Insect Systematics and Diversity, 9(3), 1; 2025, ixaf011
https://doi.org/10.1093/isd/ixaf011
Short Communication

ENTOMOLOGICAL SOCIETY OF AMERICA
SHARING INSECT SCIENCE GLOBALLY

OXFORD

Molecular phylogenetics, phylogenomics, and phylogeography

Expanded phylogeny of Nomadinae (Hymenoptera: Apidae) with integration of UCE and DNA barcode sequence data

Trevor J.L. Sless[1,2,*], Katherine A. Odanaka[1], L. Ximena Alva-Caballero[2], Jeremy B. Searle[2], Bryan N. Danforth[3], and Sandra M. Rehan[1]

[1]Department of Biology, York University, Toronto, ON, Canada
[2]Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY, USA
[3]Department of Entomology, Cornell University, Ithaca, NY, USA

[*]Corresponding author. Department of Biology, York University, Toronto, ON M3J 1P3, Canada (Email: tjlsless@yorku.ca).

Nagoya Protocol: The authors attest that all legal and regulatory requirements, including export and import collection permits, have been followed for the collection of specimens from source populations at any international, national, regional, or other geographic level for all relevant field specimens collected as part of this study.

Subject Editor: Jessica Ware

Received on 13 July 2023; revised on 15 November 2024; accepted on 19 March 2025

The apid subfamily Nomadinae is the oldest and most diverse clade of brood parasitic bees. Through the incorporation of data from a variety of sources, we generated the most detailed and taxonomically complete phylogeny of this group to date. Despite differing amounts of genetic data available for different species, the tree topology largely matched with expected relationships based on previous findings, with 95% of barcode-only taxa placed in taxonomically consistent positions and all tribes recovered as monophyletic. We further carried out divergence time estimation to investigate the evolutionary history of Nomadinae and place the phylogeny along the geological time scale, recovering an estimated age of 99 Ma for the group. Testing for the effect of barcode-only taxa on estimated dates indicated that ages for deep nodes were robust, though the inclusion of such taxa with limited sequence data tended to push shallower nodes towards older dates. Though this approach may not be appropriate for all applications, the potential for integration of cytochrome oxidase DNA barcode sequences with modern phylogenomic (ultraconserved element) sequence data is an encouraging indication that the wealth of previously published data available through sequence repositories retains the capacity to be informative to future phylogenetic studies.

Keywords: bee, phylogenetics, COI, ultraconserved element, brood parasitism.

Introduction

Modern phylogenetic studies have overwhelmingly focused on genome-wide target enrichment techniques, such as ultraconserved elements (Faircloth et al. 2015), however a wealth of existing DNA sequence data for hundreds of species is already available through repositories such as GenBank. Much of these data exist in the form of cytochrome c oxidase I (COI) barcodes, based on a methodology now entering its third decade of use (Hebert et al. 2003), which has been developed for numerous applications beyond its original scope of species identification (Adamowicz et al. 2019). In a recent study, Talavera et al. (2022) proposed a technique for integrating single-locus COI barcode data with larger multilocus or phylogenomic datasets in order to expand the taxonomic sampling available for a given phylogeny. Subsequent studies have successfully used this technique to produce phylogenies with maximal taxonomic inclusivity in frogs (Portik et al. 2023), butterflies and moths (Lukhtanov et

al. 2023, Zahiri et al. 2023), and rodents (Montgelard et al. 2023). In bees, previous work has demonstrated the incorporation of COI barcodes and small numbers of targeted nuclear loci (eg, Trunz et al. 2016), but this concept remains to be explored for larger UCE datasets.

