



**HAL**  
open science

## Phylogenetic analysis of the corbiculate bee tribes based on 12 nuclear protein-coding genes (Hymenoptera: Apoidea: Apidae)

Atsushi Kawakita, John S. Ascher, Teiji Sota, Makoto Kato, David W. Roubik

### ► To cite this version:

Atsushi Kawakita, John S. Ascher, Teiji Sota, Makoto Kato, David W. Roubik. Phylogenetic analysis of the corbiculate bee tribes based on 12 nuclear protein-coding genes (Hymenoptera: Apoidea: Apidae). *Apidologie*, 2008, 39 (1), pp.163-175. hal-00891935

**HAL Id: hal-00891935**

**<https://hal.science/hal-00891935>**

Submitted on 11 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Phylogenetic analysis of the corbiculate bee tribes based on 12 nuclear protein-coding genes (Hymenoptera: Apoidea: Apidae)\*

Atsushi KAWAKITA<sup>1</sup>, John S. ASCHER<sup>2</sup>, Teiji SOTA<sup>3</sup>, Makoto KATO<sup>1</sup>,  
David W. ROUBIK<sup>4</sup>

<sup>1</sup> Graduate School of Human and Environmental Studies, Kyoto University, Kyoto, Japan

<sup>2</sup> Division of Invertebrate Zoology, American Museum of Natural History, New York, USA

<sup>3</sup> Department of Zoology, Graduate School of Science, Kyoto University, Kyoto, Japan

<sup>4</sup> Smithsonian Tropical Research Institute, Balboa, Ancon, Panama

Received 2 July 2007 – Revised 3 October 2007 – Accepted 3 October 2007

**Abstract** – The corbiculate bees comprise four tribes, the advanced eusocial Apini and Meliponini, the primitively eusocial Bombini, and the solitary or communal Euglossini. Recovering a robust phylogeny for the four tribes is of considerable importance for understanding the evolution of eusociality, yet previous morphological and molecular studies reached strikingly different conclusions. We study an expanded data set consisting of 12 nuclear genes to explore lines of support for the molecular hypothesis. Results corroborate previous molecular studies; support increases as more genes are added. Across genes, support for the molecular hypothesis is positively correlated with the number of informative sites and the relative substitution rate. Phylogenetic signals supporting the molecular tree rest almost entirely upon synonymous changes at the first and third codon positions. We discuss possible future approaches for resolving the frustratingly persistent corbiculate bee controversy.

**advanced eusociality / bumble bee / corbiculate bee / honey bee / nuclear genes / orchid bee / phylogeny / stingless bee**

## 1. INTRODUCTION

The advanced eusocial colonies of the honey bees and stingless bees are among the most elaborate of the insect societies (Wilson, 1971; Michener, 1974; Winston, 1987). In these colonies, the morphologically differentiated queens are devoted to producing offspring, and the non-reproductive workers gather and process food, feed larvae, and build and defend nests. Workers specialize through division of labor by age, size, and genetics,

and maximize their foraging efficiency by using sophisticated communication and rapid recruitment of nestmates to food sources. A myriad of fascinating biological features of the bees and their economic importance as pollinators and honey producers have sparked numerous ecological, behavioral, neural, genetic, and genomic studies, and led the western honey bee, *Apis mellifera* L., to become one of the first insect species to have its genome sequenced (The Honeybee Genome Sequencing Consortium, 2006). Despite the widely accepted paradigms of social evolution represented by such bees (e.g., division of labor, sterile workers, overlap of generations), and several recent phylogenetic studies of tribal relationships, the origin of their advanced

Corresponding author: A. Kawakita,  
[kawakita@s01.mbox.media.kyoto-u.ac.jp](mailto:kawakita@s01.mbox.media.kyoto-u.ac.jp)

\* Manuscript editor: Eduardo A.B. Almeida

Online material is available at:  
<http://www.apidologie.org>

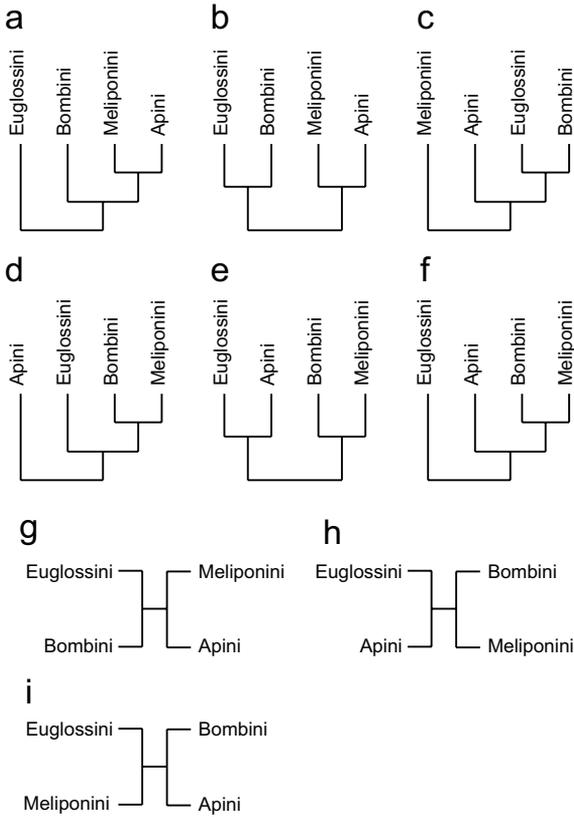
societies remain unexplained (Schultz et al., 1999, 2001; Ascher et al., 2001; Cameron and Mardulyn, 2001; Engel, 2001a; Lockhart and Cameron, 2001).

The first study that rigorously explored the phylogenetic relationships among the corbiculate tribes was by Michener (1944), who used adult external morphology to investigate the overall phylogeny of bees. He concluded that Euglossini were the first lineage to diverge, followed by Bombini, and then the advanced eusocial Meliponini and Apini (Fig. 1a), a view consistent with that of earlier authors such as Darwin (1859). Under this hypothesis, general eusociality arose once among corbiculate bees in the common ancestor of Bombini, Meliponini, and Apini. Advanced eusociality also arose only once among bees, in the common ancestor of Meliponini and Apini. Later studies explored a broad array of adult and larval morphological characters (Maa, 1953; Michener, 1974, 1990; Winston and Michener, 1977; Kimsey, 1984; Prentice, 1991; Roig-Alsina and Michener, 1993; Chavarría and Carpenter, 1994), and some of them reached more or less different conclusions (e.g., Fig. 1b, Michener, 1974; Fig. 1c, Winston and Michener, 1977). However, all recent analyses of extant corbiculate morphology (Michener, 1990; Prentice, 1991; Roig-Alsina and Michener, 1993; Chavarría and Carpenter, 1994; Schultz et al., 1999; Ascher et al., 2001) are consistent with the hypothesis proposed by Michener (1944). More recently, Engel (2001a, b) significantly expanded the morphological data set by including fossil corbiculate tribes; the combined data set lent further support to the Michener (1944) phylogeny. Cardinal and Packer (2007) studied a new set of morphological characters derived from the sting apparatus. Their results were perfectly congruent with the previous morphological studies. Noll (2002) analyzed a large series of behavioral characters and arrived at the same phylogenetic interpretation.

Despite the near unanimity of morphological, palaeontological, and behavioral data, phylogenetic analyses using DNA sequences suggest strikingly different relationships among the four tribes (Sheppard and McPheron, 1991; Cameron, 1991,

1993; Koulianos et al., 1999; Mardulyn and Cameron, 1999; Cameron and Mardulyn, 2001). Most notably, Bombini and Meliponini are nearly always recovered as sisters, a clade not supported by known morphological data (Figs. 1d–f). Sheppard and McPheron (1991) and Cameron (1991, 1993) were the first to propose this alternative phylogeny, in which sequences of nuclear and mitochondrial ribosomal RNA were used to reconstruct the corbiculate bee tribal relationships. Under this hypothesis, advanced eusociality arose twice, independently, in Meliponini and Apini; reversals from advanced eusociality to other stages, while equally parsimonious, were considered very unlikely to occur because the queens of advanced eusocial colonies have lost the ability to survive on their own without workers. Later studies utilized several other genes, from both mitochondrial and nuclear genomes (Koulianos et al., 1999; Mardulyn and Cameron, 1999; Cameron and Mardulyn, 2001; Cameron, 2003). These studies consistently recovered Bombini and Meliponini as a monophyletic group excluding Apini. The most comprehensive study was that of Cameron and Mardulyn (2001), who analyzed a data matrix of four genes (cytochrome b, 16S rRNA, opsin, and 28S rRNA). Their results gave Bombini and Meliponini strong support as sister groups, although the relationship of this clade to the remaining two tribes remained unresolved. Cha et al. (2007) recently analyzed the complete mitochondrial genome of Bombini, Meliponini, and Apini, also finding that the first two tribes are unambiguously recovered as sisters (the mitochondrial genome of Euglossini is still unavailable).

