Comprehensive phylogeny, biogeography and new classification of the diverse bee tribe Megachilini: can we use DNA barcodes in phylogenies of large genera?

V. Trunza^a, L. Packer^b, J. Vieu^{a, c}, N. Arrigo^d and C. J. Praz^{a*}

^aInstitute of Biology, University of Neuchatel, Emile-Argand 11, 2000 Neuchatel, Switzerland

^bDepartment of Biology, York University, 4700 Keele St., Toronto, ON M3J 1P3, Canada

^cPresent address: Plant Ecology, University of Basel, Hebelstrasse 1, Basel, Switzerland

^dDepartment of Ecology and Evolution, University of Lausanne, Biophore, 1015 Lausanne, Switzerland

* author for correspondence: christophe.praz@unine.ch

Abstract

Classification and evolutionary studies of particularly speciose clades pose important challenges, as phylogenetic analyses typically sample a small proportion of the existing diversity. We examine here one of the largest bee genera, the genus *Megachile* - the dauber and leafcutting bees. Besides presenting a phylogeny based on five nuclear genes (5480 aligned nucleotide positions), we attempt to use the phylogenetic signal of mitochondrial DNA barcodes, which are rapidly accumulating and already include a substantial proportion of the known species diversity in the genus. We used barcodes in two ways: first, to identify particularly divergent lineages and thus to guide taxon sampling in our nuclear phylogeny; second, to augment taxon sampling by combining nuclear markers (as backbone for ancient divergences) with DNA barcodes. Our results indicate that DNA barcodes bear phylogenetic signal limited to very recent divergences (3-4 my before present). Sampling within clades of very closely related species may be augmented using this technique, but our results also suggest statistically supported, but incongruent placements of some taxa. However, the addition of one single nuclear gene (LW-rhodopsin) to the DNA barcode data was enough to recover meaningful placement with high clade support values for nodes up to 15 million years old. We discuss different proposals for the generic classification of the tribe Megachilini. Finding a classification that is both in agreement with our phylogenetic hypotheses and practical in terms of diagnosability is particularly challenging as our analyses recover several well-supported clades that include morphologically heterogeneous lineages. We favour a classification that recognizes seven morphologically well-delimited genera in Megachilini: *Coelioxys*, *Gronoceras*, *Heriadopsis*, *Matangapis*, *Megachile*, *Noteriades* and *Radoszkowskiana*. Our results also lead to the following classification changes: the groups known as *Dinavis*, *Neglectella*, *Eurymella* and *Phaenosarus* are recognized as valid subgenera of the genus *Megachile* (new statuses), while the subgenus *Alocanthodon* is placed in synonymy with *M. (Callomegachile)*, the subgenera *Parachalicodoma* and *Largella* with *M. (Pseudomegachile)*, *Anodonteutricharaea* with *M. (Paracella)*, *Platysta* with *M. (Eurymella)*, and *Grosapis* and *Eumegachile* with *M. (Megachile)* (new synonymies). In addition, we use maximum likelihood reconstructions of ancestral geographic ranges to infer the origin of the tribe and reconstruct the main dispersal routes explaining the current, cosmopolitan distribution of this genus.

Key words: bees, *Megachile*, *Coelioxys*, species-level phylogeny, phylogenetic signal, taxon sampling.

1. Introduction

Molecular phylogenies are available for a number of bee clades (reviewed in Danforth et al., 2013) and have contributed to the resolution of many long-standing controversies in bee systematics (eg. Danforth and Ji, 2001; Praz et al., 2008; Almeida and Danforth, 2009; Cardinal et al., 2010; Rasmussen and Cameron, 2010; Praz and Packer, 2014; Romiguier et al., 2016). Together, these studies are contributing to the development of stable classifications and provide a framework for examining patterns of diversification (e.g., Hines, 2008; Litman et al., 2011, 2013), plant-bee coevolution (e.g., Sedivy et al., 2008), social evolution (e.g., Schwarz et al. 2007) and historical biogeography (e.g., Almeida et al., 2011; Chenoweth and Schwarz, 2011; Praz and Packer, 2014). For a few large bee clades, including most of the largest bee genera, however, no phylogenetic hypothesis is yet available, and classifications remain largely conjectural. Nearly ten years ago, Michener (2007: 120) listed five large complexes of problematic taxa, "for which the current generic classification is arbitrary and will probably be revised in the near future". Of these five groups, three have been examined using molecular markers (Praz et al., 2008 for the osmiine complex; Almeida and Danforth, 2009 for the *Leioproctus* and *Lonchopria*-group; and, among others, Gibbs et al., 2012 for the various lineages of the genus *Lasioglossum*). The present paper examines a fourth problematic group, the bees currently placed in the genus *Megachile* in the tribe Megachilini.

Four genera are currently recognized in this speciose tribe: *Coelioxys*, *Megachile*, *Noteriades* and *Radoszkowskiana*. The genus *Noteriades* includes only a few little-known African and Oriental species (Griswold and Gonzalez, 2011). *Coelioxys* and *Radoszkowskiana* are cleptoparasites, principally of species belonging to the genus *Megachile* (Michener, 2007). The former is distributed worldwide and includes more than 400 species (Ascher and Pickering, 2016; see Correia and Packer, submitted, for a subgeneric treatment of this genus), while the latter includes only four species limited to the Palearctic (Schwarz, 2001). With more than 1400 species (Ascher and Pickering, 2016), the cosmopolitan genus *Megachile* is among the largest bee genera worldwide and represents a significant proportion of most bee faunas, from tropical to temperate regions. Unlike most bee lineages, *Megachile* have their maximal diversity in tropical regions. Numerous species cut leaf discs that they use to line their brood cells in the

ground or in existing cavities, and are commonly referred to as the leafcutter bees. Other *Megachile* species do not cut leaves but rather build brood cells with resin or mud mixed with salivary secretions (Kronenberg and Hefetz, 1984) in cavities, in the ground, or sometimes exposed on the surfaces of stone or wood (references in Michener, 2007). While some authors have referred to these species as either resin or mason bees (e. g., O'Toole and Raw, 1991), we find these terms confusing as both are also used for Anthidiini and Osmiini, respectively. We prefer to follow Eardley (2012) and refer to these species as "dauber" bees.

The relationships among the main megachiline lineages remain largely unclear, with the exception of the genus *Noteriades*, which has been demonstrated to be the sister group to all other Megachilini based on analyses of molecular (Praz et al., 2008; Litman et al., 2011) and morphological data (Gonzalez et al., 2012). The phylogenetic position of the two cleptoparasitic genera *Coelioxys* and *Radoszkowskiana* is unclear both in terms of whether they are derived from within the genus *Megachile* (Litman et al., 2011, 2013) and whether they are sister taxa (Litman et al. 2011, 2013; Correia and Packer, submitted). Thus, it remains unknown whether cleptoparasitism has evolved once or twice within the tribe (Rozen and Kamel, 2007, 2008; Litman et al., 2013).

Michener (2007) recognizes 56 subgenera within *Megachile*. While acknowledging that the diversity observed in this genus is larger than that seen in other tribes that are divided into numerous genera (e.g. the Eucerini, Osmiini and Anthidiini), Michener refrains from recognizing distinct genera because of the morphological intergradation among the diverse lineages. Instead, he assembles the subgenera in three groups, which correspond to the genera recognized by some authors, and broadly mirror the species' nesting biology. Michener's Group 1 includes the leafcutter species, in which the female mandible mostly has a "cutting edge", a blade-like structure between some mandibular teeth. Group 2 is made of the dauber lineages, which mostly lack mandibular cutting edges and use resin or mud to build their brood cells. Group 3 only includes the subgenus *Creightonella*, whose species exhibit an intermediate morphology between the other two groups and use a combination of resin or mud and leaf discs as nesting materials. The morphological separation of these three groups is difficult (Michener, 2007): Group 2 is likely a paraphyletic assemblage from which Group 1 and possibly the

cleptoparasitic genera arose (Litman et al., 2011); and distinction between *Creightonella* and the other groups is "about as weak as that between Groups 1 and 2" (Michener, 2007: 554). Nesting biology and associated mandibular structure (presence or absence of cutting edges) do not always differ clearly by group. Most members of Groups 1 and 3 are leafcutters but cutting edges are reduced or absent in some lineages of Group 1; moreover, some subgenera included in Michener's Group 2 have distinct cutting edges, such as *Chelostomoda* and *Mitchellapis*. At least the former uses leaf discs to close its nests (Iwata, 1976). Other morphological characters used to segregate these three groups are equally ambiguous (Michener, 2007: 556).

Group 1, 2 and 3 correspond to the genera *Megachile*, *Chalicodoma* and *Creightonella* of some authors (Michener, 1965; Pasteels, 1965; Tkalců, 1969). In addition, some group 2 subgenera such as *Gronoceras* and *Thaumatossoma* have sometimes been given generic rank (e. g., Cockerell, 1935; King and Exley, 1985). The proposal of Mitchell (1980) to further divide Michener's group 1 into five genera does not appear practical (Michener, 2007: 555) and is not further discussed here. Michener (2007) suggested splitting his heterogeneous Group 2 into the following five genera: *Matangapis*, *Mitchellapis*, *Megella*, *Chelostomoides* ("including [as subgenera] *Callomegachile*, perhaps *Gronoceras* and *Thaumatossoma*"), and *Chalicodoma*. Based on cladistic analyses of morphological characters, Gonzalez (2008) (see also Engel and Gonzalez, 2011 and Gonzalez and Engel, 2012) suggested a classification that breaks up Group 2 into three genera: *Matangapis*, *Chalicodoma* (including as subgenera, among others, *Pseudomegachile*, *Gronoceras* and *Callomegachile*) and *Thaumatossoma* (including the heriadiform subgenera, such as *Hackeriapis*, *Chelostomoides* and *Maximegachile*).

The purpose of the present study is to unravel the relationships within Megachilini, especially among the subgenera of the diverse genus *Megachile*, lay the foundations for a sound classification of the group and obtain insights into their biogeographic history. We sequenced five nuclear genes for more than 100 species of *Megachile* representing most subgenera, members of all other genera of Megachilini and representatives of all other megachiline tribes. To further refine our taxon sampling in the genus *Megachile*, we also use information from trees based on DNA-barcodes (the 658 bp fragment of the mitochondrial gene cytochrome oxidase I; Hebert et al., 2003) to pinpoint particularly

divergent species and to break-up long branches within subgenera. We examine the phylogenetic signal of DNA barcodes and evaluate the possibility of combining our nuclear dataset with DNA barcodes (see Kjer et al., 2014, for a similar approach applied to the caddisfly genus *Chimarra*). We use our nuclear dataset as a backbone to infer old divergences (e.g., between subgenera and species-groups) and DNA barcodes to augment taxon sampling within more recent clades. We discuss the potential benefits and pitfalls of using DNA barcodes in combination with nuclear genes to reconstruct densely sampled phylogenies.

2. Materials and methods

2.1. Taxon sampling

In total, 127 species were included in our phylogenetic analyses of nuclear genes (Table S1). As outgroup taxa, we included representatives of the subfamilies Lithurginae and Pararhophitinae (following the classification of Gonzalez et al., 2012), as well as representatives of all tribes of the subfamily Megachilinae (Aspidosmiini, Dioxyini, Anthidiini, Osmiini) and of the genera *Ochreriades* and *Afroheriades* because of their ambiguous position within Megachilinae (Praz et al., 2008; Litman et al., 2011). Within Megachilini, we included one species each of the genera *Noteriades* and *Radoszkowskiana*, eight species of the genus *Coelioxys*, and 106 species of the genus *Megachile*. In addition, we obtained sequences of one nuclear gene (28S) using next generation sequencing (see below) for two important subgenera (*Matangapis* and *Heriadopsis*). Of the 58 subgenera of *Megachile sensu lato* currently recognized (Michener, 2007, with updates by Durante and Cabrera, 2009; Engel and Baker, 2006; Engel and Gonzalez, 2011; Gonzalez et al., 2010; Gonzalez and Engel, 2012; Raw, 2006), eleven could not be included: *Austrosarus*, *Dactylomegachile*, *Neocressoniella*, *Platysta*, *Ptilosaroides*, *Rhyssomegachile*, *Schrottkyapis*, *Trichurochile* and *Zonomegachile* in Michener's group 1 and *Cestella* and *Schizomegachile* in Michener's group 2. For subgenera present on several continents, we included members from each biogeographic area whenever possible.

