in a pair-wise manner to five alternative topologies (b–f) with various placements of *Callomelitta antipodes* and *Paracolletes* cfr. *crassipes*.

Table 5
Results of comparisons among alternative topologies for Colletidae using maximum likelihood (Kishino–Hasegawa and Shimodaira–Hasegawa tests) and Bayesian approaches. ML tests, as implemented in PAUP* using a one-tailed test of significance (*p*-value for both tests was identical in all pair-wise comparisons made). Harmonic means from MrBayes analyses and Bayes factor comparisons: $2\log_e(B_{10})$ stands for twice the log of the Bayes factor in the comparison between two topologies. All comparisons were made between the unconstrained tree topology and the alternative constrained topologies.

Tree topology	$-\ln L$	Diff. $-\ln L$	*p*-Value	Harmonic mean	$2\log_e(B_{10})$
Unconstrained tree topology (Fig. 7a)	85524.62207			−78986.35	
Paracolletes constraint 1 (Fig. 7b)	85531.55656	6.94	0.258	−79007.35	42.00***
Paracolletes constraint 2 (Fig. 7c)	85558.56353	33.93	0.020*	−79023.13	73.56***
Callomelitta constraint 1 (Fig. 7d)	85527.1818	2.56	0.264	−78990.34	7.98**
Callomelitta constraint 2 (Fig. 7e)	85521.8022	2.82	0.335	−78990.34	7.98**
Callomelitta constraint 3 (Fig. 7f)	85524.10473	0.52	0.481	−78992.79	12.88***

* Statistically significant difference of likelihood scores at $p = 0.05$.
** *Strong* evidence against null hypothesis, i.e., tree topologies equally likely—as proposed by Kass and Raftery (1995, 277).
*** *Very strong* evidence against null hypothesis (Kass and Raftery *ibid*).

Table 6
Support for the monophyly of Colletidae, colletid subfamilies, and clades formed by Euryglossinae, Hylaeinae, Scrapterinae, and Xeromelissinae, assessed using Bayesian posterior probabilities of individual gene loci ("n.r." is used to indicate clades not recovered by individual loci).

	EF-1α, exons	EF-1α introns	Opsin	*Wingless*	28S rRNA	Combined
Colletidae	100	100	100	93	100	100
Colletinae s.str.	100	100	100	97	100	100
Diphaglossinae	85	93	100	82	63	100
Euryglossinae	98	77	96	n.r.	n.r.	100
Hylaeinae	100	100	100	100	n.r.	100
"Paracolletinae"*	n.r.	n.r.	78	n.r.	n.r.	100
Scrapterinae	100	100	100	98	98	100
Xeromelissinae	100	100	100	100	n.r.	100
Euryglossinae + Scrapterinae	74	57	99	n.r.	n.r.	100
Hylaeinae + Xeromelissinae	93	89	n.r.	n.r.	n.r.	99
Eurygl. + Scrapt. + Hylaei. + Xerom.	80	n.r.	100	n.r.	n.r.	100

* Clade formed by all representatives of Paracolletinae except *Callomelita antipodes* and *Paracolletes* cfr. *crassipes*.

ported by values equal or higher than 90%. The three basal-most nodes of "Paracolletinae" in the Bayesian and ML phylograms reveals very short internodes (Figs. 5 and S2). These groupings are weakly supported by parsimony and ML bootstrap proportions, but are well-supported by Bayesian posterior probabilities. Neither ML nor parsimony analyses yield alternative arrangements of the taxa shown as basal in the "Paracolletinae" clade.

The majority of the Paracolletinae were shown to form a well-supported clade, which comprises most groups traditionally comprised in this subfamily, except *Paracolletes*, *Callomelitta*, *Anthoglossa*, *Trichocolletes*, and *Lonchopria*. This clade roughly corresponds to *Leioproctus* s.l. (e.g., Michener, 1965, 1989, 2007), although a number of paracolletine genera recognized by Michener (2007) render *Leioproctus* s.l. paraphyletic. Three main lineages recovered within this clade: (1) an entirely Australian clade which comprises *Baeocolletes*, *Goniocolletes* and relatives; (2) a Neotropical clade that includes *Eulonchopria*, *Niltonia* and related genera; and (3) another Neotropical group formed by *Halictanthrena*, *Reedapis* and related genera. Relationships among these three groups are not firmly established by any of the analytical methods employed for phylogenetic estimation. The internodes on this part of the tree are very short (Figs. 5 and S2), making the recovery of unambiguous phylogenetic information difficult. The above mentioned clade formed by Australian paracolletine bees contains a well-supported subclade which corresponds to *Leioproctus* s.str., *Goniocolletes* and related genera. This unnamed clade not only contains enough sequence synapomorphies to render high support values (Fig. 4b), but also has exceptionally long introns in the non-coding regions of EF-1α (intron 2) and opsin (intron 3) (Danforth and Almeida, unpublished data).