Over the course of their 120-million-year history, bees have evolved many variations on the central theme of pollinivory, including diverse plant associations, techniques for building nests, and social interactions (Michener 2007). One noteworthy strategy demonstrated by bees is brood parasitism, which involves the exploitation of food and nest resources from other host bee species. This behavior has evolved independently about 20 times in bees, and is proportionally more common in this group of insects than any comparably-sized clade of animals (Sless et al. 2023). Among the many independent clades of parasitic bees, by far the most species is the subfamily Nomadinae, containing over 1,500 species divided into approximately 60 genera and 17 tribes. This broadly successful group, the most diverse subfamily of Apidae, can be found anywhere with suitable hosts to exploit, including all continents except Antarctica. It is also noteworthy that Nomadinae is the oldest clade of parasitic bees by a clear margin (Litman et al. 2013, Sless et al. 2023), and therefore represents the earliest known instance of this strategy evolving among bees and possibly Hymenoptera or even all insects more broadly. Coupled with the fascinating fact that no reversals to a nonparasitic life history have ever been documented within the group, Nomadinae stands out as an important clade for the study of both the ancient origins of brood parasitism as well as the complex and derived adaptations obtained from pursuing such a strategy for tens of millions of years.

The evolutionary relationships of Nomadinae were most recently reconstructed in Sless et al. (2022), using a phylogenomic dataset composed of ultraconserved element (UCE) sequences. The phylogeny produced by this study recovered the tribe Melectini as sister to all remaining Nomadinae, which are further split into two clades: the "ericrocidine line" and "nomadine line" (the latter of these being equivalent to the previously defined Nomadinae sensu stricto used by Michener 2007). While this dataset included a wide breadth of phylogenetic diversity, additional depth of sampling, especially for particularly diverse genera, is also important for a more complete picture of the fine-scale relationships within the group. The concept discussed by Talavera et al. (2022) is well-suited to expanding the phylogeny of the Nomadinae for several reasons. Firstly, the previously published phylogeny of Sless et al. (2022) can serve as a robust "backbone" including genus-level representation of almost all members of Nomadinae. Secondly, previously published COI barcode data are available for many additional nomadine species, which in combination with the existing UCE dataset collectively represent a much broader taxonomic sampling of the clade, particularly at the species level.

In this study, we have reconstructed the phylogeny of the parasitic bee subfamily Nomadinae in unprecedented detail, incorporating essentially all species with existing sequence data at time of writing. Our analyses serve as further validation of the approach of Talavera et al. (2022) and more broadly demonstrate the potential for previously published COI barcode sequences to be integrated with modern phylogenomic datasets. To test the utility and robustness of this new dataset, we conducted divergence time estimation based on fossil calibrations to investigate the evolutionary history of this clade while also affording the opportunity to identify potential biases due to the inclusion of barcode data. The resulting expanded phylogeny reinforces our understanding of the evolutionary relationships within Nomadinae, and furthermore paves the way for more complex analyses than have previously been possible which benefit from maximizing the completeness of taxon sampling.

Materials and Methods

UCE Backbone Alignment

Previously published contigs containing UCE sequence data for Nomadinae were obtained from: Sless et al. (2022), $n = 114$; Odanaka et al. (2022), $n = 89$; Bossert et al. (2020), $n = 4$; and Bossert et al. (2019), $n = 1$, for a total of 208 species (Table S1). Additional UCE contigs for twenty outgroup taxa, representing all tribes of Apidae outside Nomadinae as well as one member of family Megachilidae, were obtained from Branstetter et al. (2017), Bossert et al. (2019), and Freitas et al. (2021).

Contigs were processed with Phyluce (Faircloth 2016) following protocols outlined by Faircloth et al. (2015). Briefly, contigs were matched to probes of the "Hym-v2-bee-ant" UCE bait set. UCE loci were then extracted and trimmed using gblocks (Talavera and Castresana 2007), further cleaned with SpruceUp (Borowiec 2019), and filtered to include loci present in at least 90% of taxa before being concatenated to create the "backbone" alignment.

Barcode Sequence Collection and Final Alignment

The GenBank (Sayers et al. 2022) and BOLD (Ratnasingham and Hebert 2007) databases were searched for COI barcode data representing species of Nomadinae not already included in the UCE dataset described above. Ultimately, barcode sequences were obtained for 150 and 40 new nomadine species from GenBank and BOLD, respectively (Supplementary Table S1). However, some barcode sequences fell outside of Nomadinae in an initial test phylogeny and were dropped from the alignment after confirming with BLAST (Boratyn et al. 2013) that they belonged to other groups, leaving a total of 187 barcode taxa.