In addition, we used trees based on 1783 mitochondrial DNA barcodes originating from all continents to further guide our taxon selection. The majority of these barcodes were generated at the Canadian Centre for DNA barcoding with material from the Packer bee collection; we also used all published barcodes from the BOLD platform (extracted January 2016; data from Sheffield et al., 2009; Magnacca and Brown, 2012; Schmidt et al., 2015). All barcodes used are available on BOLD (www.boldsystems.org). We retrieved these DNA barcodes, produced NJ trees using the BOLD platform and selected one barcode per BIN (a proxy for biological species: Ratnasingham and Hebert, 2013), favouring the longest and best quality sequences and ignoring sequences less than 500 bp. In total 460 DNA barcodes were retained and aligned. The resulting barcode matrix was used to produce trees to augment our taxon sampling. Preliminary analyses of this matrix using maximum likelihood and Bayesian inference were particularly time consuming and yielded trees with very low support values for most nodes; we therefore used neighbour joining trees to guide our taxon sampling. These trees were a surprisingly good match with the existing subgeneric classification: most existing subgenera formed single clusters. However, several large subgenera were divided among clusters. We ensured that each of these clades was represented by at least one taxon in our nuclear matrix. In total, 20 species were added to our initial dataset based on their position in trees inferred from DNA barcodes only: four species of the subgenus *Callomegachile* (taxa with voucher numbers 162, 164, 193, 319; see Table S1 and Fig. 1); three species each of *Megachile* s. str. (40, 339, 750) and *Xanthosarus* (427, 741, 744); two species each of *Chelostomoides* (202, 786), *Eutricharaea* (318, 645) and *Pseudomegachile* (24, 1134); and one species each of the subgenera *Aethomegachile* (168), *Chalicodoma* (43), *Leptorachis* (205) and *Sayapis* (199).

2.2. DNA sequencing

DNA was extracted from one leg, and the rest of the specimen was preserved as a voucher. Voucher specimens are deposited in the Praz Collection at the University of Neuchatel and in the Packer collection at York University, except if stated otherwise (Table S1). We obtained DNA using phenol-chloroform extractions (following Praz et al., 2008) or DNA extraction kits (Nucleospin tissue, Macherey-Nagel). PCR-reactions were performed with Hotstart GoTaq polymerase (Promega) in a Biometra T1 thermocycler

following standard protocols (Praz et al., 2008), with a blank sample as a negative control. PCR products were examined visually using agarose gel electrophoresis and purified enzymatically with a mix of exonuclease and FastAP thermosensitive alkaline phosphatase (Fermentas). Sequencing reactions were performed using BigDye Terminator v3.1 technologies (Applied Biosystems). Sequencing products were purified with Sephadex gel columns (GE Healthcare Life Sciences) and analysed using an ABI-3130 sequencer at the Genetic Diversity Centre at ETH Zürich.

2.3. Genes analysed

We sequenced fragments of five nuclear genes used in previous studies of Megachilidae (Litman et al., 2011): the four protein-coding genes elongation factor 1- α (hereafter EF), LW-rhodopsin (Opsin), conserved ATPase domain (CAD), sodium potassium adenosine triphosphatase (NAK), and the ribosomal gene 28S. To amplify each of these genes, we used the primer sites selected by Litman et al. (2011) and slightly modified each internal primer to optimize amplification in Megachilini based on available sequences from Litman et al. (2011). Primer sequences are given in Table S2. For the amplification of the 658 bp DNA barcoding fragment of COX1 (hereafter COX1), we used the primers LepF1 and LepR1 (Hebert et al., 2004) with standard protocols from the Canadian Centre for DNA Barcoding (<http://www.ccdb.ca/resources.php>). We used next-generation sequencing techniques to obtain sequences of 28S for two taxa for which fresh specimens were not available: *Megachile (Matangapis) alticola* (one specimen collected in 1992) and an undescribed species of *M. (Heriadopsis)* (one specimen collected in 2008, with highly degraded DNA). Both are important subgenera as they are the only species of *Megachile* having an arolium on all or some legs (Michener, 2007). The library preparation was made with a custom protocol (modified from Bronner et al., 2014) derived from the standard Illumina protocol. Details on the lab protocol and on the bioinformatics are given in the supplementary material.

2.4. Editing and aligning of sequences

Chromatograms were trimmed, assembled and edited using Geneious R6 (Kearse et al., 2012). Alignments were performed in Mafft (Katoh and Standley, 2013). 28S sequences

were aligned with consideration to secondary structure using the iterative refinement method Q-INS-i, implemented in Mafft (Kato and Toh, 2008). Alignments for each gene were examined visually and corrected when necessary in Geneious. The coding sequence of each protein-coding gene was converted to amino-acid sequence to ensure that no stop codons were found. Introns of EF and Opsin were relatively conserved and were aligned using Mafft and corrected manually in Geneious; ambiguously aligned regions were removed. The small intron of CAD, however, proved too variable to be aligned and was excluded. Each gene fragment was tested for heterogeneous base composition using the Chi-square test implemented in a beta version of Paup 4.0 (Swofford, 2002) kindly made available by D. Swofford. For protein-coding genes, the three nucleotide positions were tested separately. The third codon positions of NAK (hereafter NAK3) and COX1 were significantly heterogeneous, so we performed analyses both with and without these partitions. The different genes were concatenated into a single matrix for final analyses using Geneious.

2.5. Maximum likelihood analyses

Single gene analyses

We first performed analyses of each gene separately using maximum likelihood (ML) inference. Protein-coding genes were separated into three or four partitions corresponding to the three nucleotide positions and the introns, if applicable. ML-analyses were performed in RAxML version 8 (Stamatakis, 2014) on the CIPRES server (Miller et al., 2010), performing 1000 bootstrap replicates and applying a GTR model with a gamma distribution (G) to each partition. For all analyses performed in this study, including Beast analyses, we decided not to use the combination of a proportion of invariant sites and a gamma distribution because both parameters model rate heterogeneity in different ways (see discussion in the RAxML v8.2.X Manual, page 59).

Concatenated dataset

We concatenated the coding sequence of each nuclear gene, the nuclear introns, and COX1, and performed ML analyses on six different matrices: with and without COX1, with and without the introns, with and without significantly heterogeneous partitions. Each dataset was analysed under two different partitioning regimes: by codon position

and following the more complex partitioning regime suggested by PartitionFinder v. 1.1.0 (Lanfear et al., 2012), with 9-11 partitions, depending on whether NAK3, COX1 or the introns were included. Details on these analyses are given in the supplementary material. We repeated each analysis with the 28S sequences of *Matangapis* and *Heriadopsis* added to the concatenated matrix.

2.6. Bayesian analyses

Bayesian analyses were run in BEAST 1.8 (Drummond et al., 2012) with the same six matrices used for the ML analyses; we also applied two distinct partitioning regimes: by codon as in ML analyses, with both introns combined because initial analyses with two intron partitions yielded poor convergence of the treelikelihood parameter for the intron of Opsin; and by gene, with five partitions (or six, if COX1 was included). More complex models, such as the partitioning regime suggested by PartitionFinder (see above), resulted in poor convergence for numerous parameters and were thus not retained. These twelve Beast analyses were run without fossil calibration for 100 million generations. Additional details on these analyses are presented in the supplementary material. Maximum clade credibility trees were computed after removal of an appropriate burn-in (usually 20% of the trees) using TreeAnnotator (Drummond et al., 2012).

2.7. Divergence time estimation and biogeographic reconstruction

We repeated one selected BEAST analysis including two calibration points to estimate divergence times in our phylogeny. Only one fossil can confidently be attributed to Megachilini, *Megachile glaesaria* from Dominican amber (15-20 mya; Engel, 1999). However, its exact phylogenetic placement within Megachilini is unclear (Engel, 1999) and this fossil would at best be used as a stem group calibration point for the subgenus *Chelostomoides* given its similarity to that subgenus. Given the long branch leading to *Chelostomoides*, spanning 8-21 mya in our final dated phylogeny, this fossil is of little use for calibrating our tree. We thus used two calibration points from Cardinal and Danforth (2013): the age of the root of the tree, corresponding to the age of the node uniting Lithurginae, Pararhophitinae and Megachilinae (this node was recovered with maximal

support values in phylogenetic studies of the Megachilidae: Litman et al., 2011) and the age of the node uniting Megachilini + Osmiini; applying a prior to the age of Megachilini was not appropriate given the considerably smaller taxon sampling within this tribe in Cardinal and Danforth (2013). We placed normal priors with the following values on the ages of these nodes: mean 74, stdev 7, initial value 74 for the root; and mean 57, stdev 5, initial value 57 for the node uniting Osmiini + Megachilini. These prior distributions correspond to the means and confidence intervals found for these nodes in analysis 3 of Cardinal and Danforth (2013: Table S6). We also compared analyses with and without an additional calibration point corresponding to trace fossils putatively attributed to leafcutter bees. Fossil leaves bearing circular incisions, hereafter "fossil leafcuttings", have been reported by various authors (reviewed in Wedmann et al., 2009), but attribution to the genus *Megachile* is only hypothetical. These fossil leafcuttings are of various ages, from the Miocene to the Paleocene (Wedmann et al., 2009). We compared analyses with and without a calibration point corresponding to the oldest of these fossils (ignoring the single questionable Paleocene fossil), which originated from Messel deposits (47 mya; Wedmann et al., 2009). We applied a prior distribution (lognormal distribution, mean of 1, Stdev of 1, zero offset of 47 Mya; 95% HPD: 47-61 Mya) to the most recent common ancestor (hereafter MRCA) of all lineages having a cutting edge in the female mandible.

Probabilistic inference of ancestral range reconstruction was performed using Biogeobears (Matzke, 2013, 2014) to infer the centre of origin of the tribe and to examine the main dispersal routes explaining the cosmopolitan distribution of the genus *Megachile*. We ran the BEAST analysis twice independently, each for 100 million generations, computed a maximum clade credibility tree from both runs after removal of an appropriate burn-in, and removed non-Megachilini taxa using the "drop-tip" function of the ape package in R (Paradis et al., 2004). We recognized the following six biogeographical regions: Nearctic (A), Neotropical (B), Afrotropical (C), Palaearctic (D), Oriental (E) and Australian (F), following Figure 1a of Rueda et al (2013) for delimitations of these regions. Maximum range size was set to two and the possible ranges were the following: A, AB, AD, B, C, CD, CE, D, DE, E, EF, F. We allowed the Oriental and Afrotropical regions to be adjacent (see Fig. 2) given the great faunal similarities of these two zones in bees (Michener, 2007), and the fact that numerous

megachiline lineages (e.g., among others, the genus *Noteriades*) are distributed in both regions but are absent from the Palearctic. The coding for each terminal (Fig. 2; see supplementary material) reflects the geographic distribution of the entire subgenus or of the species-group that each terminal represents; a few exceptions are detailed in the supplementary material. We implemented only one time period in the DEC model of Biogeobears (thus with dispersal probabilities constant throughout the tree; see the supplementary material for a justification); dispersal probabilities were set to 1.0 between all adjacent regions. We also explored a model incorporating founder-event speciation (or, speciation through long-distance dispersal; Matzke, 2014), the DEC + J model implemented in Biogeobears. The nests of wood-nesting *Megachile* may be carried over water barriers as has been suggested for other twig-nesting bees (Fuller et al., 2005) and thus long-distance dispersal events between non-adjacent geographic regions are theoretically possible. In the DEC + J analyses, the dispersal probabilities were set to 1.0 between adjacent regions (AB, AD, CD, CE, DE, EF), and 0.1 between non-adjacent regions.