Five partition trees, resulting from Bayesian analyses of the individual partition data sets (introns and exons of EF-1α, exons of opsin and *wg*, and 28S rRNA), are shown in Figs. S4–S8 (Supplementary materials). All five partitions, independently (Table 6), or combined recovered colletid monophyly. The relationships between Colletidae and outgroups agree with the molecular results obtained by Danforth et al. (2006a,b). Moreover, relationships within Andrenidae and within Halictidae fully agree with more detailed studies of these two bee families conducted previously (Ascher, 2004; Danforth et al., 2004, respectively).

Additional investigation of the resolution power provided by each gene showed a different (and more fundamental) role of EF-1α and opsin compared to *wg* and 28S rRNA in the phylogenetic analyses (Fig. 7). The former two were particularly effective in resolving the earliest divergences among colletid lineages (Fig. 7: "deep nodes"). Opsin and EF-1α contained information to resolve more recent divergences as well, as measured by the congruence between paracolletine nodes in each of the gene partition trees and the combined tree (Fig. 7: "Paracolletinae"). These two genes are highly informative as compared to the other data partitions (Table 3), and high information content appears to translate into phylogenetic signal needed to resolve different regions of the colletid tree topology.

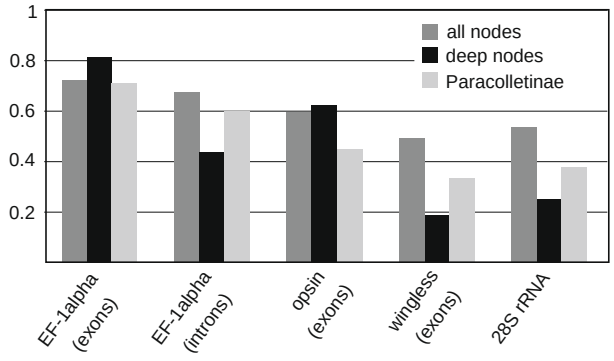

Fig. 7. Congruent topological information (CTI) of partition Bayesian trees in comparison to the combined Bayesian tree topology (see Section 2 for CTI calculation). CTI was computed for (1) the complete set of possible relationships within Colletidae ("all nodes"); (2) 17 deeper nodes within the ingroup (list of nodes in Supplementary Appendix 2); and (3) nodes within "Paracolletinae", to infer on the capacity of the individual loci to recover more recent relationships.

4. Discussion

Stable classifications are ultimately dependent on establishing a well-corroborated phylogenetic hypothesis for groups of organisms. For many taxonomically difficult groups of bees, much of the current classifications may be unsatisfying primarily because it is often not based on such a robust phylogenetic hypothesis. Molecular data can substantially clarify difficult and confusing phylogenetic matters. Colletidae and, specifically, Colletinae s.l. are examples of this. In Michener (2007, 136) words: "[i]t is clear that the Colletinae includes diverse elements. A needed step is a phylogenetic study of the forms here placed in Colletinae".

The continued addition of ideas and information to the discussion of relationships among lineages of Colletidae seems to have progressed in ways congruent with the results found by the current study. Rozen (1984) conjectured that Diphaglossinae should constitute the sister-group to all other colletid bees. Rozen's hypothesis was based on diphaglossine bees being the only colletids known to spin a cocoon, and the interpretation that this behavior (and morphology associated with it) could hardly have re-evolved from non-cocoon-spinning ancestors (Rozen, 1984). The sister-group relationship between Diphaglossinae and the remaining colletid subfamilies is strongly supported by the molecular phylogenetic results.

4.1. Molecular data and the phylogeny of Colletidae

The degree of phylogenetic resolution and the relatively strong support at many nodes obtained with the various methods tested in this study are very encouraging. This is the first phylogenetic study of Colletidae with extensive taxon sampling to provide a well-resolved hypothesis for the relationships on a worldwide scale. It is noticeable, however, that some regions of the trees are characterized by low support values. Future research may demonstrate that these weakly supported nodes are inherently hard to resolve, instead of having been caused by insufficient data. Parametric branch lengths (Figs. 5 and S2) reveal very short internodes at various parts of the tree of Colletidae, which suggest rapid diversification among the early-diverging colletid lineages. Considering that a relatively large amount of data was used and that those data came from genes appropriate for resolving the relationships among colletid subfamilies, low support may indeed derive from a rapid radiation (Whitfield and Kjer, 2008).