Resolving the controversy over tribal relationships in corbiculate bees has proven to be remarkably difficult (Schultz et al., 1999, 2001; Ascher et al., 2001; Lockhart and Cameron, 2001), and additional data from new sources are needed to broaden our perspective on the issue and to allow for more sophisticated analyses of molecular evolution. Recent publication of the honey bee genome had facilitated assembly of orthologous sequences of previously unexplored nuclear genes in many bees (Danforth et al., 2004). Because



**Figure 1.** Previously proposed rooted phylogenies of the four corbiculate tribes (a–f) and the three theoretically possible unrooted topologies for the four tribes (g–i). **a.** Phylogeny favored by the majority of morphological data sets (Michener, 1944, 1990; Maa, 1953; Prentice, 1991; Roig-Alsina and Michener, 1993; Chavarría and Carpenter, 1994; Schultz et al., 1999; Ascher et al., 2001; Engel, 2001a, b). **b.** Phylogenetic hypothesis of Michener (1974). **c.** Hypothesis proposed by Winston and Michener (1977) and Kimsey (1984). **d–f.** Phylogenetic hypotheses proposed by molecular data (Cameron, 1991, 1993; Sheppard and McPherson, 1991; Koulianos et al., 1999; Mardulyn and Cameron, 1999; Cameron and Mardulyn, 2001).

nuclear genes evolve more slowly and have a more even base composition (and thus a larger character state space) than do mitochondrial genes, they are generally better markers for inferring deeper or more ancient divergences among taxa (Lin and Danforth, 2003). In addition, the base composition of bee mitochondrial genomes has the highest AT-bias among the more than 40 complete insect mitochondrial genomes sequenced to date (Cha et al., 2007). The resulting limited character state space could easily result in spurious phylogenetic inference (Nardi et al., 2003; Simmons et al., 2004). Therefore, extensive additional data from nuclear genes are needed to help refine results obtained from previous molecular studies, which often relied on mitochondrial sequences (Cameron, 1991, 1993; Koulianos et al., 1999; Cameron and Mardulyn, 2001), ribosomal sequences for which alignment can be a confounding factor (Sheppard and McPherson, 1991), and/or

only limited nuclear data (i.e., a single gene such as opsin; Mardulyn and Cameron, 1999; Ascher et al., 2001). In this study, we construct a data matrix consisting of 12 nuclear protein-coding genes, four newly sequenced, a significant addition to molecular studies of the bee tribes considered here. Using the expanded data set, we set a goal of identifying more precisely the components of the molecular data that support the sister grouping of Bombini and Meliponini.

## 2. MATERIALS AND METHODS

### 2.1. Phylogenetic analysis of the combined 12-gene data set

Detailed information on taxon sampling, the 12 genes sequenced, and molecular protocols are provided in the supplementary methods and primer sequences are described in the supplementary tables and figure

(<http://www.apidologie.org/article/apido/olm/2008/01/m07101/m07101.html>). We first analyzed the combined 12-gene data for all taxa to determine whether or not the results are compatible with the previous molecular hypothesis. Phylogenetic trees were estimated using parsimony, maximum likelihood, and Bayesian methods following the procedures detailed in the supplementary materials. We also conducted phylogenetic analyses using only the corbiculate ingroup sequences. As discussed previously (Ascher et al., 2001), phylogenetic analysis of the corbiculate tribes using molecular data is often complicated by spurious non-monophyly of the corbiculate bees. Because the corbiculate bees are among the most well-established clades of bees, supported by a significant number of synapomorphies (Roig-Alsina and Michener, 1993), such results are likely artifacts due to systematic sources of error (but see Cameron and Mardulyn (2003), who suggested that non-monophyly of corbiculate bees may be genuine). The crux of controversy over corbiculate bee phylogeny are incongruencies of tribal relationships between the molecular and non-molecular data, regardless of where the outgroup is connected to the ingroup. Figure 1 illustrates the three theoretically possible topologies for a four-tribe phylogeny. Morphological, behavioral, and palaeontological data overwhelmingly support the first topology (Fig. 1g), whereas molecular data favor Bombini and Meliponini as sisters (Fig. 1h) regardless of root position. To explore the source of phylogenetic conflict underlying the ingroup phylogeny is our stated purpose here, while outgroup issues would more appropriately be treated separately. We therefore analyzed the combined ingroup sequences using parsimony, maximum likelihood, and Bayesian methods.

To statistically assess whether the molecular data are incongruent with traditional morphological hypothesis, we performed the likelihood-based Shimodaira-Hasegawa test (SH-test; Shimodaira and Hasegawa, 1999) using trees estimated with and without the outgroup. The alternative topologies for the tests were defined by constraining Meliponini and Apini as monophyletic and performing likelihood heuristic searches.

## 2.2. Analysis of individual gene partitions

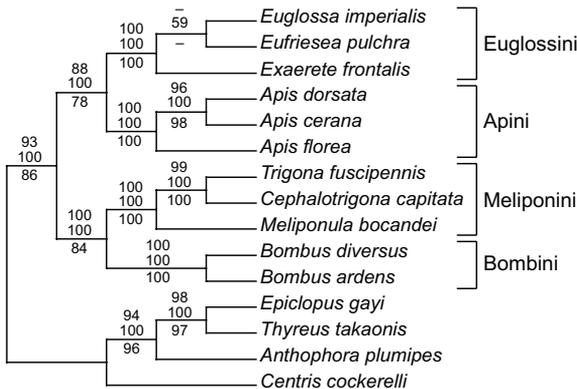
To determine whether individual genes recover different tribal relationships, we also analyzed each

gene using parsimony, likelihood, and Bayesian analyses. Only the ingroup sequences were used for the analyses. In addition, we compared nodal support values provided by each gene for the Bombini + Meliponini group, and tested whether the strength of support for this topology was correlated with the following six parameters: number of informative sites, base composition (proportion of A + T), consistency index (CI; excluding uninformative characters), relative substitution rate, shape parameter of the gamma distribution ( $\alpha$ ), and the proportion of invariable sites (PI). The following four support indices were used to assess the correlation between the above parameters and phylogenetic performance: parsimony and likelihood non-parametric bootstrap support, Bayesian posterior probability, and partitioned Bremer support (PBS; Bremer, 1994). Because PBS is dependent on the amount of signal content of a given partition, we divided PBS by the minimum number of steps for that gene (Baker et al., 2001) to obtain a standardized comparative measure of relative contribution of each gene to the node of interest. Detailed methods for calculating the parameters and support indices are given in the supplementary materials.

## 2.3. Analysis by codon positions and amino acid sequences

To determine whether molecular characters with different evolutionary constraints vary in their contributions to the Bombini + Meliponini group, we partitioned the combined 12-gene data set by codon positions and analyzed each partition under parsimony, likelihood, and Bayesian methods. Most substitutions at the third codon position do not change the amino acid and thus are synonymous, whereas most changes at the first position and all changes at the second position are non-synonymous, thus potentially constrained by protein function. We also analyzed the combined data by translating nucleotide sequences into amino acids to assess the contribution of non-synonymous substitutions to overall phylogenetic resolution (but see Simmons (2000) for confounding effect of homoplasy inherent in the amino acid coding). Methods for the analysis of amino acid sequences are given in the supplementary materials.

Because the above analysis of codon positions resulted in contrasting patterns of support for the Bombini + Meliponini group, we tested whether this variation occurs from differences in the amount



**Figure 2.** The maximum likelihood cladogram resulting from simultaneous analysis of 12 genes. Nodal support is given by likelihood bootstrap, Bayesian posterior probability, and parsimony bootstrap values, from top to bottom.

of phylogenetic signal contained in each partition by using the data decisiveness index (DD; Goloboff, 1991). Data decisiveness can be calculated as:

$$DD = (S^* - S)/(S^* - M)$$

where  $S^*$  is the mean length of 10 000 random tree topologies,  $S$  is the length of the shortest tree, and  $M$  is the minimum number of steps possible for each data set. More decisive data sets (higher DD value) allow the investigator to more confidently choose some topologies over others. In addition, we calculated PBS for each codon position to obtain a comparative measure of phylogenetic contribution, by each partition, to the Bombini + Meliponini group.

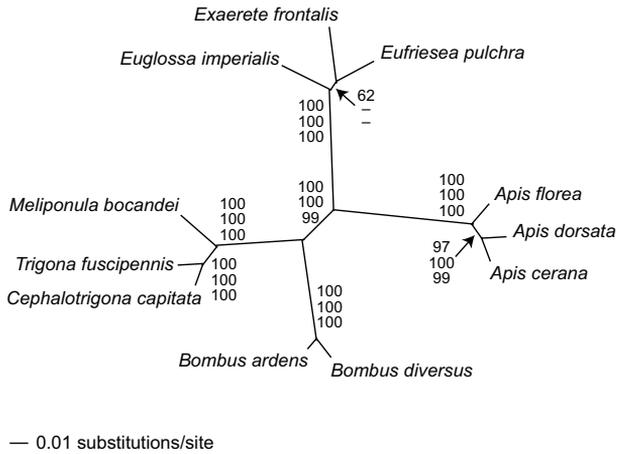
#### 2.4. Analysis of the effect of gene number on nodal support

We further explored the impact of character sampling on phylogenetic performance by analyzing the pattern of increase in nodal support as more genes were included in the data set. We created data sets consisting of one to 12 genes and calculated non-parametric bootstrap values with 1000 replications for the Bombini + Meliponini group. Twenty random gene combinations were produced for each data set size category, which were used to obtain the mean and standard deviation of support values (there are only 12 possible combinations for 1-gene and 11-gene data sets and one combination for 12-gene data). Due to computational intensity, support indices were calculated using only the parsimony criterion. Analyses were done using the data set with all positions included and those partitioned by codon positions.

### 3. RESULTS

Basic sequence statistics are provided in Supplementary Table I. Simultaneous analysis of the combined 12-gene data set produced a well-corroborated phylogeny (Fig. 2). All methods of analysis were mutually congruent except for the parsimony analysis, which differed from the likelihood and Bayesian analyses in the relative branching order within Euglossini, which was only weakly supported by any of the methods. The corbiculate bees were recovered as monophyletic, and Euglossini + Apini and Bombini + Meliponini formed two pairs of strongly supported sister tribes. Analysis performed without the non-corbiculate outgroups also produced a well-resolved topology consistent with these tribal relationships (Fig. 3). Constraining the two advanced eusocial tribes (Meliponini and Apini) as monophyletic resulted in a significant decrease in likelihood score both with and without the outgroups (SH-test,  $P < 0.001$ ).