2.8. Phylogenetic signal of DNA barcodes and combined nuclear + barcode analyses

We used our phylogenetic framework to examine the phylogenetic signal of DNA barcodes (COX1) and to evaluate the possibility of augmenting taxon sampling by combining DNA barcodes with a nuclear "backbone" phylogeny. We selected all well-supported (bootstrap support > 95% in all ML analyses, posterior probability 1.0 in all Bayesian analyses), recent clades representing either subgenera, or species-groups within subgenera in our nuclear phylogeny; older clades, that is, clades grouping several subgenera were excluded, with the exception of the genus *Coelioxys*, which was included as one clade. We analysed a matrix including only the DNA barcodes for the species included in the nuclear phylogeny (taxa with missing COX1 sequence were excluded) using both ML and Bayesian inferences (three partitions modelled with a GTR + gamma model for the ML analyses and a TN93 + G model with empirical base frequency for the Bayesian analyses). The support values for these clades were computed by filtering trees in PAUP (either the 1000 bootstrap trees in ML analyses, or all posterior trees after removal of 20% burn-in in Bayesian analyses). To evaluate variation in the phylogenetic signal of COX1 over time, these values were plotted against node age, as determined in

our dating analysis in BEAST. We also examined the support for these clades when the barcodes were used in combination with a nuclear backbone: starting with our complete nuclear + COX1 matrix, we randomly selected one species from each of the clades; for these species, the full nuclear + COX1 matrix was retained; for the other species in these clades, the nuclear matrix was deleted (thus leaving only COX1); for all other taxa (i.e., taxa not included in these clades), the full nuclear + COX1 matrix was retained. The resulting matrix was analysed using ML and Bayesian inference applying our preferred model and partitioning strategy (see below). Again, the support values for these clades were computed by filtering trees in Paup. The same procedure was repeated twice more with the same clades, leaving one nuclear marker (either 28S or LW-Rhodopsin, respectively) in addition to the mitochondrial barcode; both 28S and Rhodopsin have been used as nuclear markers to delineate species in insects (e.g., Blaimer and Fischer, 2013; Kjer et al., 2014).

Lastly, we concatenated our nucleotide matrix (including the three codon positions of COX1) with a matrix including 350 additional mitochondrial DNA barcodes of the genus *Megachile sensu lato*; these 350 barcodes correspond to the 460 barcodes mentioned above (see section entitled "Taxon sampling") minus those already represented in our dataset. Because the focus was on the genus *Megachile*, *Coelioxys* and *Radoszkowskiana* were removed, as well as taxa with missing barcode sequences. This supermatrix was subjected to Bayesian analyses, partitioning the matrix into codon position and using the substitution models mentioned above, except that we enforced a strict clock model for COX1.

3. Results

3.1 Phylogenetic analyses

After sequence editing and alignment, the concatenated, nuclear matrix included 5480 aligned bp (EF: 887, 146 bp in the intron; Opsin: 794, 122 in introns; CAD: 879; NAK: 1488; 28S: 1432). In addition, the barcode fragment of COX1 was 654 bp long. Missing data of all nucleotide sites amounted to 6.9% (Table S1, with Genbank accession numbers). Although trees were little resolved in individual gene analyses, no conflicting

topology was supported by high (>70%) bootstrap support (hereafter BS) values among the different gene trees (introns included). ML-analyses of the six different matrices all yielded highly congruent topologies (Table S3). The exclusion of NAK3 had virtually no impact on tree topology (see details in Table S3), but resulted in a higher number of nodes supported by values above 50% within Megachilini (90 versus 87 nodes). Exclusion of the introns did not significantly alter the topology or the number of nodes with BS > 50% (in both cases 90 nodes), although it resulted in a slight decrease in support values for a majority of the nodes (Table S3). The addition of COX1, either with or without third codon positions resulted in nearly identical trees, although with one node less supported by BS values > 50% (89 nodes in both cases). For analyses of these six matrices, partitioning regime had virtually no effect on tree topology or nodal support values (Table S3). Based on the number of nodes supported, and because we prefer to exclude significantly heterogeneous partitions, our preferred analysis was performed on a dataset with the introns, without both the significantly heterogeneous third positions of NAK and COX1 (see methods, section 2.4, editing and aligning of sequences), with 6 partitions (Fig. 1). Bayesian analyses performed in BEAST yielded trees that were highly similar to those found in ML analyses (Table S3; Fig. S1). Differences between the 12 Bayesian analyses and support values for taxonomically important nodes are given in Table S3.

3.3. Phylogenetic relationships within Megachilini

Important clades that do not have a taxonomic name are numbered from 1-23 for ease of reference (Fig. 1, S1; Table S3). All phylogenetic analyses recovered the tribe Megachilini as a monophyletic group with maximal support values (BS 100%, PP 1.0), with the genus *Noteriades* sister to all other Megachilini (clade 1; Table S3). Outgroup taxa have been removed from the trees shown in all figures. In all analyses, the subgenus *Gronoceras* branched next and was sister to all remaining Megachilini with maximal support values (clade 2; Table S3), including the cleptoparasitic genera *Coelioxys* and *Radoszkowskiana*. Both cleptoparasitic genera formed a monophyletic group (hereafter the cleptoparasitic clade; clade 3) in most analyses (Table S3); support for this clade was moderate to high (BS up to 83% and PP 1.0; Table S3), whereas support for any alternative placement was weak (BS < 50% in ML analyses; PP less than 0.5 in Bayesian

analyses). The cleptoparasitic clade was sister to all other Megachilini (except *Noteriades* and *Gronoceras*) in most analyses (clade 4; Table S3), although support for this relationship was weak. The inclusion of both subgenera of *Megachile* having arolia (*Heriadopsis* and *Matangapis*, represented only by 28S sequences) resulted in slightly different topologies and lower support values for nodes among the early diverging branches of Megachilini: the positions of *Noteriades* and *Gronoceras* were unchanged, but clade 4 was no longer recovered; the subgenera *Heriadopsis* and *Matangapis* were not closely related to any other lineage of *Megachile* but appeared within the cleptoparasitic clade (Fig. S2), although support for this relationship was below 50% (Fig. S2).

After the divergence of the aforementioned groups, we are left with most members of Michener's Group 2, which are the paraphyletic dauber bees from among which the leafcutter bees arose. The dauber bees are made up of a number of clades as follows: **A.** A well-supported clade (clade 5) comprising several morphologically heterogeneous lineages: one Asian lineage (clade 6; Table S3) which included taxa currently placed in the subgenus *Callomegachile* (the *ornata*-species group) and a currently unnamed species (*M. monoceros* Friese 1903, preoccupied name); both lineages were sister groups in all analyses (clade 6; Table S3); the group of species known as *Carinula*, currently considered to belong to the subgenus *Callomegachile*; and most of those Group 2 subgenera known from Australia (*Hackeriapis*, *Austrochile*, *Thaumatosoma*, *Rhodomegachile* and *Chalicodomoides*). **B.** A well-supported clade (clade 7) that includes the subgenus *Dinavis* (currently synonymized under *Pseudomegachile* from which it was distant in the molecular results), the subgenus *Cesacongoa* and the *niveofasciata*-species group (currently included in the subgenus *Chalicodoma*). **C.** A well-supported clade (clade 9) composed of the subgenera *Neglectella* (currently synonymized under *Pseudomegachile* from which it was well separated in our results) and *Maximegachile*. The position of the subgenus *Lophanthedon* was unstable; this subgenus was either sister to clade 5, 7, or 9. **D.** The subgenus *Callomegachile*, excluding *Carinula* and the species of clade 6, but including *Alocanthedon*, which rendered *Callomegachile* paraphyletic. **E.** The subgenus *Pseudomegachile*, including *Largella* and *Parachalicodoma*, both nested within *Pseudomegachile*. The position of the subgenus *Stenomegachile* was variable, although it was mostly found as sister to *Pseudomegachile*,

with weak support (clade 10 in Table S3 and Fig. 1). **F.** The subgenus *Chelostomoides*. **G.** The subgenus *Chalicodoma*.

All leafcutting subgenera including three Group 2 subgenera (*Chelostomoda*, *Mitchellapis* and *Megella*), the lone member of Michener's Group 3 (*Creightonella*), and all of Michener's Group 1, formed a well-supported monophyletic group (BS > 86%; PP 1.0; Clade 11). In all analyses, *Chelostomoda* was sister to all other leafcutting subgenera (clade 12), with high support values (BS > 89%; PP 1.0). The subgenera *Creightonella*, *Sayapis*, and *Mitchellapis*, considered as intermediate between group 1 and group 2 by Michener (2007), were among the earliest diverging lineages of group 1.

Within the rest of group 1, the subgenus *Paracella* was the first branch, the second was a well supported clade (clade 15) formed by the subgenera *Amegachile* and *Eurymella*. The five subgenera of the *Pseudocentron*-group (see Michener, 2007) formed a well-supported clade (clade 18) itself forming a well-supported clade (clade 17) along with the subgenera *Argyropile*, *Megachiloides* and *Phaenosarus*. The latter, placed in synonymy with *Xanthosarus* by Michener (2000) was not closely related to that subgenus in our analyses. Within clade 19, the subgenus *Megella* branched first, although its placement as sister to the rest of clade 19 (clade 20) was not strongly supported (Table S3). The large and diverse subgenus *Eutricharaea* was monophyletic after exclusion of *Eurymella*, except for the Australian species group allied to *M. chrysopyga*, currently placed in *Eutricharaea* but which was part of a distinct clade (clade 22) formed by *Aethomegachile*, *Xanthosarus* and *Litomegachile*. The subgenus *Megachile* was part of a poorly resolved clade (Clade 23), which also included many South American subgenera. Both *Grosapis* and *Eumegachile* were nested within the subgenus *Megachile s. str*; this paraphyly of *Megachile s. str* was highly supported in all analyses.

3.4. Dating and biogeographic analyses

Results from the BEAST analysis without the calibration point of the fossil leafcuttings are presented in Fig. 2. The inclusion of the fossil leafcuttings as a calibration point had a major impact on all inferred ages. For example, the ages (with 95% confidence interval)

of the root (Pararhophitinae + Lithurginae + Megachilinae) were 87.4 mya (79.2-95.5) and 69.1 mya (58.2-90.0), respectively, with and without this calibration point. The stem age of Megachilini varied from 61.5 to 36.9 mya, while the age of the MCRA of all species having a leafcutting edge (thus the node chosen for the calibration with the fossil leafcuttings; yellow star on Fig. 2) varied from 47.6 to 22.0, the former value mirroring exactly the youngest portion of the prior distribution enforced on this particular node. In spite of these differences, the likelihood values of both analyses were comparable (-55'628.96 and -55'627.18, with and without this calibration point, respectively).

Inferred ancestral ranges found in the DEC model (LnL -272.34) for important nodes are presented on Fig. 2. The DEC model suggested ancestral ranges including the Afrotropical and Oriental regions for the MCRA of Megachilini, of clade 2, and for most splits within Michener's Group 2 (Fig. 2). For early splits within Group 1 the DEC model suggested either an Afrotropical origin or a Palearctic origin, or an ancestor distributed over both.

According to the most likely reconstruction inferred in the DEC model, there were thirteen dispersal events from the Afrotropic to the Oriental region (Fig. 2); ten dispersal events each from the Nearctic to the Neotropic and from the Afrotropic to the Palearctic; seven from the Palearctic to the Afrotropic, six from the Palearctic into the Nearctic and five in the other direction; four from the Oriental region to the Australian region; three from the Palearctic to the Oriental region, one from the Oriental region to the Afrotropic; and none from the Oriental region to the Palearctic and from the Neotropic to the Nearctic (although see discussion).