The improved phylogenetic resolution obtained in the combined analysis of all gene data partitions as compared with the individual partitioned analyses is readily appreciated when Figs. S4–S8 and Figs. 4, S1, and S3 are contrasted. In general, individual data partitions are expected to provide lower resolution than a simultaneous analysis. Moreover, phylogenetic noise present in individual partitions may obscure the true signal that is amplified when all data are combined (Nixon and Carpenter, 1996; Baker and DeSalle, 1997).

The phylogenetic signal present in EF-1α alone recovered the relationships among the major lineages of Halictidae (Danforth, 2002), which largely agreed with the morphological results of Pesenko (1999). Later, addition of genetic data from opsin and *wg* did

not alter the deeper relationships for an analysis of basically the same set of taxa, but resulted in increased support values for most internodes (Danforth et al., 2004). Phylogeny inferred from EF-1α was, among the genes sampled by Danforth et al. (2004), the most resolved tree. Likewise, EF-1α alone was sufficient to resolve relationships within Andrenidae (Ascher *in* Rozen, 2003, Fig. 37). This result is strengthened by the high degree of observed congruence with the morphological data set collected by Ascher (2004). Andrenidae comprise approximately 2650 species, a wide distribution and a morphological diversity comparable to Colletidae. Ascher's tree based on this one locus was well-resolved and relationships were also well-supported by bootstrap proportions and Bremer support values (Ascher, 2004, 43).

In our study too, EF-1α seem to provide most of the meaningful phylogenetic signal for the analysis of colletid relationships (Fig. 7 and Table 6), followed by opsin. The parameter α of the gamma distribution is a correlate of data set quality in empirical and simulation studies (Yang, 1998; Lin and Danforth, 2004). It is higher in the exons of EF-1α and opsin compared to those of *wg* and in 28S rRNA, reflecting less heterogeneity in among-site rate variation in EF-1α and opsin as compared to *wingless* and 28S (Fig. 3c).

Differences in the diverse measures of clade support assessment were expected given their differing natures. Parsimony relies on absolute number of changes, whereas maximum likelihood and Bayesian phylogenetics depend on the likelihood function. Even though phylogenetic estimates are expected to converge in certain situations (Steel and Penny, 2000; Swofford et al., 2001), it is also true that different phylogenetic methods may extract different information from the same data set. Overall, the topologies found by the three methods are almost identical, most of the difference being on support for certain clades.

Bootstrap and posterior probabilities are not easily comparable, despite the efforts of some researchers to find ways to link them (e.g., Wilcox et al., 2002; Alfaro et al., 2003; Huelsenbeck and Ranalla, 2004). The bootstrap has long been used as an assessment of support of clades in phylogenetic analyses; posterior probabilities are more recent and less well-understood (Huelsenbeck et al., 2002, 683–684 for a discussion). Huelsenbeck and Ranalla (2004) used simulations to demonstrate that Bayesian posterior probabilities can be a useful measure of phylogenetic reliability as long as the substitution model is not violated. Empirical studies found Bayesian posterior probabilities to be useful measures of support as well (e.g., Wilcox et al., 2002). Comparisons among various models of different levels of complexity in this study showed the topology to be very resilient to these changes. Posterior probabilities changed in some cases, usually varying by less than 10%. This serves as evidence that the Bayesian analysis of the current data set was not very sensitive to model misspecification and renders credible the posterior probability values obtained.

4.2. Monophyly of Colletidae

Colletid monophyly is well-established. Sequence data from this study and larger analyses of bee relationships (Danforth et al., 2006a,b) consistently support it. All colletids whose nests have been studied line the brood cells with a polyester material unique among bees (Hefetz et al., 1979; Espelie et al., 1992; Almeida, 2008), and morphological characteristics of the colletid glossa may be associated with this lining (McGinley, 1980). Results also confirm that Strenotridae do not belong within the Colletidae. Stenotritids are different from colletids in terms of their nest construction and bionomy (Houston and Thorp, 1984), as well as morphology (McGinley, 1980). Based on larval morphological characters alone, Paracolletinae and Stenotritidae were grouped together by McGinley (1981). Stenotritidae constitutes the sister-group of Colletidae according to results of this analysis and in Danforth et al. (2006a,b).