To incorporate these barcode-only taxa with the backbone alignment of UCE taxa, COI barcodes also needed to be obtained from the latter group. For most UCE taxa, the COI sequence was extracted from contigs with MitoFinder (Allio et al. 2020), using the *Thyreus decorus (Smith 1852)* mitogenome as a template (GenBank accession NC_057192.1). In cases where MitoFinder was unsuccessful in extracting the full barcode sequence, subsequent attempts were made using Phyluce's "match_barcodes_to_contigs" script and/or local BLAST (Camacho et al. 2009) searches through the UCE contigs, using *Apis mellifera L. 1758* COI barcodes as query sequences. For a small number of cases where the resulting extracted COI barcode sequence was still incomplete, as well as some outgroups with UCEs extracted from genomes or transcriptomes that did not include any detectable mitochondrial sequence, COI barcodes were instead obtained from GenBank and grafted onto the UCE sequences in the backbone alignment.

UCE contigs containing COI barcodes were aligned with the sequences from barcode-only taxa using MAFFT (Katoh and Standley 2013), and subsequently trimmed down to the canonical 658 bp barcode region. Specifically, trimAl (Capella-Gutiérrez et al. 2009) was used to remove all columns of the alignment with <75% of taxa represented, which effectively removed any sections of contigs outside the barcode region and closed gaps within it. These UCE-derived COI barcodes were then concatenated with the backbone UCE loci using AMAS (Borowiec 2016), and the sequences from barcode-only taxa were appended to this alignment with terminal blank spaces added to bring their length in line with the concatenated UCE sequences.

Phylogenetic Analyses

An initial test phylogeny was generated with IQ-TREE v2.1.2 (Minh et al. 2020) using the unpartitioned "backbone + barcode" alignment to identify any potentially problematic barcode sequences. Subsequent analyses partitioned the alignment by the Sliding-Window Site Characteristic Entropy method implemented in PFinderUCE-SWSC-EN (Tagliacollo and Lanfear 2018), though the COI barcode locus was retained as a single partition when included. We first created a backbone-only phylogeny based on an alignment of UCE taxa and loci alone, before progressing to partitioned analyses of the combined dataset. The "-m TESTNEWMERGE" and "-rclusterf 10" options in IQ-TREE were used to group partitions with similar characteristics, and all runs were conducted with 1000 iterations of both ultrafast bootstrap (Hoang et al. 2018) and approximate likelihood ratio tests (aLRT) (Guindon et al. 2010) to assess branch support. Phylogenetic trees were visualized and exported for figures using FigTree v1.4.4 (Rambaut 2018) as well as the R packages phytools (Revell 2012) and ape (Paradis and Schliep 2019).

Divergence Time Estimation

We used MCMCtree as implemented in PAML v4.9j (Yang 2007, dos Reis and Yang 2011) to approximate divergence times for our main phylogeny. Due to computational constraints, we reduced the length of the alignment by focusing on a subset of backbone UCE loci. To accomplish this, gene trees were first generated for individual UCE loci with the "-S" option in IQ-TREE and rooted to the most distant available outgroup taxon using the "pxrr" function of the software phyx (Brown et al. 2017). Loci were then assessed for clocklikeness and bipartition support with SortaDate (Smith et al. 2018) by comparing to a pruned version of the main phylogeny retaining only the backbone UCE taxa. A reduced alignment was then created by concatenating the 50 UCE loci with highest bipartition support. Though this does represent a significant reduction in the size of the dataset, recent work (including on some members of Nomadinae) has suggested that filtering for a smaller number of loci with desirable characteristics in this way can still recover accurate divergence time estimate (Freitas et al. 2022, Almeida et al. 2023, Straka et al. 2024). We chose to filter loci based on bipartition support rather than clocklikeness, as the latter metric can be inaccurate in some cases due to favoring loci with relatively few phylogenetically informative sites (Chen et al. 2021, Koch et al. 2021).