In contrast, analyses of individual gene partitions and excluding outgroups resulted in variable results (Fig. 4), and all three theoretically possible tribal groupings (Figs. 1g–i) were recovered: two genes recovered (Euglossini + Bombini) + (Meliponini + Apini), three genes supported (Euglossini + Meliponini) + (Bombini + Apini), and the remaining seven genes favored (Euglossini + Apini) + (Bombini + Meliponini). However, statistical support for the tribal groupings by individual genes was generally very low, except for CAD, Nak, and Pol II, all of which



**Figure 3.** Unrooted maximum likelihood phylogram resulting from simultaneous analysis of 12 genes. Nodal support values as in Figure 2.

supported the group Bombini + Meliponini (Fig. 4).

We then analyzed whether support values provided by each partition to the Bombini + Meliponini group were correlated with any statistical properties of the data themselves (Fig. 5; Supplementary Tab. II). Number of informative sites and relative evolutionary rates were positively correlated with all measures of nodal support (parsimony and likelihood bootstrap, Bayesian posterior probability, and standardized PBS values; Spermann's rank correlation test,  $P < 0.05$ ). Proportion of A + T was also positively correlated with nodal support ( $P < 0.05$ ), except for the likelihood bootstrap values (Fig. 5). There was no correlation between other parameters and nodal support values provided for the Bombini + Meliponini group.

When each codon position was analyzed separately, strong support for the Bombini + Meliponini group was found only in the first and third codon positions (Tab. I). In contrast, the second codon position and also the amino acid sequences provided very weak support to this relationship. The lack of strong support in the second position data is not due solely to insufficient phylogenetic signal, because the data decisiveness index indicated that the second position data provide the most decisive data of the three codon positions (Tab. II). The second codon position had the highest proportion of A + T, and the third position had significantly heterogeneous base composition

across taxa (Tab. II). Overall, the first codon position sites contributed most strongly per site to supporting the Bombini + Meliponini group, as judged by the standardized PBS values (Tab. II).

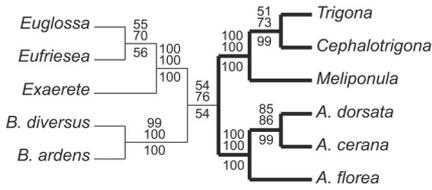
Support for the Bombini + Meliponini group provided by all-codon-position data steadily increased as more genes were added to the data set (Fig. 6). Similar trends were found for the first and third position data sets, although the third position data varied greatly in support they provided, especially with data sets consisting of fewer genes. In contrast, the data set containing only the second positions provided consistently weak support, and there was no obvious trend of increased support values (bootstrap analysis was not performed with data sets of one and two genes due to the limited number of informative sites).

## 4. DISCUSSION

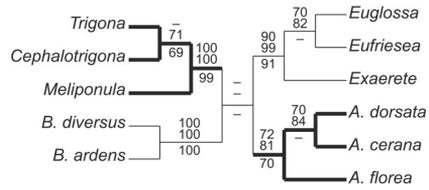
### 4.1. Phylogenetic support for the Bombini + Meliponini group

Overall, the present phylogenetic analyses based on 12 nuclear protein-coding genes strongly corroborated the results of previous molecular studies that recovered Bombini and Meliponini as a putative monophyletic group (Figs. 2, 3). Constraining Meliponini and Apini as monophyletic, consistent with

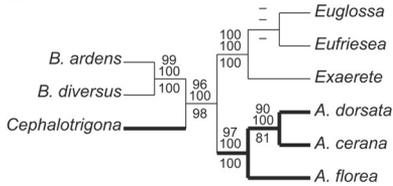
Argk



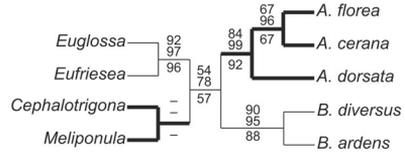
Bub3



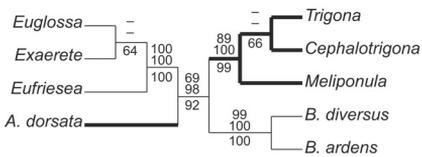
CAD



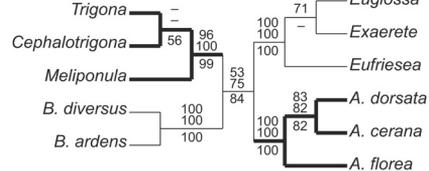
CamkII



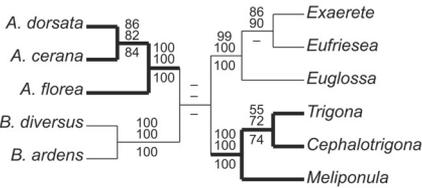
Dnk



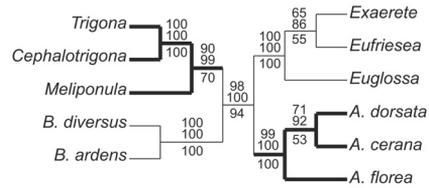
EF-1α



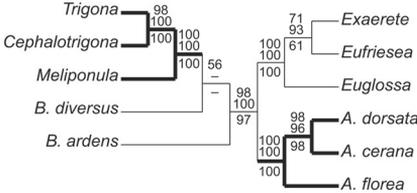
Gyk



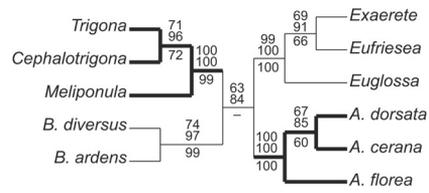
NaK



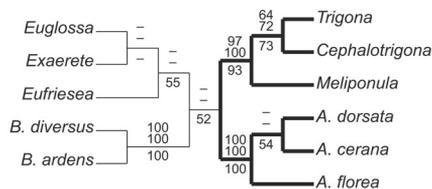
Pol II



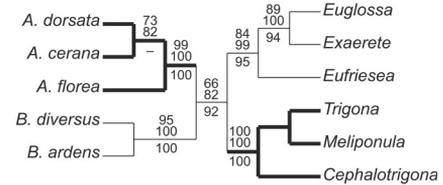
Rh



W



Wg



**Figure 4.** Maximum likelihood topology for the corbiculate ingroup based on analyses of individual gene partitions. Nodal support values as in Figure 2. The two advanced eusocial tribes are indicated in thick lines.

the traditional view of corbiculate bee relationships, resulted in significant ( $P < 0.001$ ) decrease in likelihood score, indicating that phylogenetic signals in the sequence data supporting the Bombini + Meliponini group are significantly stronger than those supporting the traditional morphological hypothesis. In contrast, phylogenetic analyses of individual gene partitions and excluding outgroups resulted in variable tree topologies, with all the three theoretically possible tribal relationships recovered (Fig. 4). Nodal support values (Fig. 6) clearly indicate that robustness of the Bombini + Meliponini group steadily improves as more genes are sampled. This suggests that an expanded molecular data set would likely converge on the same result, as long as standard phylogenetic methods are employed for reconstructing the phylogeny, but such convergence does not entail that the phylogenetic result is correct (see below).

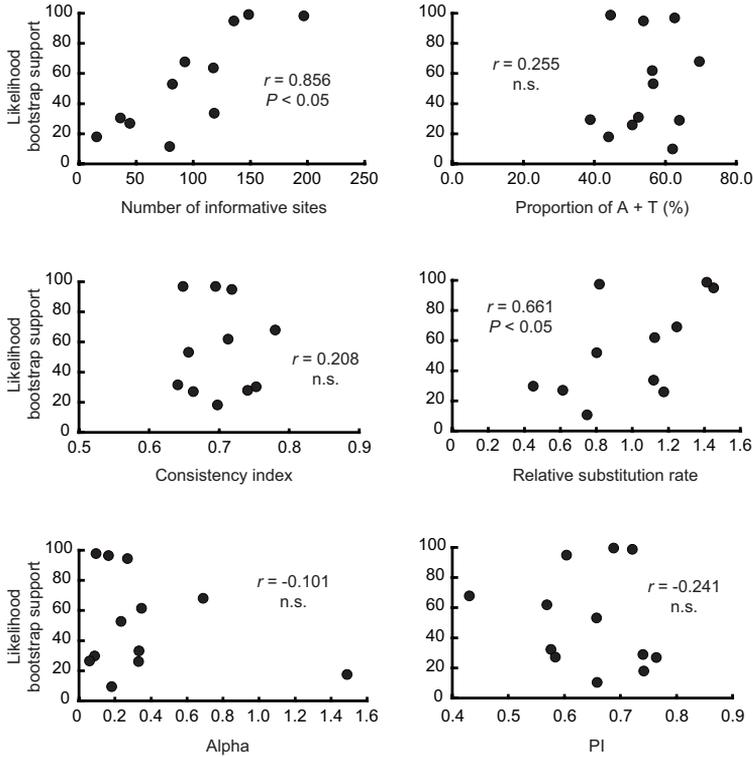
Our sampling of the 12 genes also allowed the examination of statistical properties in molecular data that are likely responsible for supporting for the Bombini + Meliponini group (Fig. 5). The tests of correlation between various parameters and support values for the Bombini + Meliponini group indicated that the number of informative sites, relative substitution rate, and to a lesser extent AT-bias are associated with the strength of nodal support (Fig. 5). The number of informative sites and substitution rate likely reflect the amount of phylogenetic information contained in each gene partition. Thus, these results suggest that nodal support becomes stronger as more information is present in the data, which is consistent with our finding that support values increased with gene number (Fig. 6). Also, an AT-bias is known to correlate positively with base substitution rate in insect DNA (Lin and Danforth, 2003), which further suggests that information content alone determines the likelihood of providing support for the Bombini + Meliponini group.