4.3. Subfamilies of Colletidae

Historically, *Colletes* and other hairy colletid groups (generally including Diphaglossinae) have been placed separately from *Hylaeus* and relatives (e.g., Ashmead, 1899). It had been supposed that Hylaeinae, Xeromelissinae, and Euryglossinae formed a distinct group within the Colletidae, despite the enormous variation in ways colletid subfamilies were imagined to relate to each other (Fig. 1). The presumed phylogenetic proximity of Hylaeinae, Xeromelissinae, and Euryglossinae was encountered in analyses of morphological data (Alexander and Michener, 1995; McGinley, 1981), but, despite their proximity, they do not form a natural group unless *Scrapter* is also included. Accumulation of evidence that *Scrapter* does not belong with the remaining Paracolletinae lead to the recent recognition of an independent monogeneric unit: Scrapterinae (Melo and Gonçalves, 2005; Ascher and Engel *in* Engel, 2005; Ascher and Engel, 2006). A sister-group relationship between *Scrapter* and Euryglossinae had been suggested by a variable but interesting morphological character: some species of both taxa possess a unique (among bees) basitibial plate not delimited by a carina, "but by a ring of large tubercles probably which are a result if subdivision of the normally elevated plate" (McGinley, 1981, 160). Additionally some larval characters support the sister-group relationship between Scrapterinae and Euryglossinae (McGinley, 1981; see comments by Engel, 2005; Ascher and Engel, 2006; Davies and Brothers, 2007), even though the analysis of the entire larval data set does not support this finding (McGinley, 1981). Results presented here unambiguously corroborate previous proposals of placing *Scrapter* in a separate subfamily and its close relationship to the Euryglossinae, despite the morphological disparity between them.

The results do not provide evidence for the grouping of Colletinae s.str. and Paracolletinae, contrary to the common and long used practice of uniting them (e.g. Michener, 1944, 1965, 1989, 2007 treated these two groups as tribes comprised into Colletinae s.l.). The proposal of keeping the two subfamilies as separate taxa (e.g., Silveira et al., 2002; Engel, 2005; Melo and Gonçalves, 2005) is compatible with the phylogenetic findings presented here (Fig. 4).

The need for additional investigation to support the separation of *Callomelitta* and *Paracolletes* from Paracolletinae becomes evident from the phylogenetic results. Although the phylogenetic affinities of *Callomelitta* and *Paracolletes* to the colletid lineages remain uncertain, there is strong evidence to assert that neither arose from within one of the formally recognized subfamilies. According to the phylogenetic analysis and the statistical topological tests (Table 5), it is possible that *Callomelitta* should be placed as sister-group to all other Paracolletinae, but this is not necessarily the case. Based on these phylogenetic results and on the observation of the distinctive morphology of these bees (Michener, 2007), the recognition of a separate subfamily to harbor *Callomelitta* would be a reasonable solution for the placement of this genus (Almeida, 2008b).

Colletinae s.str. is a clade subtended by a fairly long branch, relative to other internodes of Colletidae (Figs. 5 and S2). It is possible that the attraction of *Callomelitta* to the proximity of Colletinae was artifactual, a long branch effect. Other kinds of data are needed to elucidate this matter: morphological data should not be affected by a possible long branch effect in the same way as molecular data (Bergsten, 2005).

The placement of *Paracolletes* is another serious issue, because it entails the nomenclatural modification needed for the clade provisionally referred to by "Paracolletinae". *Paracolletes* is the type

genus of Paracolletinae and, therefore, its name-bearer taxon. Family-group names with the root "Paracollet-" must, as a result, include *Paracolletes*.

4.4. Genus-level classification of the "Paracolletinae"

Michener (2007, 145) states his awareness that *Leioproctus* s.l. most certainly does not constitute a natural group, but he acknowledges its usefulness as a diagnosable taxonomic unit. *Leioproctus* s.l. is a group recognizable on the basis of superficial similarity of its subgenera and the lack of synapomorphies present in other groups, such as *Neopasiphae* or *Eulonchopria*. Michener (2007, 141–142) resorted to caution in changing the classification: "[u]ntil more species are known and cladogeny is better studied, especially in the diverse Australian fauna, there is no point in attempting a new classification, for new combinations that would last only until the next revision would be the result".

Leioproctus s.l. (e.g. Michener, 2007) is rendered paraphyletic by a number of lineages that arise from within it. Its distinction has been traditionally made on the basis of symplesiomorphies, and the remaining groups being characterized by synapomorphies. It is not surprising, then, to find highly modified (morphologically autapomorphic) lineages arising from within it and being recognized as distinct genera, e.g., *Brachyglossula, Eulonchopria*, *Neopasiphae*, *Phenacolletes*. Part of the problem is resolved by elevating the subgenera of *Leioproctus* to the status of genus. This view has been long adopted by most South American bee taxonomists, who have traditionally treated all supra-specific taxa of South American Paracolletinae as separate genera (e.g., Moure et al., 2007; Silveira et al. 2002). In this case, the genus *Leioproctus* is restricted to the Australian region. However, the classification of Paracolletinae will require future adjustments at the generic level because the phylogenetic results very clearly show that at least two large genera constitute artificial assemblages: *Perditomorpha* and *Leioproctus* s.str. even when many subgenera are raised to generic rank.