To calibrate the dating analysis in MCMCtree, we used two secondary priors based on Almeida et al. (2023) for Apidae as well as the root of the tree (Apidae+Megachilidae), as well as five fossil calibrations (Supplementary Table S2). MCMCtreeR (Puttick 2019) was used to generate normal distributions for the secondary calibrations and skew-normal distributions for the primary calibrations. We used BASEML (Yang 2007) to estimate an overall substitution rate for the alignment, which was used to inform the "rgene_gamma" parameter of MCMCtree. The burn-in duration was set to 500,000 iterations with a total of 25,000 samples taken every 100 iterations, and four independent runs were conducted independently before assessing the outputs for convergence and merging. MCMCtree runs were assessed for convergence using Tracer v1.7.2 (Rambaut et al. 2018) and by measuring correlation of estimated parameter values across runs. For each independent run, a Hessian matrix was generated with MCMCtree in "usedata = 3" mode before carrying out the main run in "usedata = 2" mode. This entire process was repeated twice in order to compare divergence time estimates obtained from the backbone-only alignment containing just UCE taxa and sequences with the full backbone + barcode alignment of all taxa and including COI barcode sequences.

Results

Integration of Barcode and UCE Sequence Alignments

After generating our initial test phylogeny, three barcode sequences from BOLD (*Sphecodopsis vespericena* Eardley 1997, KBGPE177-18; *Thyreus somalicus (Strand 1912)*, KBGPE130-18; and *Thyreus vachali* (Friese 1802), KBGPE131-18) were recovered within outgroups rather than Nomadinae as expected. BLAST searches of these sequences indicated that they contained DNA from unknown colletid, megachilid, and allodapine bees, respectively, and they were subsequently dropped from the alignment due to evident mislabeling or contamination. Barcode taxa were distributed roughly in proportion to the diversity of each tribe. The majority of the 187 newly added species belonged to the two most diverse tribes, Nomadini ($n = 62$) and Epeolini ($n = 76$), which together account for about two-thirds of nomadine diversity (Ascher and Pickering 2020), while several small tribes were already fully represented by UCE taxa and did not receive any additional barcode sequences. Collectively, the 395 ingroup species included in our dataset represent nearly a quarter (23.8%) of all currently recognized species in the subfamily Nomadinae (Ascher and Pickering 2020).

Barcode sequences were successfully obtained from all of our 228 UCE taxa except for *Nomada stigma* Fabricius 1804 and the four outgroup taxa with UCEs extracted from genome assemblies lacking any mitochondrial content. Including these, a total of 26 UCE taxa with incomplete or missing COI barcode sequences instead used alternate sequences for the same species previously published on GenBank. The full alignment consisted of 1245 UCE loci with a combined length of 286,340 bp as well as the 658-bp COI barcode sequence, for a final length of 286,998 bp. The alignment included 228 "backbone" species (including 20 outgroups) as well as 187 "barcode" species, which only included data for COI, for a total of 415 taxa. The final alignment contained almost exactly half (49.9%) missing data, the vast majority of which was accounted for by the absence of UCE sequences for barcode-only taxa. The reduced alignment used for divergence time estimation consisting of the top 50 UCE loci identified by the bipartition support metric had a total length of 19,239 bp (18,581 bp without COI barcode sequence) and similar missing data proportion of 49.7% (11.2% without COI sequence and barcode-only taxa).

Expanded Phylogeny of Nomadinae

Overall, the phylogenies generated from this alignment matched closely with expectations from previously published studies of Nomadinae. We focus on the main results of the phylogeny generated with SWSC-EN partitioning (Fig. 1). The 3,310 initial partitions identified by SWSC-EN were merged into a final set of 117 partitions by IQ-TREE, with the COI barcode sequence remaining in a partition by itself. Substitution models selected for partitions varied, but the majority used either a transversion (TVM), general time reversible, or transition (TIM) model. Aside from the addition of new taxa, the topology of Nomadinae in our main tree perfectly matched that of the tree based only on UCE taxa and loci, which differed only at a single outgroup node (Supplementary Fig. S2). Tribal relationships were entirely consistent with Bossert et al. (2020) and Sless et al. (2022), with the exception of one barcode sequence for *Rhinepeolus rufiventris* (Friese 1908) which appeared within tribe