When the three codon positions were analyzed separately, only the first and third position data provided support for Bombini + Meliponini (Tab. I, Fig. 6). Lack of support by second position data is not due to absence of phylogenetic signal in the parti-

tion, as shown by strong decisiveness (Tab. II). In fact, the second position data were the most decisive of all codon positions (highest DD), yet provided very weak support to the tribal relationship (Tabs. I, II). This indicates that substitutions in the second position data contain signals for other nodes on the phylogeny (e.g., monophyly of each tribe) but do not contribute to robustness of the tribal relationship. The first and third codon positions contained more informative sites than the second position, which is likely one reason for their higher support values. Overall, the first codon position contributed roughly five times more signal than the third position, as judged from standardized PBS values. In turn, amino acid sequences did not provide meaningful support for the Bombini + Meliponini sister relationship. This indicates that the majority of phylogenetic signals supporting the observed tribal relationship come from synonymous substitutions at the first and third codon positions, at least with our data matrix of 12 genes. Although the vast majority (96%) of all possible first position substitutions are non-synonymous, this position yielded results much more consistent with the third position, rather than the second position for which all substitutions are non-synonymous. This further indicates that support for the Bombini + Meliponini relationship is largely contributed by synonymous substitutions. It thus remains to be determined whether non-synonymous substitutions also yield the same topology when even more characters are sampled for the analysis.

#### 4.2. Resolving the corbiculate controversy

Overall, the results of our present analysis provide strong support for the previous molecular hypothesis and therefore fuel rather than resolve the existing controversy over the phylogenetic relationships of corbiculate tribes. Resolving the controversy would ultimately require identifying the source of error that has led either morphological or molecular data to mislead phylogenetic inference. There has been a long history of, and



**Figure 5.** Relationships between various phylogenetic statistics of individual gene partitions and maximum likelihood bootstrap values provided for the Bombini + Meliponini group. Correlations were assessed by Spearman's rank correlation test.

**Table I.** Parsimony and likelihood bootstrap and Bayesian posterior probability values for the Bombini + Meliponini group in the simultaneous analysis of 12 genes using the ingroup sequences. Values are indicated in bold when 95 or above.

Partition	Parsimony	Likelihood	Bayesian
All positions	<b>99</b>	<b>100</b>	<b>100</b>
1st position	<b>99</b>	<b>97</b>	<b>100</b>
2nd position	46	51	47
3rd position	<b>95</b>	<b>100</b>	<b>100</b>
Amino acid	51	44	74

considerable debate that has been devoted to, morphological systematics of the corbiculate bees (Michener, 1944, 1974, 1990, 2000; Winston and Michener, 1977; Kimsey, 1984; Prentice, 1991; Roig-Alsina and Michener, 1993; Chavarría and Carpenter, 1994; Schultz et al., 1999; Ascher et al., 2001; Engel, 2001a, b; Cardinal and Packer, 2007). Therefore, we do not discuss the robustness of morphological hypothesis here other than to

note that no strong morphological character has been identified that might support Meliponini + Bombini. In turn, molecular phylogenetic analyses using large data sets have only recently become routine, and thus it is possible that the molecular results are misled by undetected, or as yet unknown, systematic biases.

The difficulty of inferring corbiculate phylogenetic relationships using molecular data

**Table II.** Statistics of different codon position data calculated for the ingroup sequences. Partitioned Bremer support (PBS) is given for the node supporting Bombini + Meliponini.

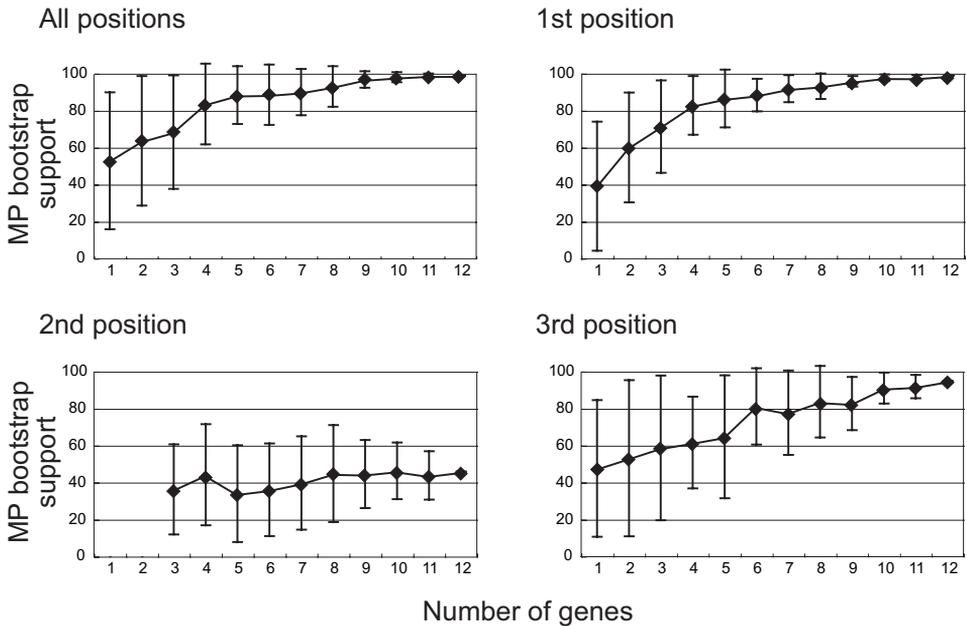
	Informative sites	Data decisiveness	PBS	PBS/min steps <sup>a</sup>	CI <sup>b</sup>	A + T (%) <sup>c</sup>	Base comp. heterogeneity <sup>d</sup>
All positions	1114	0.702	31	0.018	0.685	0.547	0.725
1st position	134	0.798	13	0.061	0.760	0.497	1.000
2nd position	54	0.865	0	0.000	0.792	0.608	1.000
3rd position	926	0.683	18	0.012	0.672	0.535	< 0.001

<sup>a</sup> Partitioned Bremer support divided by minimum number of steps.

<sup>b</sup> Consistency index excluding uninformative characters.

<sup>c</sup> Proportion of A's and T's in the partition.

<sup>d</sup> *P*-values of the chi-square test of homogeneity in base composition. across taxa.



**Figure 6.** Relationships between number of sampled genes and parsimony bootstrap support for the Bombini + Meliponini group (means  $\pm$  SD). Support values were calculated for 20 randomly sampled gene sets for each size category.

stems from the fact that there have been repeated extinctions of ancient corbiculate lineages after the extant four tribes diverged from their common ancestors (Engel, 2001a, b). This resulted in very long branches leading to the extant crown group within each tribe (the four long branches in Fig. 3), relative to short internal branches separating them. This juxtaposition of long branches connected by short internodes makes it difficult to

distinguish true historical signals in the molecular data from those originated by secondary convergence. In an extreme example where internal branches have effectively zero length with respect to the characters sampled (i.e., the four tribes diverged almost simultaneously), two tribes could still be strongly recovered as a group if they share similar patterns of base substitution that occurred *after* they had split from the common ancestor. It is therefore

important to determine if there is any significant heterogeneity in patterns of molecular evolution among tribes, which could easily lead to artifactual groupings irrespective of underlying true historical signals. Empirical examples of such biases include divergent evolutionary rate (Philippe et al., 2005) and heterogeneous base composition across taxa (Gruber et al., 2007). Although no strong evidence for such biases is detected yet in the corbiculate bees (see Supplementary Tab. IV for comparison of base composition across taxa), future studies should rigorously investigate the possibility that heterogeneous patterns of molecular evolution may have misled the interpretation of their phylogenetic relationships (see also Ascher et al., 2001).

In addition to exploration of systematic biases, studies that employ novel approaches may greatly facilitate our understandings of corbiculate relationships. With the honey bee genome now available, there is a flood of new information that can be exploited for studies of corbiculate bee phylogeny. Such novel approaches involving genome-level analyses would likely reveal characters less likely to be prone to simple artifacts and therefore particularly relevant for choosing between alternative phylogenetic hypotheses, thereby finally bring about a clear resolution to the long-lasting corbiculate controversy.

## ACKNOWLEDGEMENTS

We are grateful to Dr. Bryan Danforth for generously sharing information on CAD, NaK, Pol II, and W primers. This work was supported by Grant-in-Aid for Scientific Research (#02440217) from the Japan Ministry of Education, Science, and Culture.

**Analyse phylogénétique des tribus d'abeilles corbiculées sur la base de 12 gènes nucléaires codant pour des protéines (Hymenoptera : Apoidea : Apidae).**

**abeille domestique / Apini / bourdon / Bombini / abeille sans aiguillon / Meliponini / abeille euglossine / Euglossini / abeille corbiculée / gène nucléaire / phylogénèse / eusocialité avancée**