4.5. Biogeography

Colletidae are distributed primarily in the southern hemisphere and comprise various endemic groups in Australia, South America, and Africa. As discussed elsewhere, the current distribution of colletid bees suggests the occurrence of various events of Australian-South American interchanges in the past (Almeida et al., unpublished results). Paracolletinae is formed by various clades, some endemic to the Australian region, some South American. In a similar fashion, Hylaeinae and Xeromelissinae form a clade in which the former is primarily Australian and the latter occurs from southern South America up to Central America.

Note added on proof

Publication of the article by Almeida (2008b) while the present paper was in press made available the name *Andrenopsis micheneri-anus* Almeida for the species referred to as "Paracolletinae sp." in Table 1, and Figs. 4b, 5, S1–S8.

Acknowledgments

Our thanks for comments on earlier versions of this paper to Laurence Packer, James K. Liebherr, and Richard G. Harrison. Discussions with Charles Michener, John Ascher, Terry Houston, Glynn Maynard, Gabriel Melo, and Gregg Davies contributed to the development of this research. Seán Brady, Jenn Fang, Karl Magnacca, and Sedonia Sipes provided valuable advice concerning the molecular work. We are indebted to all who supplied specimens used for DNA extraction: Antonio Aguiar, Isabel Alves dos Santos, John Ascher, Torsten Dikow, Martin Hauser, Terry Houston, Karl Magnacca, Gabriel Melo, Claus Rasmussen, Jerome Rozen, Andrew Short, Lorenzo Zanette, and Amro Zayed. We are mostly indebted to Laurence Packer for his generosity in providing samples of a large number of rare colletid species. We are grateful for the bee taxonomists listed in the methods section for assistance with the identification of taxa included in the present study. Assistance for field trips was given by Paulo Almeida, Ricardo Ayala, Ricardo Botelho, Guillermo Debandi, Mariano Devoto, Eduardo Dominguez, Connal Eardley, Dimitri Forero, Terry Houston, Laurence Packer, Charlotte Skov. Collecting permits in Argentina and Australia were facilitated by the competent work of Roberto Salinas (Servicio de Fauna Silvestre, Catamaca, Argentina); Gustavo Solá (Dirección de Recursos Naturales Renovables, Mendoza, Argentina); Ing. Ana Inés Arce (Secretaría de Medio Ambiente y Desarollo Sustentable, Salta, Argentina); Lic. Claudia Perez Miranda (Dirección de Recursos Naturales y Suelos, Tucumán, Argentina), and Mr. Danny Stefony (CALM Wildlife Branch, Western Australia). The majority of funding for this research came from a National Science Foundation Research Grant in Systematic Biology to B.N. Danforth and E.A.B. Almeida (DEB-0412176). Additional funding was provided by the A.C.Rawlins Endowment (Cornell University); Grace Griswold Fund (Cornell University); Tinker Foundation & Latin American Studies Program (Cornell University). E.A.B. Almeida is grateul to *Fundação de Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* (CAPES: Brazilian Ministry of Education) for a Ph.D. scholarship (BEX-0967/01-7) and to CNPq for a post-doctoral fellowship (PDJ-150626/2007-0).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2008.09.028.

References

Alexander, B.A., Michener, C.D., 1995. Phylogenetic studies of the families of short-tongued bees (Hymenoptera: Apoidea). Univ. Kans. Sci. Bull. 55, 377–424.
Alfaro, M.E., Zoller, S., Luzzoni, F., 2003. Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. Mol. Biol. Evol. 20, 255–266.
Almeida, E.A.B., 2008a. Colletidae nesting biology (Hymenoptera: Apoidea). Apidologie 39, 16–29.
Almeida, E.A.B., 2008b. Revised species checklist of the Paracolletinae (Hymenoptera, Colletidae) of the Australian Region, with the description of new taxa. Zootaxa 1891, 1–24.
Almeida, E.A.B., Packer, L., Danforth, B.N., 2008. Phylogeny of the Xeromelissinae (Hymenoptera: Colletidae) based upon morphology and molecules. Apidologie 39, 75–85.
Altekar, G., Dwarkadas, S., Huelsenbeck, J.P., Ronquist, F., 2004. Parallel Metropolis-coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. Bioinformatics 20, 407–415.
Ashmead, W.H., 1899. Classification of the bees. Trans. Am. Entomol. Soc. 26, 49–100.
Ascher, J.S., 2004. Systematics of the Bee Family Andrenidae (Hymenoptera: Apoidea). Ph.D. Dissertation, Cornell University, Ithaca.
Ascher, J.S., Engel, M.S., 2006. On the availability of family-group names based on *Scrapter* (Hymenoptera: Colletidae). Entomol. News 117, 117–119.
Ascher, J.S. et al., 2008. Apoidea Species Guide. Available from: <http://www.discoverlife.org/mp/20q?guide=Apoidea_species> (accessed 22.04.08).
Baker, R.H., DeSalle, R., 1997. Multiple sources of character information and the phylogeny of Hawaiian drosophilids. Syst. Biol. 46, 654–673.
Belshaw, R., Quicke, D.L.J., 1997. A molecular phylogeny of the Aphidiinae (Hymenoptera: Braconidae). Mol. Phylogenet. Evol. 7, 281–293.
Bergsten, J., 2005. A review of long-branch attraction. Cladistics 21, 163–193.
Brady, S.G., Danforth, B.N., 2004. Recent intron gain in elongation factor-1α of colletid bees (Hymenoptera: Colletidae). Mol. Biol. Evol. 21, 691–696.
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, 73–82.

Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multimodel Inference. A Practical Information-Theoretic Approach, second ed. Springer-Verlag, New York.
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.
Chang, B.S.W., Ayers, D., Smith, W.C., Pierce, N.E., 1996. Cloning of the gene encoding honeybee long-wavelength rhodopsin: a new class of insect visual pigments. Gene 173, 215–219.
Daly, H.V., Magnacca, K.N., 2003. Insects of Hawaii. Volume 17: Hawaiian Hylaeus (Nesoprosopis) Bees (Hymenoptera: Apoidea). University Hawai'i Press, Honolulu.
Danforth, B.N., 1999. Phylogeny of the bee genus *Lasioglossum* (Hymenoptera: Halictidae) based on mitochondrial cytochrome oxidase. Syst. Entomol. 24, 377–393.
Danforth, B.N., 2002. Evolution of sociality in a primitively eusocial lineage of bees. Proc. Natl. Acad. Sci. USA 99, 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. 55, 309–326.
Danforth, B.N., Fang, J., Sipes, S., 2006a. 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.
Danforth, B.N., Sipes, S., Fang, J., Brady, S.G., 2006b. The history of early bee diversification based on five genes plus morphology. Proc. Natl. Acad. Sci. USA 103, 15118–15123.
Danforth, B.N., Lin, C.P., Fang, J., 2005. How do insect nuclear ribosomal genes compare to protein-coding genes in phylogenetic utility and nucleotide substitution patterns? Syst. Entomol. 30, 549–562.
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-1 sequence data. Mol. Phylogenet. Evol. 13, 605–618.
Davies, G.B.P., Brothers, D.J., 2007. Morphology of *Scrapter* (Hymenoptera: Anthophila: Colletidae), with description of three new species and taxonomic status of five Cockerell taxa. Afr. Invert. 47, 135–183.
Davies, G.B.D., Eardley, C.D., Brothers, D.J., 2005. Eight new species of *Scrapter* (Hymenoptera: Apoidea: Colletidae), with description of *S. albifumus* and *S. albitarsus* females and a major range extension of the genus. Afr. Invert. 46, 141–179.
Doyle, J.J., Doyle, J.L., 1990. A rapid total DNA preparation procedure for fresh plant tissue. Focus 12, 13–15.
Eardley, C., 1996. The genus *Scrapter* Lepeletier and Serville (Hymenoptera: Colletidae). Afr. Entomol. 4, 37–92.
Engel, M.S., 2001. A monograph of the Baltic amber bees and evolution of the Apoidea (Hymenoptera). Bull. Am. Mus. Nat. Hist. 259, 1–192.
Engel, M.S., 2005. Family-group names for bees (Hymenoptera: Apoidea). Am. Mus. Novit. 3476, 1–33.
Espelie, K.E., Cane, J.H., Himmelsbach, D.S., 1992. Nest cell lining of the solitary bee *Hylaeus bisinuatus* (Hymenoptera: Colletidae). Experientia 48, 414–416.
Farris, J.S., 1989. The retention index and the rescaled consistency index. Cladistics 5, 417–419.
Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.
Felsenstein, J., 2004. Inferring Phylogenies. Sinauer, Sunderland.
Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. Stat. Sci. 7, 457–472.
Hancock, J.M., Tautz, D., Dover, G.A., 1988. Evolution of the secondary structures and compensatory mutations of the ribosomal RNAs of *Drosophila melanogaster*. Mol. Biol. Evol. 5, 393–414.
Hefetz, A., Fales, H.M., Batra, S.W.T., 1979. Natural polyesters: Dufour's gland macrocyclic lactones form brood cell laminesters in *Colletes* bees. Science 204, 415–417.
Hillis, D.M., Dixon, M.T., 1991. Ribsomal DNA: molecular evolution and phylogenetic inference. Q. Rev. Biol. 66, 411–453.
Houston, T.F., Thorp, R.W., 1984. Bionomics of the bee *Stenotritus greavesi* and ethological characteristics of Stenotritidae (Hymenoptera). Rec. West. Aust. Mus. 11, 375–385.
Huelsenbeck, J.P., Larget, B., Miller, R.E., Ronquist, F., 2002. Potential applications and pitfalls of Bayesian inference of phylogeny. Syst. Biol. 51, 673–688.
Huelsenbeck, J.P., Ranalla, B., 2004. Properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. Syst. Biol. 53, 904–913.
Huelsenbeck, J.P., Ronquist, F., 2005. MrBayes: Bayesian inference of phylogeny. Bioinformatics 17, 754–755.
Jander, R., 1976. Grooming and pollen manipulation in bees (Apoidea): the nature and evolution of movements involving the foreleg. Physiol. Entomol. 1, 179–194.
Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Am. Stat. Assoc. 90, 773–795.
Kishino, H., Hasegawa, M., 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominioidea. J. Mol. Evol. 29, 170–179.
Kishino, H., Miyata, T., Hasegawa, M., 1990. Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. J. Mol. Evol. 31, 151–160.
Lin, C.P., Danforth, B.N., 2004. How do insect nuclear and mitochondrial gene substitution patterns differ? Insights from Bayesian analyses of combined datasets. Mol. Phylogenet. Evol. 30, 686–702.
Linares, A.R., Hancock, J.M., Dover, G.A., 1991. Secondary structure constraints on the evolution of *Drosophila* 28S ribosomal RNA expansion segments. J. Mol. Biol. 219, 381–390.
Maddison, D.R., Maddison, W.P., 2005. MacClade 4.08. Sinauer Associates, Sunderland.