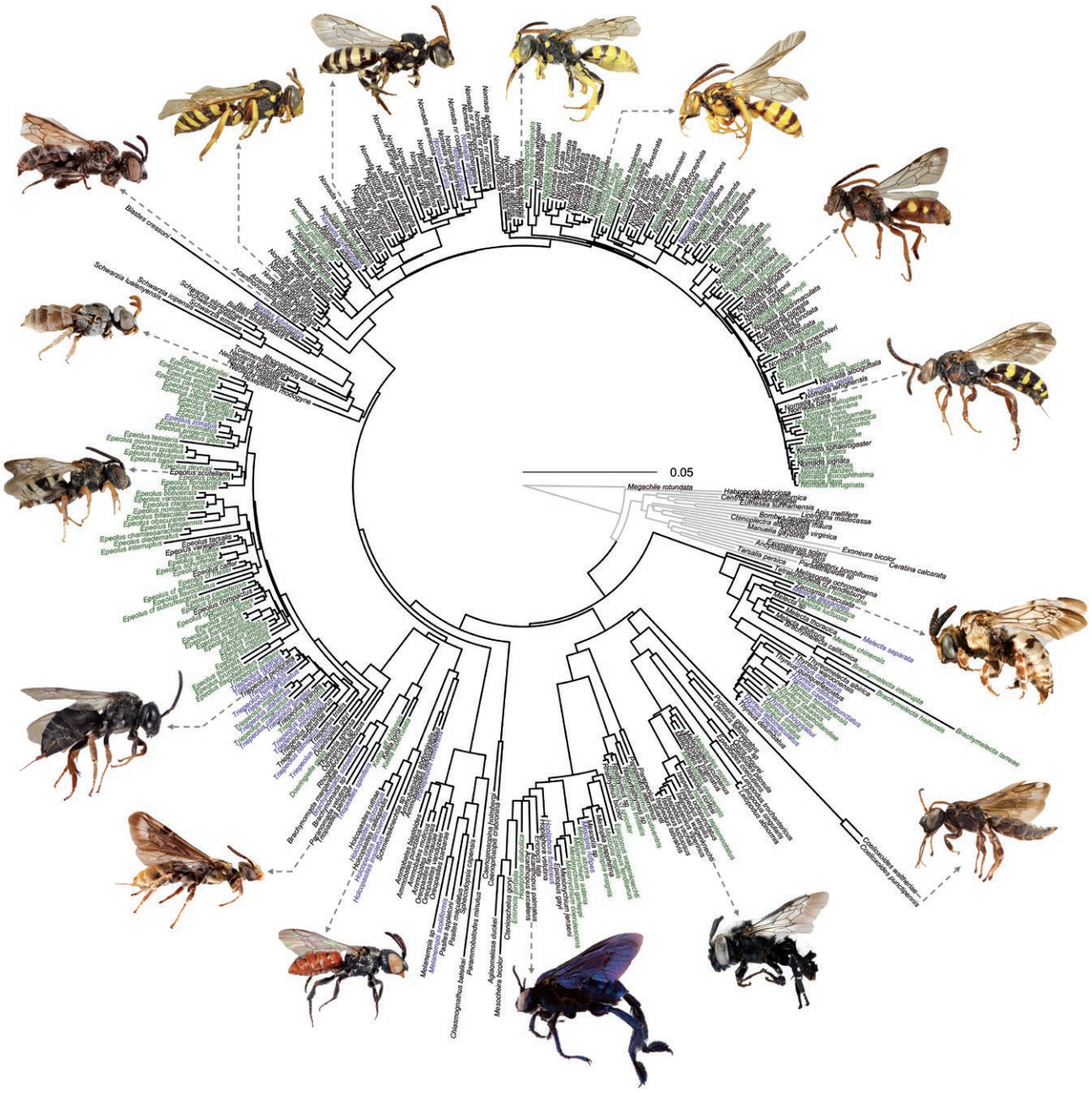

Fig. 1. Maximum likelihood phylogeny of 395 species of Nomadinae and 20 outgroup taxa from Apidae and Megachilidae (gray branches), generated from the combined alignment of backbone and barcode taxa with IQ-TREE 2. Tips are color-coded according to data type and source: black = backbone UCE taxa, green = GenBank barcodes, blue = BOLD barcodes. Scale bar for branch lengths (middle) indicates expected number of nucleotide substitutions per site. Images depict the following species (not to scale) clockwise from middle right: *Zacosmia maculata* (Cresson 1879), *Coelioxoides waltheriae* (Ducke 1908), *Isepeolus wagenknechti* Toro & Rojas 1968, *Acanthopus* sp. Klug 1807, *Holcopasites calliopsidis* (Linsley 1943), *Paranomada velutina* Linsley 1939, *Triepeolus pectoralis* (Robertson 1897), *Epeolus scutellaris* Say 1824, *Neolarra vigilans* (Cockerell 1895), *Biastes* (*Neopasites*) sp. Ashmead 1898, *Nomada vincta* Say 1837, *Nomada mutans* Cockerell 1910, *Nomada edwardsii* Cresson 1878, *Nomada luteoloides* Robertson 1895, *Nomada pygmaea* Cresson 1863, and *Nomada vicina* Cresson 1863. Images courtesy of Laurence Packer.

Ammobatoidini rather than Epeolini as expected. Briefly, Melectini forms the sister group to all other Nomadinae, which then split into the two lineages containing eight tribes each: the ericrocidine line (Coelioxoidini, Osirini, Epeoloidini, Protepeolini, Isepeolini, Parepeolini, Rhathymini, and Ericrocidini) and the nomadine line (Caenoprosopidini, Ammobatini, Ammobatoidini, Brachynomadini, Epeolini, Hexepeolini, Neolarrini, and Nomadini).

Below the tribe level, some genera as currently recognized by Ascher and Pickering (2020) were not recovered as monophyletic. Namely: *Melecta* Latreille 1802 and *Brachynomada* Holmberg 1886 are each rendered paraphyletic due to the placement of *Brachymelecta* Linsley 1939 and *Paranomada* Linsley & Michener 1937 respectively within them (Fig. 1; Supplementary Fig. 1); the barcode-only specimen *Tetralonioidella himalayana* (Bingham