In contrast, the DEC + J model (Fig. S3) allowed long-distance dispersal between non-adjacent regions, that is, long-distance dispersal over vast oceanic distances (> 1000 km). This model had substantially higher likelihood (LnL - 246.49); ancestral range reconstructions for the tribe and for most nodes within Group 2 also included the Afrotropical and Oriental regions (Fig. S3); however, reconstructions for early splits within Group 1 favoured an Afrotropical origin. Other differences were mostly minor (Fig. S3). In contrast to the DEC model in which the only dispersal route allowed into the Nearctic was from the Palearctic, the Nearctic was colonized three times through long-

distance dispersal events, twice from the Afrotropic (MRCA of *Chelostomoides* and of clade 17) and once from the Oriental region (MRCA of *Sayapis*).

3.5. Use of DNA barcodes in combination with a nuclear backbone

Thirty-four recent, well-supported clades were retained for the examination of the phylogenetic signal of COX1; they are indicated on Fig. S1. Of these 34 clades, the phylogenetic placement of only three was recovered with BS values > 75% in ML analyses (Fig. 3, closed circles) of the mitochondrial DNA barcodes only. Crown ages of these three clades were between 0.8 and 4.2 million years according to our dating analyses. By 5 million years all BS values were below 50% (Fig. 3). Of all 34 clades, 12 were recovered in the best tree while 22 were not and thus placement of their species in the barcode tree contradicted results of our nuclear phylogenies. Incongruent placements were never supported by >50% BS values. Results were largely similar in Bayesian analyses of the barcode data: 12 clades were recovered with values above 0.5; support mostly dropped to below 0.5 PP after 4 million years. Twenty taxa had a contradictory placement compared to our final phylogenetic hypothesis, and support for the incongruent placement was commonly above 0.5 (13 of the 20 contradictory placements) and even up to 0.95 (one case) and 0.98 (one case). Having a nuclear backbone for all species outside these 34 clades and one randomly selected species within each clade did not improve the phylogenetic signal of DNA barcodes: support dropped to below 50% BS values beyond a divergence time of approximately 4 million years, and 22 clades were not recovered in the best tree.

This failure of accurate placement of taxa based only on barcode data was easily overcome if only one nuclear marker (LW-Rhodopsin) was added (Fig. 3, open circles): all 34 clades were recovered in ML analyses with BS values above 50% and only four nodes had BS values below 70% (Fig. 3). In Bayesian analyses, one clade was not recovered in the best tree; four further nodes had PP between 0.6 and 0.85; all other nodes (30) had PP above 0.95. Having a nuclear backbone for all species outside these 34 clades and for one randomly selected species within each of them resulted in trees highly similar to those derived from analyses of the full nucleotide matrix, with each of the 34 clades recovered as a monophyletic group and placed in agreement with our

phylogenetic analyses of the complete nuclear dataset. 28S was less informative than LW-Rhodopsin when analysed in combination with COX1: only 31 clades were recovered in ML analyses, six of them with support below 50% BS.

Our combined analysis including 413 taxa (outgroup excluded) is presented in Fig. S4. In this analysis, all terminals had a barcode sequence; 99 ingroup terminals had a full nuclear matrix (in red on Fig. S4), 12 had only one nuclear marker, LW-Rhodopsin (in blue), and 302 terminals had only the DNA barcode (black). While not all nodes are fully resolved, all subgenera (as reassigned in the present study, see discussion and Table 2) appear monophyletic and most of them have high statistical support, with the exception of *Hackeriapis* (see below, discussion: subgeneric classification). Morphological examination of all terminals with barcode sequences only, for which the vouchers were available, confirmed their molecular placement. A list of 34 barcodes not placed in sampled clades is given in the supplementary material; most of them are unidentified species.

Discussion

Our phylogenetic analysis provides the first comprehensive phylogeny of the species-rich bee tribe Megachilini, with a particularly dense taxon sampling that covers most of the diversity observed within the tribe: all genera, 81% of currently recognized subgenera of the genus *Megachile*, including all three of Michener's groups. A large majority of the clades are well supported and many of these clades represent morphologically well-defined species-groups or subgenera. However, the relationships among these well-defined groups, and especially among the different lineages of Michener's Group 2 subgenera are more difficult to resolve, rendering classification attempts particularly challenging.

Nevertheless, our phylogeny provides meaningful insights into the evolution of nesting biology within the tribe. First, both cleptoparasitic lineages were found to form a monophyletic group (Clade 3) in the majority of our analyses. Support for this clade was moderate to high, whereas support for any alternative placement was weak in all cases. Based on larval anatomy and nesting biology of *Radoszkowskiana rufiventris* and of some

species of *Coelioxys*, Rozen and Kamel (2008) suggested that the ancestor of both genera shared a cleptoparasitic ancestor. An open question is the ancestral mode of parasitism: Rozen and Kamel (2008) suggested a similar mode of parasitism (the host egg is killed by the modified first larval instar) in *Radoszkowskiana* and *C. (Allocoelioxys)*, but not in *C. (Liothyrapis) decipiens*, with a larval anatomy and behaviour more similar to other members of *Coelioxys* (the host larva is killed by the modified third larval instar). Second, our analyses recover a well-supported clade that includes all leafcutting members of the tribe (clade 11), including taxa from Michener's Groups 2. A topology that groups the leafcutting subgenus *Chelostomoda* with Michener's group 1 has not been suggested before; it indicates that the sophisticated ability to cut circular leaf discs has evolved only once within the tribe.

Generic classification of the Megachilini

Our phylogeny does not support recent, morphology-based classification proposals to break up Michener's group 2 subgenera into several genera (Gonzalez, 2008; Engel and Gonzalez, 2011; Gonzalez and Engel, 2012). In particular, our results strongly indicate that the elongate ("hoplitiform" or "heriadiform") subgenera, placed into one genus in Gonzalez (2008), are scattered among lineages with a more robust body form. Some superficially similar, heriadiform subgenera such as *Chelostomoides* or *Hackeriapis*, long-thought to be closely related (see biogeographic analyses below; Michener, 1979, 2007) do not form a monophyletic group in our analyses and uniting these subgenera is not appropriate. An elongate body form is most likely an adaptation to nesting in narrow cavities (Praz et al., 2008, Sedivy et al., 2013, Rozen et al., 2015).

Based on the phylogenetic results presented here, we envision four possible classifications for the tribe Megachilini (Table 1): 1. Recognition of only two genera in Megachilini: *Noteriades* and *Megachile*, the latter including the cleptoparasitic genera *Radoszkowskiana* and *Coelioxys*, which would be downgraded to subgeneric level within a heterogeneous genus *Megachile* (clade 1 in Fig. 1); 2. Recognition of seven genera: *Noteriades*, *Gronoceras*, *Heriadopsis*, *Matangapis*, *Coelioxys*, *Radoszkowskiana* and *Megachile*, the latter including all other subgenera currently assigned to the genus *Megachile* (clade 4 in Fig. 1); *Heriadopsis* and *Matangapis* are recognized as distinct

genera because of their tentative phylogenetic position (Fig. S2), the presence of an arolium and their straight-forward morphological diagnosis from all other genera of the Megachilinae; 3. Sixteen genera, with the seven genera recognized under proposal 2, plus the following: *Thaumatoma* (clade 5), *Dinavis* (clade 7), *Lophanthodon*, *Maximegachile* (clade 9, thus with *Neglectella* included as a subgenus of *Maximegachile*), *Stenomegachile*, *Callomegachile*, *Pseudomegachile*, *Chelostomoides*, *Chalicodoma*; in this scheme the genus *Megachile* would include all leafcutting subgenera, including *Chelostomoda* (Clade 11); 4. Twenty genera, with the morphologically heterogeneous clades 7, 9 and 11 further divided (see Table 1) to improve morphological diagnosis of the different genera.

The establishment of new classifications (see Vences et al., 2013 for a careful review) should balance practical aspects (diagnosis of the groups; number of species; homonymies) with phylogenetic considerations. In the case of large and diverse groups such as Megachilini, accounting for these two aspects proves extraordinarily challenging. The advantages and disadvantages of our four classification proposals are summarized in Table 1. Proposals 1 (two genera), 3 (16 genera) and 4 (20 genera) are those best in agreement with our phylogenetic hypotheses: each of the resulting genera would be well supported and the classification might have greater stability. However, there are practical drawbacks to these proposals. The first proposal suggests downgrading the cleptoparasitic lineages (the genera *Coelioxys* and *Radoszkowskiana*) to subgenera within a large and heterogeneous genus *Megachile*. Our results indeed suggest that the cleptoparasitic lineages are derived from nest-building lineages of *Megachile sensu latu*, given the well-supported position of the subgenus *Gronoceras* as the sister group to all other Megachilini (with the exception of *Noteriades*). A similar approach has been adopted for some bee groups (e.g., *Bombus* and the social parasite species included in the subgenus *Psithyrus*: Williams, 1994; Cameron et al., 2007; the cleptoparasitic *Hoplitis*: Sedivy et al., 2013). However, this approach is not desirable for Megachilini because of the strong morphological differences between *Megachile* and *Coelioxys* and the high number of species within both genera. Including both groups in the same genus would result in many homonymies and would likely not gain acceptance among bee taxonomists. The third proposal (16 genera) would also result in genera whose monophyly is well established, but whose diagnosis is challenging. Of the 14 non-

parasitic genera recognized, four would be difficult to delimit using morphological criteria (Table 1). Particularly problematic are clades 5, 7 and 9, as they include morphologically heterogeneous species and are, to our knowledge, not supported by a single diagnostic morphological character. Modifying this proposal to end up with morphologically well-defined genera would result in 20 genera (proposal 4), of which a few would still be difficult to diagnose (most notably *Thaumatosoma*, Clade 5), at least under current morphological knowledge.

Our favoured classification, given our phylogeny and the poor current knowledge of the African and Asian faunas, is our proposal two, which results in the fewest changes to the present classification, reflects our phylogeny, and is practical in term of diagnosis. Consequently, *Gronoceras*, *Heriadopsis* and *Matangapis* are recognized here as valid genera of the tribe Megachilini (new statuses) in addition to the genera *Radoszkowskiana*, *Coelioxys*, *Noteriades* and *Megachile*, resulting in the following new combinations (based on Ascher and Pickering, 2016): *Gronoceras africanibia* (Strand, 1912), *G. angolensis* Cockerell, 1935, *G. armipygata* (Strand, 1911), *G. bombiformis* (Gerstäcker, 1857), *G. catulus* (Cockerell, 1910), *G. chapini* Cockerell, 1935, *G. cincta* (Fabricius, 1781), *G. felina* Gerstäcker, 1857, *G. mabirensis* (Cockerell, 1937), *G. mutuala* (Strand, 1912), *G. praetexta* (Vachal, 1910), *Heriadopsis striatulus* Cockerell, 1931, *H. whiteanus* (Cameron, 1905) and *Matangapis alticola* (Cameron, 1902). The main drawback of this classification is the lack of strong support for clade 4, ie, the genus *Megachile* as established here, thus raising concerns about the stability of this classification. Once the African and Asian faunas are better described, it may be meaningful to adopt a classification closer to our proposals 3 or 4. Conversely, if more phylogenetic data confirms the monophyly of clade 4, we would strongly favour a classification close to our proposal 2. The new proposed classification is presented in Table 2.