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.
Mardulyn, P., Whitfield, J.B., 1999. Phylogenetic signal in the COI, 16S, and 28S genes for inferring relationships among genera of microgastrinae (Hymenoptera: Braconidae): evidence of a high diversification rate in this group of parasitoids. Mol. Phylogenet. Evol. 12, 282–294.
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 if apoid larvae (Hymenoptera: Apoidea). Univ. Calif. Publ. Entomol. 91, 1–307 + vxi.
Melo, G.A.R., Gonçalves, R.B., 2005. Revised bee classifications (Hymenoptera, Apoidea, Apidae sensu lato). Rev. Bras. Zool. 22, 153–159.
Michener, C.D., 1944. Comparative external morphology, phylogeny, and a classification of the bees (Hymenoptera). Bull. Am. Mus. Nat. Hist. 82, 151–326.
Michener, C.D., 1965. A classification of the bees of the Australian and South Pacific regions. Bull. Am. Mus. Nat. Hist. 130, 1–362. pls. 1–15.
Michener, C.D., 1974. The Social Behavior of the Bees: A Comparative Study. The Belknap Press of Harvard University Press, Cambridge, USA.
Michener, C.D., 1979. Biogeography of the bees. Ann. MO Bot. Gard. 66, 277–347.
Michener, C.D., 1986. A review of the tribes Diphaglossini and Dissoglottini. Univ. Kans. Sci. Bull. 53, 183–214.
Michener, C.D., 1989. Classification of the American Colletinae. Univ. Kans. Sci. Bull. 53, 622–703.
Michener, C.D., 2007. The Bees of the World, 2nd ed. The John Hopkins University Press, Baltimore.
Minin, V., Abdo, Z., Joyce, P., Sullivan, J., 2003. Performance-based selection of likelihood models for phylogeny estimation. Syst. Biol. 52, 674–683.
Morgenstern, B., 1999. DIALIGN 2: improvement of the segment-to-segment approach to multiple sequence alignment. Bioinformatics 15, 211–218.
Moure, J.S., Graf, V., Urban, D., 2007. Paracolletini Cockerell, 1934. In: Moure, J.S., Urban, D., Melo, G.A.R. (Eds.), Catalogue of Bees (Hymenoptera, Apoidea) in the Neotropical Region. Sociedade Brasileira de Entomologia, Curitiba, pp. 723–743.
Nixon, K.C., Carpenter, J.M., 1993. On outgroups. Cladistics 9, 413–426.
Nixon, K.C., Carpenter, J.M., 1996. On simultaneous analysis. Cladistics 12, 221–241.
Nixon, K.C., 2002. WinClada v1.00.08. Computer program distributed by the author. Available from: <http://www.cladistics.com/wincDownload.htm>.
Nylander, J.A.A. 2004a. MrAIC.pl. Computer program distributed by the author. Evolutionary Biology Centre, Uppsala University. Available from: <http://www.ebc.uu.se/systzoo/staff/nylander/MrAIC142.zip>.
Nylander, J.A.A., 2004b. MrModeltest v2. Computer program distributed by the author. Evolutionary Biology Centre, Uppsala University. Available from: <http://www.ebc.uu.se/systzoo/staff/nylander/MrMod22.sit>.
Nylander, J.A.A., Ronquist, F., Huelsenbeck, J.P., Nieves-Aldrey, J.L., 2004. Bayesian phylogenetic analysis of combined data. Syst. Biol. 53, 47–67.
O'Toole, C., Raw, A., 1991. Bees of the World. Blanford, London, UK.
Packer, L., 2008. Phylogeny and classification of the Xeromelissinae (Hymenoptera: Apoidea, Colletidae) with special emphasis on the genus *Chilicola*. Syst. Entomol. 33, 72–96.
Pesenko, Y.A., 1999. Phylogeny and classification of the family Halictidae revised (Hymenoptera: Apoidea). J. Kans. Entomol. Soc. 72, 104–123.
Posada, D., Buckley, T.R., 2004. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. Syst. Biol. 53, 793–808.
Posada, D., Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14, 817–818.
Posada, D., Crandall, K.A., 2001. Selecting the best-fit model of nucleotide substitution. Syst. Biol. 50, 580–601.
Rozen Jr., J.G., 1984. Nesting biology of diphaglossine bees (Hymenoptera: Colletidae). Am. Mus. Novit. 2786, 1–33.
Rozen Jr., J.G., 2003. A new tribe, genus, and species of South American panurgine bee (Andrenidae: Panurginae), oligolectic on Nolana (Nolanaceae); with and appendix by J.S. Ascher: "Evidence for the phylogenetic position of Nolanomelissa from nuclear EF-1a sequence data". In: Melo, G.A.R., Alves-dos-Santos, I. (Eds.), Apoidea Neotropica: Homenagem aos 90 Anos de Jesus Santiago Moure. Editora da UNESC, Criciúma, pp. 3–19.
Rozen Jr., J.G., Michener, C.D., 1968. The biology of *Scrapter* and its cuckoo bee, *Pseudodichroa* (Hymenoptera: Colletidae and Anthophoridae). Am. Mus. Novit. 2335, 1–13.
Shimodara, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16, 1114–1116.
Silveira, F.A., Melo, G.A.R., Almeida, E.A.B., 2002. Abelhas Brasileiras. Sistemática e Identificação. Published by the authors, Belo Horizonte, Brazil.
Swofford, D.L., 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods), v4.0b10. Sinauer Associates, Sunderland, MA.
Swofford, D.L., Waddell, P.J., Huelsenbeck, J.P., Foster, P.G., Lewis, P.O., Rogers, J.S., 2001. Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. Syst. Biol. 50, 525–539.
Steel, M., Penny, D., 2000. Parsimony, likelihood, and the role of models in molecular phylogenetics. Mol. Biol. Evol. 17, 839–850.