**Zusammenfassung – Phylogenetische Analyse der corbiculaten Bienen auf der Grundlage von**

**12 Kernprotein-Genen (Hymenoptera: Apoidea: Apidae).** Die Völker der hocheusozialen Honigbienen und der Stachellosen Bienen zählen zu den höchstentwickeltesten Insektensozialitäten. In diesen Völkern fällt den morphologisch differenzierten Königinnen die Aufgabe der Reproduktion zu, während die nichtreproduktiven Arbeiterinnen die Larven füttern, das Nest bauen und verteidigen und mittels fortgeschrittener Kommunikationssysteme Futter sammeln. Die Honigbienen und Stachellosen Bienen sind eng verwandt mit den primitiv eusozialen Hummeln und den solitären oder kommunalen Prachtbienen. Diese vier Gruppen bilden die corbiculaten Bienen, benannt nach der auf Pollentransport spezialisierten Struktur auf den Hinterbeinen, dem Pollenkörbchen oder Corbicula. Die Aufdeckung der Beziehungen innerhalb dieser vier Gruppen mittels einer robusten Phylogenie ist wichtig für das Verständnis des Ursprungs der fortgeschrittenen Eusozialität. Morphologische, paläontologische, molekulare und Verhaltensstudien kamen hier jedoch zu sehr unterschiedlichen Schlüssen. Während die nichmolekularen Studien im allgemeinen eine Schwesterngruppenbeziehung von Honigbienen und Stachellosen Bienen stützen und damit auf einen einmaligen Ursprung fortgeschrittener Sozialverhaltensweisen, lieferten molekulare Studien Hinweise auf eine enge Beziehung von Hummeln und Stachellosen Bienen und damit auf einen zweifachen Ursprung hoher Eusozialität. Wir führten hier eine molekulare phylogenetische Analyse mit einem erweiterten Set aus 12 für Kernproteine kodierenden Genen durch. Die Ergebnisse bestätigen nicht nur die früheren molekularen Analysen, sondern dieses erweiterte Genet erhöht sogar die Signifikanzwerte für die Richtigkeit der molekularen Hypothese. Die Schwesterngruppenbeziehung von Hummeln und Stachellosen Bienen ist positiv korreliert mit der Zahl phylogenetisch informativer Positionen und mit den relativen Basensubstitutionsraten. Das phylogenetische Signal, das die Hummeln + Stachellosen Bienen Gruppierung stützt, beruht vor allem auf der ersten und dritten Kodon-Position, mit geringem Beitrag seitens der zweiten Position. Auch die Aminosäuresequenzen liefern keine starke Unterstützung für diese Gruppierung. Der grösste Teil der phylogenetischen Information kommt daher von synonymen Basensubstitutionen in der ersten und zweiten Position, die keine Veränderungen in der Aminosäuresequenz zur Folge haben. Wir diskutieren mögliche zukünftige Ansätze für die Klärung dieser frustrierenden Kontroverse zur Phylogenie der corbiculaten Bienen, die aus der Inkongruenz zwischen Morphologie (einschliesslich Verhalten und Paläontologie) und DNA hervorgeht.

**Fortgeschrittene Eusozialität / Hummeln / Corbiculate Bienen / Honigbiene / Kerngene / Prachtbienen / Phylogenie / Stachellose Bienen**

## REFERENCES

- Ascher J.S., Danforth B.N., Ji S. (2001) Phylogenetic utility of the major opsin in bees (Hymenoptera: Apoidea): a reassessment, *Mol. Phylogenet. Evol.* 19, 76–93.
- Baker R.H., Wilkinson G.S., Desalle R. (2001) Phylogenetic utility of different types of data used to infer evolutionary relationships among stalk-eyed flies (Diopsidae), *Syst. Biol.* 50, 87–105.
- Bremer K. (1994) Branch support and tree stability, *Cladistics* 10, 295–304.
- Cameron S.A. (1991) A new tribal phylogeny of the Apidae inferred from mitochondrial DNA sequences, in: Smith D.R. (Ed.), *Diversity in the genus, Apis*, Westview Press, Oxford, pp. 71–87.
- Cameron S.A. (1993) Multiple origins of advanced eusociality in bees inferred from mitochondrial DNA sequences, *Proc. Natl. Acad. Sci. USA* 90, 8687–8691.
- Cameron S.A. (2003) Data from the elongation factor-1a gene corroborates the phylogenetic pattern from other genes revealing common ancestry of bumble bees and stingless bees (Hymenoptera: Apinae), III Seminario Mesoamericano sobre Abejas sin Aguijón, Tapachula, Chiapas, Mexico.
- Cameron S.A., Mardulyn P. (2001) Multiple molecular data sets suggest independent origins of highly eusocial behavior in bees (Hymenoptera: Apinae), *Syst. Biol.* 50, 194–214.
- Cameron S.A., Mardulyn P. (2003) The major opsin gene is useful for inferring higher level phylogenetic relationships of the corbiculate bees, *Mol. Phylogenet. Evol.* 28, 610–613.
- Cardinal S., Packer L. (2007) Phylogenetic analysis of the corbiculate Apinae based on morphology of the sting apparatus (Hymenoptera: Apidae), *Cladistics* 23, 99–118.
- Cha S.Y., Yoon H.J., Lee E.M., Yoon M.H., Hwang J.S., Jin B.R., Han Y.S., Kim I. (2007) The complete nucleotide sequence and gene organization of the mitochondrial genome of the bumblebee, *Bombus ignitus* (Hymenoptera: Apidae), *Gene* 392, 206–220.
- Chavarría G., Carpenter J.M. (1994) “Total evidence” and the evolution of highly social bees, *Cladistics* 10, 229–258.
- Danforth B.N., Fang J., Sipes S., Brady S.G., Almeida E. (2004) Phylogeny and molecular systematics of bees (Hymenoptera: Apoidea), Cornell University, Ithaca, NY, [online] <http://www.entomology.cornell.edu/BeePhylogeny/> (accessed on 25 October 2007).
- Darwin C. (1859) *On the origin of species by means of natural selection*, John Murray, London.
- Engel M.S. (2001a) Monophyly and extensive extinction of advanced eusocial bees: Insights from an unexpected Eocene diversity, *Proc. Natl. Acad. Sci. USA* 98, 1661–1664.
- Engel M.S. (2001b) A monograph of the Baltic amber bees and evolution of the Apoidea (Hymenoptera), *Bull. Am. Mus. Nat. Hist.* 259, 1–192.
- Goloboff P.A. (1991) Homoplasy and the choice among cladograms, *Cladistics* 7, 215–232.
- Gruber K.F., Voss R.S., Jansa S.A. (2007) Base-compositional heterogeneity in the RAG1 locus among didelphid marsupials: implications for phylogenetic inference and the evolution of GC content, *Syst. Biol.* 56, 83–96.
- Kimsey L.S. (1984) A re-evaluation of the phylogenetic relationships in the Apidae (Hymenoptera), *Syst. Entomol.* 9, 435–441.
- Koulianos S., Schmid-Hempel R., Roubik D.W., Schmid-Hempel P. (1999) Phylogenetic relationships within the corbiculate Apinae (Hymenoptera) and the evolution of eusociality, *J. Evol. Biol.* 12, 380–384.
- Lin C.P., Danforth B.N. (2003) How do insect nuclear and mitochondrial gene substitution patterns differ? Insights from Bayesian analysis of combined datasets, *Mol. Phylogenet. Evol.* 30, 686–702.
- Lockhart P.J., Cameron S.A. (2001) Trees for bees, *Trends Ecol. Evol.* 16, 84–88.
- Maa T. (1953) An inquiry into the systematics of the tribus Apidini or honeybees (Hym.), *Treubia* 21, 525–640.
- Mardulyn P., Cameron S.A. (1999) The major opsin in bees (Insecta: Hymenoptera): a promising nuclear gene for higher level phylogenetics, *Mol. Phylogenet. Evol.* 12, 168–176.
- Michener C.D. (1944) Comparative external morphology, phylogeny, and classification of bees, *Bull. Am. Mus. Nat. Hist.* 82, 151–362.
- Michener C.D. (1974) *The social behavior of the bees: a comparative study*, The Belknap Press, Cambridge, MA.
- Michener C.D. (1990) Classification of the Apidae (Hymenoptera), *Univ. Kans. Sci. Bull.* 54, 75–164.
- Michener C.D. (2000) *The bees of the world*, John Hopkins University Press, Baltimore, MD, pp. xiv + 913.
- Nardi F., Spinsanti G., Boore J.L., Carapelli A., Dallai R., Frati F. (2003) Hexapod origins: monophyletic or paraphyletic? *Science* 299, 1887–1889.
- Noll F.B. (2002) Behavioral phylogeny of corbiculate Apidae (Hymenoptera: Apinae), with special reference to social behavior, *Cladistics* 18, 137–153.
- Philippe H., Lartillot N., Brinkmann H. (2005) Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa, and Protosmia, *Mol. Biol. Evol.* 22, 1246–1253.
- Prentice M. (1991) Morphological analysis of the tribes of Apidae, in: Smith D.R. (Ed.), *Diversity in*

- the genus *Apis*, Westview Press, Oxford, pp. 51–69.
- Roig-Alsina A., Michener C.D. (1993) Studies of the phylogeny and classification of long-tongued bees (Hymenoptera: Apoidea), Univ. Kans. Sci. Bull. 55, 124–162.
- Schultz T.R., Engel M.S., Prentice M. (1999) Resolving conflict between morphological and molecular evidence for the origin of eusociality in the ‘corbiculate’ bees (Hymenoptera: Apidae): a hypothesis-testing approach, Univ. Kans. Nat. Hist. Mus. Spec. Publ. 24, 125–138.
- Schultz T.R., Engel M.S., Ascher J.S. (2001) Evidence for the origin of eusociality in the corbiculate bees (Hymenoptera: Apidae), J. Kans. Entomol. Soc. 74, 10–16.
- Sheppard W.S., McPheron B.A. (1991) Ribosomal DNA diversity in Apidae, in: Smith D.R. (Ed.), Diversity in the genus *Apis*, Westview Press, Oxford, pp. 89–102.
- Shimodaira H., Hasegawa M. (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference, Mol. Biol. Evol. 16, 114–116.
- Simmons M.P. (2000) A fundamental problem with amino-acid-sequence characters for phylogenetic analyses, Cladistics 16, 274–282.
- Simmons M.P., Reeves A., Davis J.I. (2004) Character-state space versus rate of evolution in phylogenetic inference, Cladistics 20, 194–204.
- The Honeybee Genome Sequencing Consortium (2006) Insights into social insects from the genome of the honeybee *Apis mellifera*, Nature 443, 931–949.
- Wilson E.O. (1971) The insect societies, Harvard University Press, Cambridge, MA.
- Winston M.L. (1987) The biology of the honey bee, Harvard University Press, Cambridge, MA.
- Winston M.L., Michener C.D. (1977) Dual origins of highly social behavior among bees, Proc. Natl. Acad. Sci. USA 74, 1135–1137.