Subgeneric classification

Our phylogeny strongly supports the following modifications at the subgeneric level. The groups known as *Carinula*, *Dinavis*, *Neglectella*, *Eurymella* and *Phaenosarus* are here reestablished as valid subgenera within the genus *Megachile* (new statuses). These

groups have been placed in synonymy with various subgenera of *Megachile* (Michener, 2000), but consistently emerged as distinct clades with high support values in all analyses. A list of the species included in these groups is provided by Pasteels (1965) and Mitchell (1980). The subgenus *Alocanthodon* is newly synonymized with the subgenus *Callomegachile*, and the subgenera *Largella* and *Parachalicodoma* are synonymized under the subgenus *Pseudomegachile* (new synonymies). Gonzalez et al. (2010) indicated that the subgenus known as *Platysta* was derived from *Eurymella*, which they kept in synonymy with the subgenus *Eutrichararea*. Although our sampling did not include any member of *Platysta*, we agree, based on morphological examination of both of its species (C. Praz, unpublished) with their treatment and place *Platysta* as a junior synonym of *Eurymella* (new synonymy). Michener (2000) placed the group known as *Anodonteutrichararea* (type species *M. lanigera*, included in Fig. S4) in synonymy with the subgenus *Eutrichararea*; in our analyses, *Anodonteutrichararea* (represented by *M. villipes*) was either sister to the subgenus *Paracella* (ML analyses; Fig. 1) or nested within *Paracella* (Fig. S4) and we place *Anodonteutrichararea* in synonymy with *Paracella* (new synonymy). Based on their phylogenetic position in all our analyses, *Eumegachile* and *Grosapis* are newly synonymized with the subgenus *Megachile* (new synonymies). The subgenus *Aethomegachile*, described from a single species from Thailand (Engel and Baker, 2006) appears to be a large Oriental group, in line with a recent treatment (Ascher et al., 2016). This subgenus was represented in our nuclear phylogeny by *M. conjuncta* and *M. kohtaoensis*, and several additional species clustered with them in the combined barcode + nuclear phylogeny (e.g., *M. laticeps* and *M. remota*; Fig. S4). The subgenus *Hackeriapis* was polyphyletic in our analyses (Fig. 1; Fig. S4); eventually several Australian, Group 2 subgenera may be united into a single subgenus, but we refrain from doing so given the poor resolution within clade 5 in our analyses. Lastly, our analyses also highlight several groups of species currently not placed into any existing subgenus: the *niveofasciata*-group of species (currently included in the subgenus *Chalicodoma*); the *ornata*-group of species (and *M. monoceros* Friese 1903, which showed affinities to this group), currently included in the subgenus *Callomegachile*; and the Australian group known as the *chrysopyga*-group of species, currently included in the subgenus *Eutrichararea*. All these groups appear distantly related to the subgenera in which they are currently placed.

Dating analyses

Our dating analyses indicate that the inclusion of the leafcutting trace fossils as a calibration point greatly altered ages throughout the trees. The ages inferred in analyses with this calibration point were nearly twice as old as those found in analyses without it. They actually correspond to analysis 1 (i.e., without prior distribution on the root) of Cardinal and Danforth (2013), while analyses without this calibration corresponded to their analysis 3 (with prior distribution on the root). In the absence of other fossils within Megachilini (see comments above regarding *Megachile glaesaria*), it is hard to evaluate our two alternative dating analyses. We favour the analysis without inclusion of these leafcutting trace fossils, for the following reasons. First, the assumption that these leaf incisions have been made by *Megachile* bees is only hypothetical given that many other insects produce arcuate excisions with sharp and cusped margins (Labandeira, 2002: 50). In our opinion, and based on observations of excisions made by extant leafcutting bees, excisions can be confidently attributed to leafcutting bees only if they are oblong, cuspidate; leaves bearing only small, circular incisions are often found in habitats without leafcutting bees (C. Praz, pers. observation) and are likely made by other herbivorous insects. All Eocene fossil leafcuttings are rather small and circular and none bear the typical, oblong shape made by leafcutting bees. Second, dating analyses including the leafcutting fossils are less in keeping with the broader picture of hymenopteran phylogeny. Ronquist et al., 2012 suggested a crown age for Apoidea (thus the node including the bees and the apoid wasps) of 150 my, in line with Cardinal and Danforth's analysis 3. In contrast, our analyses with the leafcuttings included as a calibration point are in line with Cardinal and Danforth's analysis 1, which recovers a crown age for Apoidea of 181 my. Additionally, analyses without the leafcutting fossils are in line with our biogeographic scenario (detailed below) and the general fossil record for megachilid bees, in particular the absence of Megachilini from Eocene amber in spite of the widespread use of resin in early diverging lineages. In summary, we do not believe these fossils are indicative of the existence of leaf-cutting Megachilini in the Eocene.

Biogeography

Ancestral range inference under both the DEC and the DEC + J models suggests that the tribe Megachilini and most lineages of the dauber bees (Fig. 2, S3) originated either in the Afrotropic or as a widespread ancestral taxon that spanned both the Afrotropic and the Oriental region. The preferred biogeographical reconstruction under both the DEC and DEC + J models suggests a higher number of dispersal events from the Afrotropic to the Oriental (14) than between any other pair of geographic regions. This high dispersal confirms the statement of Michener (1979: 331) that the oriental faunal region "is inhabited by a depauperate African bee fauna". In *Megachile*, most groups present in both regions are indeed more diverse in the Afrotropic than in the Oriental region. Exchanges appear to have been nearly uninterrupted between both regions from at least 27.2 mya to the present, e.g., at the base of clade 2 (crown age 27.2 mya), and more recently, from the Afrotropic to the Oriental zone within *Eurymella* (crown age 14.78), *Callomegachile* (crown age 11.34) and *Amegachile* (crown age 11.26), and in the other direction at least within *Carinula* (crown age 13.59). The most recent documented exchanges between the Afrotropical and Oriental regions are found in *Eurymella* (0.82 and 3.79 my). For all these exchanges, it is likely that geodispersal over the Arabian Peninsula, and then through southern Iran and Pakistan, was possible under less arid climatic conditions (Michener, 1979; see Parker, 2009).

Both the DEC and the DEC + J model also agree on the mode of colonization of the Australian faunal region, namely, from the Oriental zone. Colonization of the Australian region is found three times in our phylogeny, although at least six further, more recent colonization events are implied in clades present in Northern Australia (*Coelioxys* and the subgenera *Callomegachile*, *Amegachile*, *Chelostomoda*, *Creightonella* and *Eutricharaea*) for which no Australian specimens were sampled. In the latter six cases, colonization was probably comparatively recent as the Australian members are morphologically similar to (or even conspecific with) species present in the Oriental zone (Michener, 1965, 1979; see also Correia and Packer, submitted, for Australian *Coelioxys*). The earliest arrival of *Megachile* bees into the Australian region is found at the MRCA of a group of 5 (or 6) Australian subgenera of clade 5: *Hackeriapis*, *Thaumatoma*, *Rhodomegachile*, *Chalicodomoides* and *Austrochile*; *Schizomegachile* also likely belongs to this group (Michener, 1965: 199). These subgenera colonized Australia

once or more (phylogenetic relationships within clade 5 were poorly resolved) some 21 mya from the Oriental zone and considerable diversification occurred within these lineages in the temperate area of Australia. This time period is close to the Oligocene-Miocene boundary, which is posterior to the collision of the Australian plate with the Asian plate (around 25 mya: Hall, 2002), and which coincides with a period of global cooling and increase in the Antarctic Ice Sheet (Beddow et al., 2016). This would have reduced the oceanic barriers to dispersal from Asia to Australia. Our dating analyses with the calibration point corresponding to the fossil leafcuttings suggest an older age for clade 5 (40.5 mya), less in line with faunal exchanges between Asia and Australia. Fuller et al. (2005) and Chenoweth and Schwarz (2011) present and discuss several hypotheses for how allodapine bees may have reached Australia. The allodapine genus *Braunsapis* originated in Africa ca. 20.7 mya (26.1-17.7), dispersed into Asia 16.5 mya (20.5-13.3) and then into Australia 8.8 mya (11.9-7.0) (Fuller et al., 2005). This genus is still restricted to non-temperate regions of Northern Australia (Fuller et al., 2005). In contrast, the exoneurine genera of Allodapini (*Exoneura*, *Exoneurella*, *Brevineura* and *Inquilina*) are restricted to the temperate areas of southern Australia and are derived from a single, older dispersal into Australia (42-34 mya) (Chenoweth and Schwarz, 2011). Three scenarios for how exoneurine bees may have reached Australia were presented and discussed (Chenoweth and Schwarz, 2011): from southern Asia, as in *Braunsapis*; through a long-distance dispersal event from Africa through the Indian Ocean (Schwarz et al., 2006); or from Africa via Antarctica, also through long-distance dispersal. The biogeographic scenario inferred here for *Megachile* is interesting as it appears intermediate between the cases of *Braunsapis* (distributed in the north of Australia and likely diverged from an Oriental clade) and of the exoneurine genera (distributed in the south of Australia): Australian lineages of clade 5 are particularly diverse in temperate, southern Australia, but are inferred to have reached Australia from the Oriental Region, suggesting secondary adaptation to temperate climate. It is worth noting that inferred divergence times may substantially differ across studies: Chenoweth and Schwarz's inferred age for the crown age of Allodapini was 48.8 my (95% confidence interval 42-56), while the age for the same node was inferred to be 38 my (29-49) in analysis 3 of Cardinal and Danforth (2013), upon which our age estimates are based. Consequently the arrival of clade 5 *Megachile* into Australia may be contemporary to the arrival of the exoneurine bees.

The New World was colonized at least five times independently in non-parasitic Megachilini. These five colonization events are found at the base of the following clades: 1. the subgenus *Chelostomoides* (crown age 13.8 my, stem age 24.6 my), which has its centre of diversity in the Madrean region of North America and few species in northern South America; 2. the subgenus *Sayapis* (crown age 5.84, stem age 19.17), equally diverse in North and South America (Michener 2007); 3. clade 17 (crown age 11.9, stem age 15.3), restricted to the New World, with early diverging lineages (*Argyropile*, *Phaenosarus*, *Megachiloides*) predominantly or exclusively found in the Nearctic region; clade 17 also includes a particularly large and diverse group of predominantly South American lineages, the *Pseudocentron* group of subgenera (clade 18, with 150 described species; Ascher and Pickering, 2016). 4. the clade containing the subgenera *Litomegachile* (Nearctic) and *Xanthosarus* (Holarctic), with crown age 5.65 and stem age 7.69; 5. clade 23 (crown age 8.51, stem age 12.3), with several predominantly Neotropical lineages and one Holarctic subgenus (*Megachile* s. str.); phylogenetic relationships within clade 23 were not confidently established.

Regarding the dispersal routes associated with these five instances of colonization of the New World, both biogeographic models explored here suggest distinct scenarios. In the DEC model, colonization of the New World was only possible from the Palearctic, which resulted in ancestral ranges including the Palearctic for all clades from which the New World lineages were issued. In the DEC + J model, long-distance dispersal was possible and this model suggested that three of the five colonizations of the New World were achieved through long-distance dispersal, twice from the Afrotropic and once from the Oriental region. In contrast to other studies (see especially Almeida et al., 2011), the cases of long-distance dispersal events revealed in our DEC + J models are not strongly favoured over alternate scenarios. The first of these events was found at the base of *Chelostomoides*. This subgenus consistently nests in wood and its closed nests may have been carried over long distances but our data do not firmly exclude alternate scenario as the phylogenetic position of *Chelostomoides* was not confidently settled within Group 2. The second long-distance dispersal from Africa inferred under the DEC + J model is even more doubtful because it involves primarily ground-nesting lineages (Neff and Simpson, 1991; Sheffield et al., 2011, and references therein) at the base of clade 17. Although

clade 17 includes a diverse and predominantly neotropical clade, (clade 18, the *Pseudocentron*-group), early-diverging lineages are found in temperate climates of North America and thus dispersal from the temperate Palearctic region appears more likely. The DEC + J model also suggested a dispersal event from the Oriental region into the New World at the MRCA of *Sayapis*. As discussed above, the relative phylogenetic placement of *Sayapis*, *Mitchellapis* and *Creightonella* were only poorly supported, and again, ancestral state reconstructions were ambiguous for these nodes.