Sullivan, J., Joyce, P., 2005. Model selection in phylogenetics. Annu. Rev. Ecol. Evol. Syst. 36, 445–466.
Talavera, G., Castresana, J., 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst. Biol. 56, 564–577.
Tautz, D., Hancock, J.M., Webb, D.A., Tautz, C., Dover, G.A., 1988. Complete sequences of the rRNA genes of *Drosophila melanogaster*. Mol. Biol. Evol. 5, 366–376.
Uzvölgyi, E., Kiss, I., Pitt, A., Arsenian, S., Ingvarsson, S., Udvardy, A., Hamada, M., Klein, G., Sumegi, J., 1988. *Drosophila* homolog of the murine Int-1 protooncogene. Proc. Natl. Acad. Sci. USA 85, 3034–3038.
Walldorf, U., Hoverman, B.T., 1990. *Apis mellifera* cytoplasmatic elongation factor 1 alpha (EF-1a) is closely related to *Drosophila melanogaster* EF-1a. FEBS 267, 245–249.
Ward, P.S., Brady, S.G., 2003. Phylogeny and biogeography of the ant subfamily Myrmeciinae (Hymenoptera: Formicidae). Invertebr. Syst. 17, 361–386.
Wcislo, W.T., Cane, J.H., 1996. Floral resource utilization by solitary bees (Hymenoptera: Apoidea) and exploitation of their stored foods by natural enemies. Annu. Rev. Entomol. 41, 257–286.
Whitfield, J.B., Kjer, K.M., 2008. Ancient rapid radiations of insects: challenges for phylogenetic analysis. Annu. Rev. Entomol. 53, 449–472.
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.
Wilcox, T.P., Zwickl, D.J., Heath, T.A., Hillis, D., 2002. Phylogenetic relationships of the dwarf boas and a comparison of Bayesian and bootstrap measures of phylogenetic support. Mol. Phylogenet. Evol. 25, 361–371.
Yang, Z., 1998. On the best evolutionary rate for phylogenetic analysis. Syst. Biol. 47, 125–133.
Zwickl, D.J., 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. Dissertation, University Texas, Austin. [Computer program distributed by the author at: <http://www.bio.utexas.edu/faculty/antisense/garli/Garli.html>].