# **Online Material**

## SUPPLEMENTARY METHODS

### 1. Taxonomic sampling

The corbiculate bees are a monophyletic group of long-tongued bees within Apinae, defined most notably by the specialized pollen transporting structures on the hind legs known as pollen baskets or corbicula (Sakagami and Michener, 1987; Roig-Alsina and Michener, 1993; Michener, 2000). The corbiculate bees include four tribes: the solitary or communal Euglossini (orchid bees; 210 species in five genera; Roubik and Hanson, 2004; Ascher, 2007), the primitively eusocial Bombini (bumble bees; 246 species in the genus *Bombus*; Williams, 1998; Ascher, 2007), and the advanced eusocial Meliponini (stingless bees; 471 species in 20–40 or more genera depending on the classification; Michener, 2000; Camargo and Roubik, 2005; Ascher, 2007) and Apini (honey bees; six to 11 species in the genus *Apis*; Engel, 1999; Michener, 2000).

A list of taxa examined in the present study is given in Supplementary Table I. The monophyly of each extant corbiculate tribe is well established both by morphology (Roig-Alsina and Michener, 1993; Michener, 2000) and molecular data (Cameron and Mardulyn, 2001; Michel-Salzat et al., 2004; Arias and Sheppard, 2005; Cameron et al., 2007; Rasmussen and Cameron, 2007), so use of exemplars is appropriate. Because our focus was to sequence more genes rather than to sample more taxa, sampling was limited to two or three exemplars per tribe. However, an effort was made to sample across a greater diversity of genera and species, as corroborated by recent molecular phylogenetic analyses of each tribe (Kawakita et al., 2004; Michel-Salzat et al., 2004; Arias and Sheppard, 2005; Cameron et al., 2007; Rasmussen and Cameron, 2007). Outgroup taxa were sampled from four tribes that are suggested as closely related to the corbiculate bees by morphology (Michener, 1990; Roig-Alsina and Michener, 1993) and a recent molecular phylogenetic analysis (Danforth et al., 2006a).

### 2. Genes analyzed and primers for the polymerase chain reaction (PCR)

We assembled sequences of 12 protein-coding nuclear genes for this study (Supplementary Tab. II). The following seven genes have been previously utilized in bee phylogenetics: arginine kinase (Argk; Kawakita et al., 2003), CAD (Danforth

et al., 2006b), elongation factor 1- $\alpha$  F2 copy (EF-1 $\alpha$ ; Danforth and Ji, 1998), sodium-potassium ATPase (NaK; Danforth et al., unpublished), RNA polymerase II (Pol II; Danforth et al., 2006a,b), long-wavelength rhodopsin (Rh; Mardulyn and Cameron, 1999) and wingless (Wg; Danforth et al., 2004a). The primer sequences have been previously published and those used in the present study are listed in Supplementary Table II. The white (W) gene has been used in insect systematics, e.g., mosquitoes (Besansky and Fahey, 1997); Danforth et al. (2004b) designed the primer for bees. The remaining four genes have not been explored previously for phylogenetic analysis and were newly developed for this study: mitotic checkpoint control protein (Bub3), Ca<sup>2+</sup>/calmodulin-dependent protein kinase II (CamkII), deoxyribonucleoside kinase (Dnk), and glycerol kinase (Gyk). For these four loci, we compared the sequences derived from the genomic databases of *Apis mellifera* and *Drosophila melanogaster* and those contained in the ETS database of *Heliconius* butterflies (Papanicolaou et al., 2005) to identify conserved regions within the exons for designing degenerate primers. Because our aim was to sample more genes but not longer sequences from individual loci, care was taken to keep the length of a PCR fragment, including introns, to about 1000 bp in order to facilitate direct sequencing. Primer sequences and maps showing primer and intron positions for these four genes are provided as Supplementary Figure.

The Argk, EF-1 $\alpha$ , and Rh sequences of the following species have been previously published and were retrieved from the GenBank database: *Euglossa imperialis* (AY267176/AY267144/AY267160), *Exaerete frontalis* (AY267175/AY267143/AY267159), *Bombus ardens* (AF492897/AF492964/AF493031), *Bombus diversus* (AF492894/AF492961/AF493028), *Meliponula bocandei* (AY267177/AY267145/AY267161), *Apis dorsata* (AY267178/AY267146/AY267162), and *Centris cockerelli* (AY267180/AY267148/AY267164). The remaining sequences were newly obtained in this study and deposited in the GenBank database under accession numbers EU184704–EU184857.

### 3. Molecular methods

Genomic DNA was extracted from thoracic tissue of ethanol-preserved or pinned specimens with

**Supplementary Table I.** Taxa examined.

Tribe	Species	Locality
Ingroup		
Euglossini	<i>Euglossa imperialis</i> Cockerell, 1922	Panama
	<i>Exaerete frontalis</i> (Guérin-Méneville, 1845)	Panama
	<i>Eufriesea pulchra</i> (Smith, 1854)	Panama
Bombini	<i>Bombus ardens</i> Smith, 1879	Kyoto, Japan
	<i>Bombus diversus</i> Smith, 1869	Kyoto, Japan
Meliponini	<i>Trigona fuscipennis</i> Friese, 1900	Costa Rica
	<i>Cephalotrigona capitata</i> (Smith, 1854)	Costa Rica
	<i>Meliponula bocandei</i> (Spinola, 1853)	Gabon
Apini	<i>Apis florea</i> Fabricius, 1787	Mahaxai, Laos
	<i>Apis dorsata</i> Fabricius, 1793	Laksao, Laos
	<i>Apis cerana</i> Fabricius, 1793	Kyoto, Japan
Outgroup		
Centridini	<i>Centris cockerelli</i> Fox, 1899	New York, USA
Ericrocidini	<i>Epiclopus gayi</i> Spinola, 1851	Chile
Melectini	<i>Thyreus takaonis</i> (Cockerell, 1911)	Okinawa, Japan
Anthophorini	<i>Anthophora plumipes</i> (Pallas, 1772)	Kyoto, Japan

standard phenol/chloroform methods. PCR amplifications were carried out with the following reaction conditions: initial denaturation at 94 °C for 5 min, 30 cycles of denaturation at 94 °C for 30 sec, annealing for 30 sec, and extension at 72 °C for 1 min, and final extension at 72 °C for 7 min. Annealing temperatures depended on primers used and are provided in Supplementary Table II. All PCR amplifications were aided by Ex Taq polymerase (TaKaRa, Otsu, Japan) and carried out using a GeneAmp PCR System 9700 (Perkin-Elmer, Foster City, CA). Amplified PCR products were purified using a NucleoSpin Extract II Kit (Macherey-Nagel, Düren, Germany). Sequencing reactions were performed using an ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit (Perkin-Elmer), and electrophoresis was conducted on an ABI 3100 sequencer (Perkin-Elmer). The CamkII gene displayed extensive length variation in the first intron; full-length sequences could not be obtained by using only the PCR primers in some taxa. We therefore designed internal sequencing primers located at the boundaries between the first intron and the flanking exons, and used them to obtain sequences of only exons and the second intron.

#### 4. Phylogenetic analysis

Because the intron sequences varied greatly in length among taxa, only exon sequences were used

for analyses. Alignment of the exon sequences was straightforward. Only a single, three-base insertion in the CAD gene of *Thyreus takaonis* was necessary. The number of nucleotide sites for each gene partition varied between 166 and 942 (Supplementary Tab. III), and the combined data matrix consisted of 6018 aligned nucleotide sites. The number of parsimony-informative characters for the corbiculate ingroup ranged between 16 and 196 for each partition, and in total, 1387 sites were informative (Supplementary Tab. III). Within the ingroup, sequences could not be obtained in *Trigona fuscipennis* and *Meliponula bocandei* for CAD, *Trigona fuscipennis* and *Euglossa imperialis* for CamkII, and *Apis cerana* and *Apis florea* for Dnk. In the outgroup, *Epiclopus gayi* could not be sequenced for Dnk, EF-1 $\alpha$ , and Gyk, *Thyreus takaonis* for Argk and EF-1 $\alpha$ , and *Anthophora plumipes* for Argk and Dnk.

All phylogenetic analyses were done using PAUP\* version 4.0b10 (Swofford, 2002), unless otherwise stated. Parsimony, likelihood, and Bayesian analyses for the nucleotide characters were performed with the following procedures. We constructed the most parsimonious trees by heuristic searches with 100 random addition analyses and tree bisection-reconnection (TBR) branch-swapping. Robustness of the trees was validated with non-parametric bootstrap analysis with 1000 replications. For the maximum likelihood analysis, we used the program Modeltest 3.0 (Posada and

**Supplementary Table II.** Primers used for this study. Superscript letters after primer name indicate annealing temperatures; a, 40 °C; b, 45 °C; c, 50 °C; d, 55 °C; e, used for sequencing only. When using primers with different superscript letters, the lower annealing temperature was used.