All New World clades of *Megachile*, including the subgenera *Sayapis*, *Chelostomoides*, *Litomegachile* and the early diverging lineages of clade 17 (subgenera *Argyropile*, *Phaenosarus*, *Megachiloides*), are present in North America as far north as southern Canada (Ascher and Pickering, 2016), with mean annual temperature (MAT) around 10° C (Hijmans et al., 2005). The faunal boundaries between the Oriental and Palearctic zones are not clear-cut within group 1 *Megachile*; many Oriental, subtropical subgenera are present in southern Japan or in central China, thus in the Palearctic region in areas with MAT around 12 °C. This pattern is especially pronounced in the leafcutting subgenera (*Chelostomoda*, *Amegachile*, *Megella*, *Aethomegachile*), the group within which most dispersal events into the New World have occurred. Wolfe (1994) suggested a MAT as high as 12 °C at paleolatitude 60°N (the latitude of Beringia) between 20 and 14 mya. Consequently the eastern Palearctic, subtropical region harbouring a particularly diverse megachiline fauna was likely more extensive in the Early Miocene and may have offered a corridor through Beringia into the New World. These ages and climatic conditions are perfectly in keeping with the oldest dates of arrival of Megachilini into the New World and suggest that lineages currently restricted to latitudes around 40°N and MAT 12 °C could have migrated through Beringia before 14 mya. Several additional exchanges between the Palearctic and the Nearctic regions are likely found in the Holarctic subgenera *Xanthosarus* and *Megachile* (crown ages 5.65 and 6.05 my, respectively); these two subgenera have their centre of diversity in boreal climates and thus are expected to have been able to cross Beringia after 14 mya. In keeping with conclusions inferred for Lycaenid butterflies (Vila et al., 2011), Beringia appears to have acted as a climate-regulated dispersal route for bees.

Use of barcodes in phylogenetic studies

Our attempts to use DNA barcodes as a tool in phylogenetic reconstruction yield mixed results. On the one hand, our approach to using DNA barcodes to guide taxon sampling was promising: it highlighted several members of the heterogeneous subgenera *Callomegachile* (numbers 162, 164) and *Eutricharaea* (number 645), which, based upon our analyses of the entire DNA dataset and future morphological research, will likely eventually constitute new subgenera. In the case of the diverse and polyphyletic subgenus *Callomegachile*, our combined nuclear + barcode analysis (Fig. S4) also enabled the placement of the type species of *Callomegachile* (*M. mystaceana*) with high support values, facilitating future systematic treatment and reclassification of this group of bees.

However, the addition of COX1 to our nuclear data (ML analyses 5 and 6) hardly affected bootstrap support, and if it did, it mostly lowered support values (Table S3), confirming that this marker is of very limited utility for old divergences (e.g., Kjer et al., 2001, 2014; Mueller, 2006; Klopstein et al., 2010; Ekrem et al., 2010). Moreover, our attempts to recover 34 confidently established clades with a matrix of the DNA barcodes demonstrate that COX1 alone contains very little phylogenetic signal beyond an approximate age of 4 my (Fig. 3). A few isolated clades with older crown ages were recovered in Bayesian analyses, but these placements were accompanied by many contradictory placements with equally high posterior probabilities. Such contradictory placements are particularly worrisome with respect to the use of barcodes in phylogenetics, as they are difficult to distinguish from reliable phylogenetic signal. There are several issues with relying on a single mitochondrial marker for phylogenetic placement, such as saturation in the phylogenetic signal, biased compositional frequencies (e.g., Timmermans et al., 2016), or, worse, horizontal transfer of mitochondria between distantly related species (Nicholls et al. 2012; Klopstein et al. 2016). Our limited data do not enable us to distinguish between these hypotheses; however, base composition of the third nucleotide position of COX1 was strongly biased in our dataset, with a GC content ranging from 0.8% to 22.9%; all erroneous placements with support above 0.8 in Bayesian analyses involved the grouping of unrelated taxa with similarly biased base composition.

In conclusion, we do not recommend the use of COX1 alone in phylogenetics; however, the approach of combining a solid backbone inferred with a nuclear dataset with two markers used as a barcode (eg., COX1 and LW-Rhodopsin or ribosomal DNA) has great potential (see Kjer et al., 2014). Given the limitations in the use of COX1 to delineate species due to mitochondrial introgression, heteroplasmy (Magnacca and Brown, 2010), or within-species divergences not associated with morphological divergences (eg., Huemer et al. 2014; Kekkonen et al. 2015; Schmidt et al. 2015), having two barcoding markers, one mitochondrial and one nuclear would offer numerous advantages both for species delineation and for enabling their placement in a phylogenetic framework (Kjer et al. 2014).

Conclusion

Our study is a first step towards a phylogeny-based classification of the Megachilini yet further work is needed. First, our dataset should be completed for missing subgenera, especially from Michener's group 2: *Heriadopsis* and *Matangapis*, whose position was not well established here, and *Cestella* and *Schizomegachile*. Second, other phylogenetic markers, such as ultraconserved elements (Faircloth et al., 2014) may help settle relationships among the various lineages of the dauber bees. Once the generic classification is better established, much work remains to be undertaken at the subgeneric level: the subgeneric placement of many *Megachile* species remains unclear, especially in the diverse tropical faunas. Our approach of combining a core nuclear phylogenetic backbone with DNA barcodes represents a novel application of increasingly abundant barcode data, providing a testable hypothesis for their phylogenetic placement. In our study, over 400 species of *Megachile* (Fig. S4), more than a quarter of the known species diversity, were placed in a phylogenetic context.

Acknowledgement

CP wishes to express his deep gratitude to Maximilian Schwarz for his continuous support and for putting his large collection at our disposal for ongoing studies on the genus *Megachile*. We thank all people listed in Table S1 who contributed specimens for

this study, in particular, Andreas Müller, Claudio Sedivy, Sam Droege (US Geological Survey), Achik Dorchin, Terry Griswold (USDA ARS Bee Biology & Systematics Laboratory), Jerry Rozen and John Ascher (American Museum of Natural History), Philippe Sagot, Peter Kwapong, Alana Pindar and the parataxonomists associated with the TIGER project. Nicolai de Silva and Natalia Veiga were largely responsible for processing Thai and Argentinean bees for barcoding respectively. LP's sampling in Chile was supported by the Natural Sciences and Engineering Research Council of Canada through research and discovery grants, the National Geographic Society and his own salary. Sequencing was funded by the University of Neuchatel (among other, through an "Overhead" grant to CP) and the Linnean Society (grant to VT), and was performed at the Genetic Diversity Center of ETH Zurich. We thank Stefan Risch and Connal Eardley for identifying many specimens, Nadir Alvarez and Tomasz Suchan for assistance with the NGS data, and Jessica Litman and two anonymous reviewers for numerous comments on this manuscript.

References

- Almeida, E.A.B., Danforth, B.N., 2009. Phylogeny of colletid bees (Hymenoptera: Colletidae) inferred from four nuclear genes. *Mol. Phylogenet. Evol.* 50, 290–309. doi:10.1016/j.ympev.2008.09.028
- Almeida, E.A.B., Pie, M.R., Brady, S.G., Danforth, B.N., 2011. Biogeography and diversification of colletid bees (Hymenoptera: Colletidae): emerging patterns from the southern end of the world. *J. Biogeogr.* 39, 526–544. doi:10.1111/j.1365-2699.2011.02624.x
- Ascher, J.S., Pickering, J., 2016. Discover Life bee species guide and world checklist (Hymenoptera: Apoidea: Anthophila). [WWW Document]. URL <http://www.discoverlife.org/20/q?search=Apoidea> (accessed 1.25.16).
- Ascher, J.S., Risch, S., Soh, Z.W.W., Lee, J.X.Q., Soh, E.J.Y. 2016. *Megachile* leaf-cutter and resin bees of Singapore (Hymenoptera: Apoidea: Megachilidae). *Raffles Bull. Zool. Suppl.* 32, 33-55.
- Beddow, H.M., Liebrand, D., Sluijs, A., Wade, B.S., 2016. Global change across the Oligocene - Miocene transition: High - resolution stable isotope records from IODP Site U1334 (equatorial Pacific Ocean). *Paleoceanography* 31. doi:10.1002/(ISSN)1944-9186
- Blaimer, B.B., Fisher, B.L., 2013. How much variation can one ant species hold? Species delimitation in the *Crematogaster kelleri*-group in Madagascar. *PLoS ONE* 8, e68082. doi:10.1371/journal.pone.0068082
- Bronner, I.F., Quail, M.A., Turner, D.J., Swerdlow, H., 2014. Improved protocols for Illumina sequencing, in: *Current protocols in human genetics*. John Wiley & Sons, Inc., p. Unit 18.2. doi:10.1002/0471142905.hg1802s62
- Cameron, S.A., Hines, H.M., Williams, P.H., 2007. A comprehensive phylogeny of the bumble bees (*Bombus*). *Biol. J. Lin. Soc.* 91, 161–188.
- Cardinal, S., Danforth, B.N., 2013. Bees diversified in the age of eudicots. *P. Roy. Soc. B-Biol. Sci.* 280, 20122686. doi:10.1098/rspb.2012.2686
- Cardinal, S., Straka, J., Danforth, B.N., 2010. Comprehensive phylogeny of apid bees reveals the evolutionary origins and antiquity of cleptoparasitism. *P. Natl. Acad. Sci. U.S.A.* 107, 16207–16211. doi:10.1073/pnas.1006299107
- Chenoweth, L.B., Schwarz, M.P., 2011. Biogeographical origins and diversification of the exoneurine allodapine bees of Australia (Hymenoptera, Apidae). *J. Biogeogr.* 38, 1471–1483. doi:10.1111/j.1365-2699.2011.02488.x
- Cockerell, T., 1935. African bees of the genus *Gronoceras*. *Am. Mus. Novit.* 1–6.
- Danforth, B.N., Cardinal, S., Praz, C., Almeida, E.A.B., Michez, D., 2013. The impact of molecular data on our understanding of bee phylogeny and evolution. *Annu. Rev. Entomol.* 58, 57–78. doi:10.1146/annurev-ento-120811-153633
- Danforth, B.N., Ji, S., 2001. Australian *Lasioglossum*+ *Homalictus* form a monophyletic group: resolving the “Australian enigma.” *Syst. Biol.* 50, 268–283.
- Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29, 1969–1973. doi:10.1093/molbev/mss075

- Durante, S., Cabrera, N., 2009. Cladistic analysis of *Megachile (Chrysosarus)* Mitchell and revalidation of *Megachile (Dactylomegachile)* Mitchell (Hymenoptera, Megachilidae). *Zootaxa* 2284, 48–62.
- Eardley, C., 2012. A taxonomic revision of the southern African species of dauber bees in the genus *Megachile* Latreille (Apoidea: Megachilidae). *Zootaxa* 3460, 1–139.
- Ekrem, T., Willassen, E., Stur, E., 2010. Phylogenetic utility of five genes for dipteran phylogeny: a test case in the Chironomidae leads to generic synonymies. *Mol. Phylogenet. Evol.* 57, 561–571.
- Engel, M.S., 1999. *Megachile glaesaria*, the first megachilid bee fossil from amber (Hymenoptera, Megachilidae). *Am. Mus. Novit.* 3276, 1–13.
- Engel, M.S., Baker, D.B., 2006. A remarkable new leaf-cutter bee from Thailand (Hymenoptera: Megachilidae). *Beitr. Ent.* 56, 69–74.
- Engel, M.S., Gonzalez, V.H., 2011. *Alocanthodon*, a new subgenus of *Chalicodoma* from Southeast Asia (Hymenoptera, Megachilidae). *Zookeys* 101, 51–79. doi:10.3897/zookeys.101.1182
- Faircloth, B.C., Branstetter, M.G., White, N.D., Brady, S.G., 2014. Target enrichment of ultraconserved elements from arthropods provides a genomic perspective on relationships among Hymenoptera. *Mol. Ecol. Resour.* 15, 489–501. doi:10.1111/1755-0998.12328
- Fuller, S., Schwarz, M., Tierney, S., 2005. Phylogenetics of the allodapine bee genus *Braunsapis*: historical biogeography and long-range dispersal over water. *J. Biogeogr.* 32, 2135–2144. doi:10.1111/j.1365-2699.2005.01354.x
- Gibbs, J., Brady, S.G., Kanda, K., Danforth, B.N., 2012. Phylogeny of halictine bees supports a shared origin of eusociality for *Halictus* and *Lasioglossum* (Apoidea: Anthophila: Halictidae). *Mol. Phylogenet. Evol.* 65, 926–939. doi:10.1016/j.ympev.2012.08.013
- Gonzalez, V.H., 2008. Phylogeny and classification of the bee tribe Megachilini (Hymenoptera: Apoidea, Megachilidae), with emphasis on the genus *Megachile*. PhD Thesis, University of Kansas, Kansas, USA.
- Gonzalez, V.H., 2013. Taxonomic comments on *Megachile* subgenus *Chrysosarus* (Hymenoptera: Megachilidae). *Journal of Melittology* 5, 1–6.
- Gonzalez, V.H., Engel, M.S., 2012. African and Southeast Asian *Chalicodoma* (Hymenoptera: Megachilidae): new Subgenus, new Species, and notes on the composition of *Pseudomegachile* and *Largella*. *Ann. Zool.* 62, 599–617. doi:10.3161/000345412X659669
- Gonzalez, V.H., Engel, M.S., Hinojosa-Díaz, I.A., 2010. A new species of *Megachile* from Pakistan, with taxonomic notes on the subgenus *Eutricharaea* (Hymenoptera: Megachilidae). *J. Kansas. Entomol. Soc.* 83, 58–67.
- Gonzalez, V.H., Griswold, T., Praz, C.J., Danforth, B.N., 2012. Phylogeny of the bee family Megachilidae (Hymenoptera: Apoidea) based on adult morphology. *Syst. Entomol.* 37, 261–286. doi:10.1111/j.1365-3113.2012.00620.x