1897) formed a polytomy rather than unambiguously grouping with the other *Tetralonioidella* specimen; individual barcode specimens of *Melectoides niveiventris* (Friese 1925) and *Kelita* aff. *chilensis* (Friese 1916) grouped near but not within the clades defined by other members of their respective genera; and finally several genera of Ericrocidini (*Epiclopus* Spinola 1851, *Ericrocis* Cresson 1887, *Hopliphora* Lepeletier 1841, and *Mesonychium* Lepeletier & Audinet-Serville 1825) were rendered paraphyletic due to the likely erroneous placements of *Epiclopus gayi* Spinola 1851, *Ericrocis pintada* Snelling & Zavortink 1984, and *Hopliphora diabolica* (Friese 1900).

Within the genus *Nomada* Scopoli 1770, which represents nearly half of all species in Nomadinae, our topology closely follows that of Odanaka et al. (2022), and is also consistent with the recent subgeneric classification of Straka et al. (2024). All species groups other than the traditional *ruficornis* group sensu Alexander (1994) are recovered as monophyletic, with the exception of the barcode-only samples *N. nobilis* Herrich-Schäffer 1839 and *N. atroscutellaris* Strand 1921 which both appear in the main *ruficornis* group clade (= *Nomada* sensu stricto in Straka et al. 2024) in our tree rather than the *basalis* group (= *Holonomada*) or *armata* + *trispinosa* group (= *Gestamen*) as expected. As reported by Lim and Lee (2023), *N. emarginata* Morawitz 1877 and *flavopicta* (Kirby 1802) do not group with the main *ruficornis* clade in our phylogeny, although we recover these two tips as the closest relatives of the *belfragei* species group (= *Phelonomada*) rather than the *basalis* group. Additionally, we recover the same clade of *N. ginran* Tsuneki 1973, *kaguya* Hirashima 1953, *aswensis* Tsuneki 1973, and *taicho* Tsuneki 1973 as identified by Lim and Lee (2023), but in our tree this group forms the sister clade to the *armata* + *trispinosa* + *furva* clade (*Gestamen* + *Mininomada* in Straka et al. 2024) rather than being nested within the *armata* species group.