Gene	Primer name	Strand	Primer sequence (5' to 3')	Reference
Argk	ArgKfor2 <sup>c</sup>	F	GAC AGC AAR TCT CTG CTG AAG AA	Kawakita et al. (2003)
	ArgKrev2 <sup>c</sup>	R	GGT YTT GGC ATC GTT GTG GTA GAT AC	this study
Bub3	Bub3for1 <sup>a</sup>	F	GTC AAT GCH ATM TTV ACK GGH GGH TG	this study
	Bub3f2 <sup>c</sup>	F	GAT CCY AGA ACA CCY ACY TGT GTW GG	
	Bub3rev1 <sup>a</sup>	R	CG YTT YTT RTT RAA RCC ATC CC	
	Bub3r2 <sup>c</sup>	R	GC AAA YGT RTT ATA MGT WGA ATG AA	
CAD	apCADfor1 <sup>c</sup>	F	GGW TAT CCC GTD ATG GCB MGW GC	Danforth et al. (2006b)
	apCADrev1mod <sup>c</sup>	R	GC CAT YRC YTC BCC YAC RCT YTT CAT	this study
CamkII	CamkIIP <sup>b</sup>	F	TTY GCY GGG ACR CCM GGT TAY CT	
	CamK_int_F <sup>e</sup>	F	G CAG TAY CCR AGY CCR GAA TG	
Dnk	CamkIIR <sup>b</sup>	R	TT GCG CCT YGC RIT RAA TTT CT	this study
	CamK_int_R <sup>e</sup>	R	CT TAC GTC GTA YGW TCC	
EF-1 $\alpha$	dNKf1 <sup>c</sup>	F	GAR GGY AAY ATM GGH AGC GGK AAR AC	this study
	dNKf2 <sup>c</sup>	F	TAS AAG AAC CAR TAG AAC TTT GGC G	
	dNKr1 <sup>c</sup>	R	AG CCA KTC MTC RTG MAT MTT RTG MAG YT	
	dNKr2 <sup>c</sup>	R	T TTC TTC TTT YCT YGC TCT KGY TT	
Gyk	ef1aF2 <sup>pl</sup>	F	GGa CAC AGA GAT TTC ATC AAR AA	Kawakita et al. (2003)
	ef1aF2 <sup>pd</sup>	R	TTG CAA AGC TTC RTG RTG CAT TT	
NaK	GlyK-F <sup>d</sup>	F	ACG GGY GGS AAR GAC GGT GG	this study
	GlyKF2 <sup>c</sup>	F	TGG GAT CCC ACV TTR TGC AGA TAT TT	
	GlyK-R <sup>d</sup>	R	AA AGC TGG YAC RAA RTA TAC GTC	
	GlyKr2 <sup>c</sup>	R	AG TTT CAT ATT ATC TCG YAR CCA	
Pol II	NaKfor2 <sup>d</sup>	F	GCS TTC TTC TCB ACS AAC GCC GTY GAR GG	Danforth et al. (2004b)
	NaKrev2 <sup>d</sup>	R	AC CTT GAT RCC GGC YGA WCG GCA CTT GGC	Danforth et al. (2006b)
Rh	polfor2a <sup>c</sup>	F	AAY AAR CCV GTY ATG GGT ATT GTR CA	
	polrev2a <sup>c</sup>	R	AG RTA NGA RIT CTC RAC GAA TCC TCT	
W	LWRhF <sup>c</sup>	F	AAT TGC TAT TAY GAR ACN TGG GT	Danforth et al. (2004a)
	LWRhR <sup>c</sup>	R	ATA TGG AGT CCA NGC CAT RAA CCA	
Wg	LwRhFor3 <sup>c</sup>	F	AGA TAC AAC GTR ATC GTS AAR GGT	Danforth et al. (2004b)
	whiterev2 <sup>c</sup>	F	GCT GAR GGW AGA GTA GCY TTC ATG GG	
Wg	LepWG1a <sup>b</sup>	R	C SGC RAA RAC GTT CTG GAA RGT CAT ATT	Brower and DeSalle (1998)
	LepWG2a <sup>b</sup>	R	GAR TGY AAR TGY CAY GGY ATG TCT GG	
	beewgFor <sup>b</sup>	F	AC TIC GCA RCA CCA RTG GAA TGT RCA	Danforth et al. (2004a)
		F	TGC ACN GTS AAG ACC TGY TGG ATG AG	

**Supplementary Table III.** Summary of basic sequence statistics and phylogenetic results (analyses were done without the outgroups). Likelihood bootstrap values and partitioned Bremer support (PBS) are provided for the node supporting Bombini + Meliponini.

Partition	Total		Informative		A + T		Likelihood		PBS/ min steps <sup>c</sup>	Relative rate <sup>d</sup>	Alpha <sup>d</sup>	PI <sup>d</sup>
	sites	sites	sites	(%) <sup>a</sup>	CI <sup>b</sup>	bootstrap	PBS	bootstrap				
Argk	530	119		52.6	0.641	32	2	0.012	1.122	0.337	0.583	
Bub3	413	38		64.2	0.754	30	1	0.016	0.453	0.096	0.746	
CAD	594	136		54.0	0.719	96	9	0.037	1.455	0.272	0.605	
CamkII	332	16		44.2	0.696	19	-1	-0.024	0.368	1.492	0.743	
Dnk	428	94		69.8	0.781	69	6	0.035	1.258	0.703	0.435	
EF-1 $\alpha$	529	81		56.6	0.660	53	5	0.037	0.806	0.218	0.657	
Gyk	422	79		62.2	0.767	11	-1	-0.009	0.747	0.193	0.663	
NaK	942	196		44.8	0.649	98	9	0.027	1.417	0.109	0.720	
Pol II	766	148		62.8	0.696	98	9	0.048	0.822	0.168	0.689	
Rh	502	117		56.5	0.714	63	-3	-0.016	1.128	0.349	0.570	
W	166	45		50.9	0.662	27	0	0.000	1.179	0.335	0.591	
Wg	394	45		38.9	0.740	28	-5	0.068	0.618	0.062	0.765	

<sup>a</sup> Average proportion of A's and T's across sequences of the ingroup taxa.

<sup>b</sup> Consistency index excluding uninformative characters.

<sup>c</sup> Partitioned Bremer support divided by minimum number of steps for each partition.

<sup>d</sup> Relative substitution rate, shape parameter of the gamma distribution, and proportion of invariable sites as estimated from combined Bayesian analysis.

**Supplementary Table IV.** Comparison of base compositions among the ingroup taxa for the combined 12-gene data set. There is no significant heterogeneity in base composition as judged by the chi-square test ( $\chi^2 = 25.0096$ ,  $df = 30$ ,  $P = 0.7246$ ).

Taxon	Sites	A	C	G	T
<i>Euglossa imperialis</i>	5620	0.2886	0.2155	0.2354	0.2605
<i>Exaerete frontalis</i>	5786	0.2814	0.2157	0.2421	0.2608
<i>Eufriesea pulchra</i>	6003	0.2909	0.2096	0.2351	0.2645
<i>Bombus ardens</i>	5896	0.2808	0.2144	0.2388	0.2660
<i>Bombus diversus</i>	5841	0.2851	0.2111	0.2371	0.2668
<i>Trigona fuscipennis</i>	5084	0.2797	0.2182	0.2469	0.2552
<i>Cephalotrigona capitata</i>	5944	0.2784	0.2221	0.2470	0.2525
<i>Meliponula bocandei</i>	5379	0.2814	0.2172	0.2460	0.2554
<i>Apis florea</i>	5420	0.2832	0.2137	0.2391	0.2639
<i>Apis dorsata</i>	5820	0.2944	0.2049	0.2298	0.2709
<i>Apis cerana</i>	5568	0.2827	0.2132	0.2378	0.2664

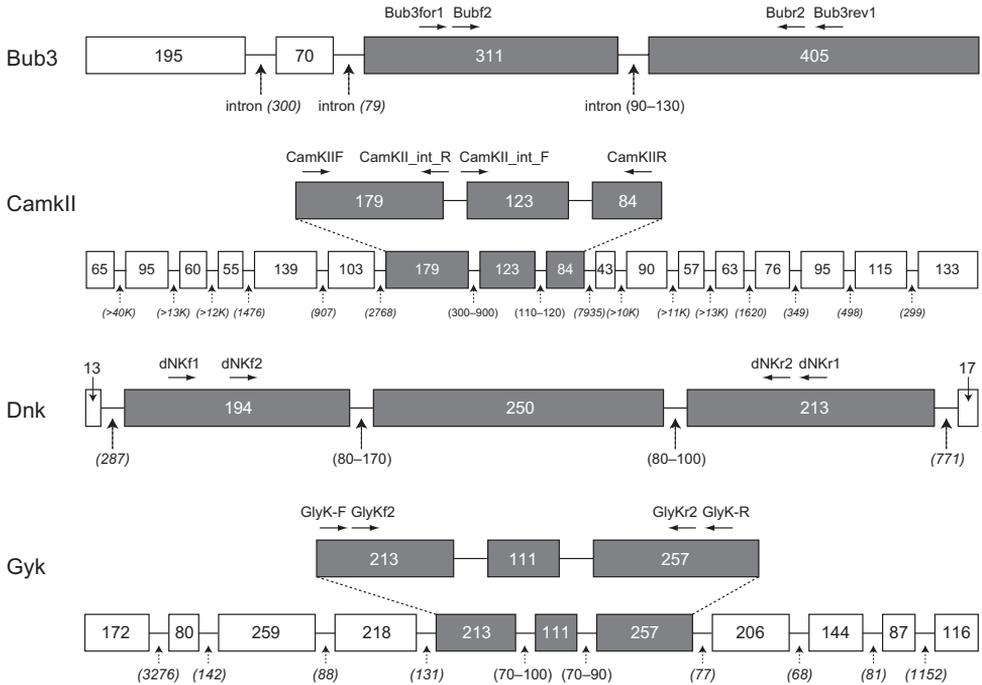
Crandall, 1998) to select an appropriate model of base substitution and to estimate model parameters, which were used in maximum likelihood heuristic searches with 10 random addition analyses and TBR branch-swapping. Likelihood node support was assessed by performing 100 non-parametric bootstrap replications. The Bayesian analysis was performed using MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003) and appropriate substitution models selected by MrModeltest 2.2 (Nylander, 2004). In the simultaneous analysis of the 12 genes, we partitioned the data set by genes, and substitution models were selected separately by using MrModeltest. The unlink command was used to obtain parameter estimates separately for each partition. In the Bayesian tree searches, we ran four simultaneous chains for  $5 \times 10^5$  generations, saving trees every 100 generations for a total of 5001 trees. We plotted ln-likelihood of the sampled trees against generation time to identify the region of the analysis in which the parameter estimates were stable. We discarded the burn-in region (trees and parameter estimates obtained before equilibrium; the initial 1001 tree), and the remaining 4000 samples were used to obtain the posterior distribution of tree topology, branch lengths, and substitution parameters. To ensure that analyses were not trapped on local optima, we carried out two independent runs and compared the results for consistency.