- Griswold, T., Gonzalez, V., 2011. New species of the Eastern Hemisphere genera *Afroheriades* and *Noteriades* (Hymenoptera, Megachilidae), with keys to species of the former. *Zookeys* 159, 65–80. doi:10.3897/zookeys.159.2283
- Hall, R., 2002. Cenozoic geological and plate tectonic evolution of SE Asia and the SW Pacific: computer-based reconstructions, model and animations. *J Asian Earth Sci* 20, 353–431.
- Hebert, P.D.N., Penton, E.H., Burns, J.M., Janzen, D.H., Hallwachs, W., 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *P. Natl. Acad. Sci. U.S.A.* 101, 14812–14817. doi:10.1073/pnas.0406166101
- Hebert, P.D.N., Ratnasingham, S., deWaard, J., 2003. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *P. Roy. Soc. B-Biol. Sci.* 270, S96–S99. doi:10.1098/rsbl.2003.0025
- Hines, H.M., 2008. Historical biogeography, divergence times, and diversification patterns of bumble bees (Hymenoptera: Apidae: *Bombus*). *Syst. Biol.* 57, 58–75. doi:10.1080/10635150801898912
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25, 1965–1978.
- Huemer, P., Mutanen, M., Sefc, K.M., Hebert, P.D.N., 2014. Testing DNA barcode performance in 1000 species of European Lepidoptera: large geographic distances have small genetic impacts. *PLoS ONE* 9, 1–21. doi:10.1371/journal.pone.0115774.s003
- Iwata, K., 1976. *Evolution of instinct*. Amerind Publishing Co., New Delhi.
- Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi:10.1093/molbev/mst010
- Katoh, K., Toh, H., 2008. Improved accuracy of multiple ncRNA alignment by incorporating structural information into a MAFFT-based framework. *BMC Bioinformatics* 9, 212. doi:10.1186/1471-2105-9-212
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi:10.1093/bioinformatics/bts199
- Kekkonen, M., Mutanen, M., Kaila, L., Nieminen, M., Hebert, P.D.N., 2015. Delineating species with DNA barcodes: a case of taxon dependent method performance in moths. *PLoS ONE* 10, e0122481. doi:10.1371/journal.pone.0122481.s008
- King, J., Exley, E.M., 1985. A reinstatement and revision of the genus *Thaumatostoma* Smith (Apoidea: Megachilidae). *J. Aust. ent. Soc.* 24, 87–92.
- Kjer, K.M., Blahnik, R.J., Holzenthal, R.W., 2001. Phylogeny of Trichoptera (caddisflies): characterization of signal and noise within multiple datasets. *Syst. Biol.* 50, 781–816.

- Kjer, K.M., Zhou, X., Frandsen, P.B., 2014. Moving toward species-level phylogeny using ribosomal DNA and COI barcodes: an example from the diverse caddisfly genus *Chimarra* (Trichoptera: Philopotamidae). *Arthropod Syst. Phylo.* 72, 345–354.
- Klopfstein, S., Kropf, C., Baur, H., 2016. Wolbachia endosymbionts distort DNA barcoding in the parasitoid wasp genus *Diplazon* (Hymenoptera: Ichneumonidae). *Zool. J. Linn. Soc.-Lond.* 1–17.
- Klopfstein, S., Kropf, C., Quicke, D.L.J., 2010. An evaluation of phylogenetic informativeness profiles and the molecular phylogeny of Diplazontinae (Hymenoptera, Ichneumonidae). *Syst. Biol.* 59, 226–241. doi:10.1093/sysbio/syp105
- Kronenberg, S., Hefetz, A., 1984. Role of labial glands in nesting behaviour of *Chalicodoma sicula* (Hymenoptera; Megachilidae). *Physiol. Entomol.* 9, 175–179.
- Labandeira, C.C., 2002. Paleobiology of middle Eocene plant-insect associations from the Pacific Northwest: A preliminary report. *Rocky Mountain Geology* 37, 31–59.
- Lanfear, R., Calcott, B., Ho, S.Y.W., Guindon, S., 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol. Biol. Evol.* 29, 1695–1701. doi:10.1093/molbev/mss020
- Litman, J.R., Danforth, B.N., Eardley, C.D., Praz, C.J., 2011. Why do leafcutter bees cut leaves? New insights into the early evolution of bees. *P. Roy. Soc. B-Biol. Sci.* 278, 3593–3600. doi:10.1098/rspb.2011.0365
- Litman, J.R., Praz, C.J., Danforth, B.N., Griswold, T.L., Cardinal, S., 2013. Origins, evolution, and diversification of cleptoparasitic lineages in long-tongued bees. *Evolution* 67, 2982–2998. doi:10.1111/evo.12161
- Magnacca, K.N., Brown, M.J.F., 2010. Mitochondrial heteroplasmy and DNA barcoding in Hawaiian *Hylaeus* (*Nesoprosopis*) bees (Hymenoptera: Colletidae). *BMC Evol. Biol.* 10, 1–16. doi:10.1186/1471-2148-10-174
- Magnacca, K.N., Brown, M.J., 2012. DNA barcoding a regional fauna: Irish solitary bees. *Mol. Ecol. Resour.* 12, 990–998.
- Matzke, N.J., 2013. BioGeoBEARS: Biogeography with Bayesian (and likelihood) evolutionary analysis in R scripts. R package, version 0.2 1.
- Matzke, N.J., 2014. Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. *Syst. Biol.* 63, 951–970. doi:10.1093/sysbio/syu056
- Michener, C.D., 1965. A classification of the bees of the Australian and South Pacific regions. *Bulletin of the American Museum of Natural History* 130, 1–362.
- Michener, C.D., 1979. Biogeography of the bees. *Ann. Mo. Bot. Gard.* 66, 277–347.
- Michener, C.D., 2000. The bees of the world. The Johns Hopkins University Press, Baltimore, Maryland.
- Michener, C.D., 2007. The bees of the world, second ed. The Johns Hopkins University Press, Baltimore, Maryland.

- Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES science gateway for inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop (GCE), New Orleans LA, USA.
- Mitchell, T.B., 1980. A generic revision of the Megachilinae bees of the Western Hemisphere (Hymenoptera: Megachilidae). Department of Entomology, North Carolina State University, Raleigh NC, USA.
- Mueller, R.L., 2006. Evolutionary rates, divergence dates, and the performance of mitochondrial genes in Bayesian phylogenetic analysis. *Syst. Biol.* 55, 289–300. doi:10.1080/10635150500541672
- Neff, J.L., Simpson, B.B., 1991. Nest biology and mating behavior of *Megachile fortis* in central Texas (Hymenoptera: Megachilidae). *J. Kansas. Entomol. Soc.* 64, 324–336.
- Nicholls, J.A., Challis, R.J., Mutun, S., Stone, G.N., 2012. Mitochondrial barcodes are diagnostic of shared refugia but not species in hybridizing oak gallwasps. *Mol. Ecol.* 21, 4051–4062. doi:10.1111/j.1365-294X.2012.05683.x
- O'Toole, C., Raw, A., 1991. Bees of the World. Blanford, London.
- Paradis, E., Claude, J., Strimmer, K., 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20, 289–290. doi:10.1093/bioinformatics/btg412
- Parker, A.G., 2009. Pleistocene climate change in Arabia: developing a framework for hominin dispersal over the Last 350 ka. Springer Netherlands, Dordrecht. doi:10.1007/978-90-481-2719-1_3
- Pasteels, J.J., 1965. Revision des Megachilidae (Hymenoptera apoidea) de l'Afrique noire. Annales du Musée Royal de L'Afrique Centrale, Tervuren, Belgium.
- Praz, C.J., Müller, A., Danforth, B.N., Griswold, T.L., Widmer, A., Dorn, S., 2008. Phylogeny and biogeography of bees of the tribe Osmiini (Hymenoptera: Megachilidae). *Mol. Phylogenet. Evol.* 49, 185–197. doi:10.1016/j.ympev.2008.07.005
- Praz, C.J., Packer, L., 2014. Phylogenetic position of the bee genera *Ancyla* and *Tarsalia* (Hymenoptera: Apidae): a remarkable base compositional bias and an early Paleogene geodispersal from North America to the Old World. *Mol. Phylogenet. Evol.* 81, 258–270. doi:10.1016/j.ympev.2014.09.003
- Rasmussen, C., Cameron, S.A., 2010. Global stingless bee phylogeny supports ancient divergence, vicariance, and long distance dispersal. *Biol. J. Lin. Soc.* 99, 206–232. doi:10.1111/j.1095-8312.2009.01341.x
- Ratnasingham, S., Hebert, P.D.N., 2013. A DNA-Based Registry for All Animal Species: The Barcode Index Number (BIN) System. *PLoS ONE* 8, e66213. doi:10.1371/journal.pone.0066213.s001
- Raw, A., 2006. A new subgenus and three new species of leafcutter bees, *Megachile* (*Austrosarus*) (Hymenoptera, Megachilidae) from central Brazil. *Zootaxa* 1228: 25–34.
- Romiguier, J., Cameron, S.A., Woodard, S.H., Fischman, B.J., Keller, L., Praz, C.J., 2016. Phylogenomics controlling for base compositional bias reveals a single origin of eusociality in corbiculate bees. *Mol. Biol. Evol.* 33, 670–678. doi:10.1093/molbev/msv258

- Ronquist, F., Klopstein, S., Vilhelmsen, L., Schulmeister, S., Murray, D.L., Rasnitsyn, A.P., 2012. A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. *Syst. Biol.* 61, 973–999. doi:10.1093/sysbio/sys058
- Rozen, J.G., Kamel, S.M., 2007. Investigations on the biologies and immature stages of the cleptoparasitic bee genera *Radoszkowskiana* and *Coelioxys* and their *Megachile* hosts (Hymenoptera : Apoidea : Megachilidae : Megachilini). *Am. Mus. Novit.* 3573, 1–43.
- Rozen, J.G., Kamel, S.M., 2008. Hospicidal behavior of the cleptoparasitic bee *Coelioxys* (*Allocoelioxys*) *coturnix*, including descriptions of its larval instars (Hymenoptera: Megachilidae). *Am. Mus. Novit.* 3636, 1–15.
- Rozen, J.G., Pisanty, G., Trunz, V., Bénon, D., Dorchin, A., Praz, C.J., 2015. Nesting Biology, Flower Preferences, and Larval Morphology of the Little-Known Old World Bee *Ochreriades fasciatus* (Apoidea: Megachilidae: Megachilinae). *Am. Mus. Novit.* 3830, 1–18. doi:10.1206/3830.1
- Rueda, M., Rodríguez, M.Á., Hawkins, B.A., 2013. Identifying global zoogeographical regions: lessons from Wallace. *J. Biogeogr.* 40, 2215–2225. doi:10.1111/jbi.12214
- Schmidt, S., Schmid-Egger, C., Morinière, J., Haszprunar, G., Hebert, P.D.N., 2015. DNA barcoding largely supports 250 years of classical taxonomy: identifications for Central European bees (Hymenoptera, Apoidea partim). *Mol. Ecol. Resour.* 15, 985–1000. doi:10.1111/1755-0998.12363
- Schwarz, M. 2001. Revision der Gattung *Radoszkowskiana* Popov 1955 und ein Beitrag zur Kenntnis der Gattung *Coelioxys* Latreille 1809 (Hymenoptera: Apidae: Megachilinae). *Linzer biol. Beitr.* 33, 1267–1286.
- Schwarz, M., Fuller, S., Tierney, S., Cooper, S.J., 2006. Molecular phylogenetics of the exoneurine allodapine bees reveal an ancient and puzzling dispersal from Africa to Australia. *Syst. Biol.* 55, 31–45. doi:10.1080/10635150500431148
- Schwarz, M.P., Richards, M.H., Danforth, B.N., 2007. Changing paradigms in insect social evolution: insights from halictine and allodapine bees. *Annu. Rev. Entomol.* 52, 127–150. doi:10.1146/annurev.ento.51.110104.150950
- Sedivy, C., Dorn, S., Widmer, A., 2013. Host range evolution in a selected group of osmiine bees (Hymenoptera: Megachilidae): the Boraginaceae - Fabaceae paradox. *Biol. J. Lin. Soc.* 108, 34–54.
- Sedivy, C., Praz, C.J., Müller, A., Widmer, A., Dorn, S., 2008. Patterns of host-plant choice in bees of the genus *Chelostoma*: the constraint hypothesis of host-range evolution in bees. *Evolution* 62, 2487–2507. doi:10.1111/j.1558-5646.2008.00465.x
- Sheffield, C.S., Hebert, P.D.N., Kevan, P.G., Packer, L., 2009. DNA barcoding a regional bee (Hymenoptera: Apoidea) fauna and its potential for ecological studies. *Mol. Ecol. Resour.* 9, 196–207. doi:10.1111/j.1755-0998.2009.02645.x
- Sheffield, C.S., Ratti, C., Packer, L., 2011. Leafcutter and mason bees of the genus *Megachile* Latreille (Hymenoptera: Megachilidae) in Canada and Alaska. *Canadian Journal of Arthropod Identification* 18, 1–107. doi:10.3752/cjai.2011.18
- Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi:10.1093/bioinformatics/btu033

- Swofford, D.L., 2002. Phylogenetic analysis using parsimony (* and other methods). Version 4. Sinauer Associates, Sunderland MA, USA.
- Timmermans, M.J.T.N, Barton, C., Haran, J., Ahrens, D., Culverwell, C.L., Ollikainen, A., Dodsworth, S., Foster, P.G., Bocak, L., Vogler, A.P. 2016. Family-level sampling of mitochondrial genomes in Coleoptera: compositional heterogeneity and phylogenetics. *Mol. Biol. Evol.* 8, 161-175. doi: 10.1093/gbe/evv241
- Tkalců, B., 1969. Beitrage zur Kenntnis der Fauna Afganistans - *Chalicodoma* Lep., Megachilidae, Apoidea. *Act. Mus. Morav. Sci. na* 54, 347–384.
- Vences, M., Guayasamin, J.M., Miralles, A., la Riva, de, I., 2013. To name or not to name: Criteria to promote economy of change in Linnaean classification schemes. *Zootaxa* 3636, 201–244. doi:10.11646/zootaxa.3636.2.1
- Vila, R., Bell, C.D., Macniven, R., Goldman-Huertas, B., Ree, R.H., Marshall, C.R., Balint, Z., Johnson, K., Benyamini, D., Pierce, N.E., 2011. Phylogeny and palaeoecology of *Polyommatus* blue butterflies show Beringia was a climate-regulated gateway to the New World. *P. Roy. Soc. B-Biol. Sci.* 278, 2737–2744. doi:10.1126/science.285.5431.1265
- Wedmann, S., Wappler, T., Engel, M.S., 2009. Direct and indirect fossil records of megachilid bees from the Paleogene of Central Europe (Hymenoptera: Megachilidae). *Naturwissenschaften* 96, 703–712. doi:10.1007/s00114-009-0525-x
- Williams, P.H., 1994. Phylogenetic relationships among bumble bees (*Bombus* Latr.): a reappraisal of morphological evidence. *Syst. Entomol.* 19.
- Wolfe, J.A., 1994. An analysis of Neogene climates in Beringia. *Palaeogeogr. Palaeocl.* 108, 207–216.

Tables

Table 1. Four classification proposals for the tribe Megachilini; genera under each proposal are presented in bold, with the estimated included number of species (based on Ascher and Pickering, 2016) indicated in parentheses. Morphologically heterogeneous genera which would be particularly challenging to identify under each proposal are indicated with an asterisk.

Proposal	Number of genera	Agreement with phylogeny	Diagnosability	Changes compared to actual classification	Included Genera
1	2	Good	Good	Major	Noteriades (16), Megachile [incl. <i>Coelioxys</i> and <i>Radoszkowskiana</i>] (2014)
2	7	Moderate; monophyly of <i>Megachile</i> not recovered in all analyses	Good	Minor	Noteriades (16), Gronoceras (12), Matangapis (1), Heriadopsis (2), Coelioxys (475), Radoszkowskiana (4), Megachile [clade 4] (1520).
3	16	Good (placement and status of <i>Cestella</i> unknown)	Poor	Major	Noteriades (16), Gronoceras (12), Matangapis (1), Heriadopsis (2), Coelioxys (475), Radoszkowskiana (4), Dinavis * [clade 7, incl. <i>Cesacongoa</i> and the <i>niveofasciata</i> -gr] (7), Lophanthedon (5), Thaumatossoma * [clade 5] (117), Maximegachile * [clade 9; incl. <i>Neglectella</i>] (13), Callomegachile (98), Stenomegachile (4), Pseudomegachile (73), Chalicodoma (61), Chelostomoides (34), Megachile * [incl. <i>Chelostomoda</i> , <i>Creightonella</i> , <i>Mitchellapis</i> and <i>Megella</i>] (1117)
4	20	Good (placement and status of <i>Cestella</i> unknown)	Acceptable	Major	Noteriades (16), Gronoceras (12), Matangapis (1), Heriadopsis (2), Coelioxys (475), Radoszkowskiana (4), Dinavis (2), niveofasciata -gr (4), Cesacongoa (1), Lophanthedon (5), Thaumatossoma * [clade 5] (117), Maximegachile (3), Neglectella (10), Callomegachile (98), Pseudomegachile (73), Stenomegachile (4), Chalicodoma (61), Chelostomoides (34), Chelostomoda (22), Megachile (1095)

Table 2. New classification proposed for the bee tribe Megachilini. Genera are indicated in bold; new synonymies are indicated in parenthesis.

Genus *Noteriades*

Genus ***Gronoceras*** new status

Genus ***Matangapis*** new status

Genus ***Heriadopsis*** new status

Genus ***Radoszkowskiana***

Genus ***Coelioxys***

Genus ***Megachile***

Group 1 subgenera		Group 2 subgenera:
<i>Acentron</i>	<i>Melanosarus</i>	<i>Austrochile</i>
<i>Aethomegachile</i>	<i>Mitchellapis</i> ¹	<i>Carinula</i> new status
<i>Amegachile</i>	<i>Moureapis</i>	<i>Callomegachile</i> (incl. <i>Alocanthodon</i>)
<i>Argyropile</i>	<i>Neochelynia</i>	<i>Cesacongoa</i> (= <i>Cuspidella</i>)
<i>Austromegachile</i>	<i>Neocressoniella</i>	<i>Cestella</i>
<i>Austrosarus</i> ³	<i>Paracella</i> (incl. <i>Anodonteutricharaea</i>)	<i>Chalicodoma</i>
<i>Chelostomoda</i> ¹	<i>Phaenosarus</i>	<i>Chalicodomoides</i>
<i>Chrysosarus</i>	<i>Pseudocentron</i>	<i>Chelostomoides</i>
<i>Creightonella</i> ²	<i>Ptilosaroides</i>	<i>Dinavis</i> new status
<i>Cressoniella</i>	<i>Ptilosarus</i>	<i>Hackeriapis</i>
<i>Dactylomegachile</i> ³	<i>Rhyssomegachile</i>	<i>Maximegachile</i>
<i>Dasymegachile</i>	<i>Sayapis</i>	<i>Neglectella</i> new status
<i>Eurymella</i> new status (incl. <i>Platysta</i>)	<i>Schrotkyapis</i>	<i>Pseudomegachile</i> (incl. <i>Largella</i> and <i>Parachalicodoma</i>)
<i>Eutricharaea</i>	<i>Stelodides</i> ³	<i>Rhodomegachile</i>
<i>Leptorachis</i>	<i>Trichurochile</i>	<i>Schizomegachile</i>
<i>Litomegachile</i>	<i>Tylomegachile</i>	<i>Stenomegachile</i>
<i>Megachile</i> s. str (incl <i>Grosapis</i> and <i>Eumegachile</i>)	<i>Xanthosarus</i>	<i>Thaumatoma</i>
<i>Megachiloides</i>	<i>Zonomegachile</i> ³	
<i>Megella</i> ¹		

¹subgenera included in group 2 in Michener (2007)

²lone subgenus of group 3 in Michener (2007)

³placed in synonymy with *Megachile* (*Chrysosarus*) by Gonzalez (2013)

Figure legends

Figure 1. Best tree found in maximum likelihood analyses of the matrix with significantly heterogeneous partitions removed (third codon positions of NAK and COX1), with the introns, partitioned by codon position; number above branches represent bootstrap support values (only values $\geq 50\%$) based on 1000 bootstrap replicates. Outgroup taxa have been removed and the branch leading to the ingroup has been shortened for better graphic representation. Numbers in blue squares indicate 23 nodes discussed in the text for which no taxonomic name is available. Pictures by Dino Martins (*Gronoceras* sp), Andreas Müller (*Coelioxys afra*, *Megachile parietina*) and Albert Krebs (*Megachile alpicola*).

Figure 2. Maximum clade credibility tree found in the dated BEAST analysis of the matrix with significantly heterogeneous partitions removed (third codon positions of NAK and COX1), with the introns, without the calibration point corresponding to the fossil leafcuttings (yellow star), partitioned by codon position. The pie-charts on the nodes represent the ancestral range reconstructions performed under the DEC model in Biogeobears; the six geographic ranges A-F are indicated on the map; colour codes for polymorphic ranges are indicated above the map. The arrows on the diagram above the map represent all exchanges suggested by the most likely reconstruction in the DEC model; the width of the arrow is proportional to the number of exchanges documented, which are indicated by numbers above the arrow.

Figure 3. Support values (bootstrap support in maximum analyses with 1000 bootstrap replicates) for 34 well established clades in maximum likelihood analyses of DNA barcodes (COX1) only (closed circles) or DNA barcodes + the nuclear marker LW-Rhodopsin (open circles). Values are plotted against the crown age of the clade in our BEAST analysis.

Figure 1

Figure 2

Figure 3

“backbone”
phylogeny

+

DNA barcodes

=

densely sampled
phylogeny

+

=

→

new classification:

Genus *Gronoceras*

Cleptoparasitic clade:
Genus *Coelioxys*
Genus *Radoszkowskiana*

“Dauber” subgenera
of *Megachile*
(paraphyletic)

“Leafcutter” subgenera
of *Megachile*
(monophyletic)