Parsimony analysis of amino acid sequences was performed using PAUP\* as described for nucleotide

sequences. The best-fit model of amino acid sequence for the likelihood analysis was selected using Modelgenerator (Keane et al., 2006), and the likelihood search and 100 replications of bootstrap analysis were performed with MultiPhyl (Keane et al., 2007) using the nearest neighbor interchange (NNI) swapping algorithm. The Bayesian search was performed using MrBayes and by setting the prior for the amino acid model to “mixed” under which the program estimates the relative contribution of different amino acid models according to their posterior probabilities. The shape parameter of the gamma distribution and proportion of invariable sites were also estimated as suggested by Modelgenerator.

#### 4.3. Analysis of correlation between nodal support values and statistical properties of individual gene partitions

Because individual gene partitions supported variable tribal relationships, we compared nodal support values provided by each gene for the Bombini + Meliponini group, and tested whether the strength of support for this topology was correlated with the following six parameters: number of informative sites, base composition (proportion of A + T), consistency index (CI; excluding uninformative characters), relative substitution rate, shape parameter of the gamma distribution



**Supplementary Figure.** Gene map showing positions of exons (boxes) and introns (lines), and the primers newly designed in this study. Exons are proportional to their lengths, and the number of nucleotide sites is given for each exon. Regions sequenced in this study are indicated in gray. Approximate lengths of introns in species examined in this study are given for the regions sequenced. Lengths for other introns are those of *Apis mellifera*.

( $\alpha$ ), and the proportion of invariable sites (PI). The first three parameters were calculated using PAUP\*. The remaining three parameters were estimated within the Bayesian framework. The advantage of the Bayesian approach is that various substitution parameters can be estimated (without assuming a particular tree topology a priori) and compared across different partitions when analyzed simultaneously using the same model (Lin and Danforth, 2003; Danforth et al., 2005). We first estimated relative substitution rate and  $\alpha$  by employing the GTR+G model to all gene partitions and performing Bayesian searches. Parameter estimates were obtained separately for each gene by using the unlink command. We then obtained estimates of PI using the GTR+I model across all genes. We did not estimate  $\alpha$  and PI simultaneously (GTR+G+I model) because the two parameters are dependent on each other and thus would more appropriately be compared across partitions when estimated separately.

The following four support indices were used to assess the correlation between the above parameters and phylogenetic performance: parsimony and likelihood bootstrap support, Bayesian posterior probability, and partitioned Bremer support (PBS; Bremer, 1994). The command file for calculating PBS was produced using TreeRot version 2 (Sorenson, 1999). Because PBS is dependent on the amount of signal content of a given partition, we divided PBS by the minimum number of steps for that gene (Baker et al., 2001) to obtain a standardized comparative measure of relative contribution of each gene to the node of interest.

## REFERENCES

- Arias M.C., Sheppard W.S. (2005) Phylogenetic relationships of honey bees (Hymenoptera: Apinae: Apini) inferred from nuclear and mitochondrial DNA sequence data, *Mol. Phylogenet. Evol.* 37, 25–35.

- Ascher J.S. (2007) Apoidea species, Discover Life, [online] <http://stri.discoverlife.org/mp/20q?search=apoidea> (accessed on 18 Sep 2007).
- Baker R.H., Wilkinson G.S., Desalle R. (2001) Phylogenetic utility of different types of data used to infer evolutionary relationships among stalk-eyed flies (Diopsidae), *Syst. Biol.* 50, 87–105.
- Besansky N.J., Fahey G.T. (1997) Utility of *white* gene in estimating phylogenetic relationships among mosquitoes (Diptera: Culicidae), *Mol. Biol. Evol.* 14, 442–454.
- Bremer K. (1994) Branch support and tree stability, *Cladistics* 10, 295–304.
- 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.
- Camargo J.M.F., Roubik D.W. (2005) Neotropical Meliponini: *Paratrigonoides mayri*, new genus and species from western Colombia (Hymenoptera, Apidae, Apinae) and phylogeny of related genera, *Zootaxa* 1081, 33–46.
- Cameron S.A., Mardulyn P. (2001) Multiple molecular data sets suggest independent origins of highly eusocial behavior in bees (Hymenoptera: Apinae), *Syst. Biol.* 50, 194–214.
- Cameron S.A., Hines H.M., Williams P.H. (2007) A comprehensive phylogeny of the bumble bees (*Bombus*), *Biol. J. Linn. Soc.* 91, 161–188.
- Danforth B.N., Ji S. (1998) Elongation factor-1 $\alpha$  occurs as two copies in bees: implications for phylogenetic analysis of EF-1 $\alpha$  sequences in insects, *Mol. Biol. Evol.* 15, 225–235.
- Danforth B.N., Brady S.G., Sipes S., Pearson A. (2004a) Single-copy nuclear genes recover Cretaceous-age divergences in bees, *Syst. Biol.* 53, 309–326.
- Danforth B.N., Fang J., Sipes S., Brady S.G., Almeida E. (2004b) Phylogeny and molecular systematics of bees (Hymenoptera: Apoidea), Cornell University, Ithaca, NY, [online] <http://www.entomology.cornell.edu/BeePhylogeny/>.
- Danforth B.N., Lin C.P., Fang J. (2005) How do insect nuclear ribosomal genes compare to protein-coding genes in phylogenetic utility and nucleotide substitution patterns? *Syst. Entomol.* 30, 549–562.
- Danforth B.N., Sipes S., Fang J., Brady S.G. (2006a) The history of early bee diversification based on five genes plus morphology, *Proc. Natl. Acad. Sci. USA* 103, 15118–15123.
- Danforth B.N., Fang J., Sipes S. (2006b) 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.
- Engel M.S. (1999) The taxonomy of recent and fossil honey bees (Hymenoptera: Apidae; *Apis*), *J. Hym. Res.* 8, 165–196.
- Kawakita A., Sota T., Ascher J.S., Ito M., Tanaka H., Kato M. (2003) Evolution and phylogenetic utility of alignment gaps within intron sequences of three nuclear genes in bumble bees (*Bombus*), *Mol. Biol. Evol.* 20, 87–92.
- Kawakita A., Sota T., Ito M., Ascher J.S., Tanaka H., Kato M., Roubik D.W. (2004) Phylogeny, historical biogeography, and character evolution in bumble bees (*Bombus*: Apidae) based on simultaneous analysis of three nuclear genes sequences, *Mol. Phylogenet. Evol.* 31, 799–804.
- Keane T.M., Creevey C.J., Pentony M.M., Naughton T.J., McInerney J.O. (2006) Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified, *BMC Evol. Biol.* 6, 29.
- Keane T.M., Naughton T.J., McInerney J.O. (2007) MultiPhyl: A high-throughput phylogenomics webserver using distributed computing, *Nucleic Acids Res.* 35, W33–W37.
- Lin C.P., Danforth B.N. (2003) How do insect nuclear and mitochondrial gene substitution patterns differ? Insights from Bayesian analysis of combined datasets, *Mol. Phylogenet. Evol.* 30, 686–702.
- Mardulyn P., Cameron S.A. (1999) The major opsin in bees (Insecta: Hymenoptera): a promising nuclear gene for higher level phylogenetics, *Mol. Phylogenet. Evol.* 12, 168–176.
- Michel-Salzat A., Cameron S.A., Oliveira M.L. (2004) Phylogeny of the orchid bees (Hymenoptera: Apinae: Euglossini): DNA and morphology yield equivalent patterns, *Mol. Phylogenet. Evol.* 32, 309–323.
- Michener C.D. (2000) *The bees of the world*, John Hopkins University Press, Baltimore, MD, pp. xlv + 913.
- Nylander J.A.A. (2004) MrModeltest, version 2, Uppsala University, Uppsala, Sweden.
- Papanicolaou A., Joron M., McMillan O., Blaxter M.L., Jiggins C.D. (2005) Genomic tools and cDNA derived markers for butterflies, *Mol. Ecol.* 14, 2883–2897.
- Posada D., Crandall K.A. (1998) Modeltest; testing the model of DNA substitution, *Bioinformatics* 14, 817–818.
- Rasmussen C., Cameron S.A. (2007) A molecular phylogeny of the Old World stingless bees (Hymenoptera: Apidae: Meliponini) and the non-monophyly of the large genus *Trigona*, *Syst. Entomol.* 32, 26–39.
- Roig-Alsina A., Michener C.D. (1993) Studies of the phylogeny and classification of long-tongued bees

- (Hymenoptera: Apoidea), Univ. Kans. Sci. Bull. 55, 124–162.
- Ronquist F., Huelsenbeck J.P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models, *Bioinformatics* 19, 1572–1574.
- Roubik D.W., Hanson P.E. (2004) Orchid bees of tropical America. Biology and field guide, Instituto Nacional de Biodiversidad, Santo Domingo de Heredia, Costa Rica, p. 370.
- Sakagami S.F., Michener C.D. (1987) Tribes of Xylocopinae and origin of the Apidae (Hymenoptera: Apoidea), *Ann. Entomol. Soc. Am.* 80, 439–450.
- Sorenson M.D. (1999) TreeRot, version 2, Boston University, Boston, MA.
- Swofford D.L. (2002) PAUP\*: phylogenetic analysis using parsimony (\*and other methods), version 4.0b10, Sinauer Associates, Sunderland, MA.
- Williams P.H. (1998) An annotated checklist of bumblebees with an analysis of patterns of description (Hymenoptera: Apidae, Bombini), *Bull. Nat. Hist. Mus. Lond. (Entomol.)* 67, 79–152.