In summary, barcode taxa generally integrated well with UCE taxa, showing no obvious clustering by data source/type (Fig. 1), and though some appeared in likely erroneous locations, the vast majority (178/187, 95%) were recovered in positions consistent with current taxonomy. Branch support as indicated by ultrafast bootstrap and aLRT values was broadly high for all nodes informed by the backbone UCE dataset, though unsurprisingly was more variable for nodes leading to barcode taxa, including a few cases where individual barcode sequences created nodes that were essentially not resolvable (Supplementary Fig. S1).

Divergence Time Estimation

We carried out dating analyses using MCMCtree on the subsampled alignment of the 50 UCE loci with highest bipartition support for both the full set of taxa and a second set limited to taxa with UCE sequence data (Fig. 2; Supplementary Figs. 3 and 4). Though analysis of the latter dataset was interrupted after approximately 4.2 million iterations, both datasets reached a high degree of convergence in final parameter values across runs (all pairwise $R^2 > 0.999$) and achieved effective sample size (ESS) scores greater than 150 for all parameters except for two nodes in the full taxon set with ESS values of 148 and 149, respectively. The crown age of Nomadinae was estimated at 99.44 Ma with the full taxon set (95% HPD 89.86 to 108.75) vs. 100.92 Ma in the UCE-only tree (95% HPD 91.62 to 110.04). Median ages for major subgroups of Nomadinae in the full dataset were approximately 80.3 Ma for Melectini (95% HPD 62.9 to 97.3), 88.6 Ma for the ericrocidine line (95% HPD 76.7 to 100.4), and 92.8 Ma for the nomadine line (95% HPD 82.7 to 102.5). Corresponding ages in the UCE-only tree without barcodes were 64.8 Ma (95% HPD 44.2 to 86.4) for Melectini, 89.4 Ma (95% HPD 78.1 to 100.1) for the ericrocidine line, and 93.0 Ma (95% HPD 83.2 to 102.8) for the nomadine line. The tribe Nomadini, representing nearly half of the diversity of Nomadinae as a whole, was recovered with a crown age of 64.7 Ma (95% HPD 52.2 to 78.3) in the full dataset, compared to 60.1 Ma (95% HPD 46.2 to 73.4) with UCEs only.

In direct comparison, both datasets tended to agree well with respect to ages near the root of the tree, but the full dataset incorporating both barcode and backbone UCE taxa consistently recovered older dates than the UCE-only dataset for later-diverging nodes (Supplementary Fig. S4). This effect was particularly noticeable in clades with a higher proportion of barcode-only taxa. For example, 22/32 (68.75%) of taxa within Melectini are represented by barcode sequences, and this group shows a gap of ~15 Ma between the estimated ages from both datasets (Supplementary Figs. S3 and S4). The tribe Epeolini, which had the highest proportion of barcode-only taxa at 89/103 (86.4%), similarly showed a disparity of approximately 20 million years in the age estimates from the two datasets (67 Ma vs. 47 Ma, respectively).

Discussion

Utility of DNA Barcodes in Modern Phylogenetics

Ultimately, the combination of UCE backbone taxa with barcode-only taxa performed satisfactorily in reconstructing an expanded phylogenetic framework of the subfamily Nomadinae. This finding provides additional support to those of Talavera et al. (2022) and other recent studies that have used similar methods (eg, Lukhtanov et al. 2023, Montgelard et al. 2023, Portik et al. 2023, Zahiri et al. 2023), suggesting that DNA barcode data can complement modern phylogenomics. However, several important caveats need to be considered for the proper integration of such sequences with ultraconserved element datasets.

As noted by Talavera et al. (2022), a robust backbone consisting of species with the full complement of sequence data in the alignment is an important prerequisite. By including at least one such representative per genus and carefully selecting barcode-only taxa to be interspersed among them, the deep relationships of the phylogeny may be recovered much more accurately than if backbone and barcode taxa were simply distributed randomly throughout the tree. The overall proportion of barcode-only taxa in the matrix should be minimized, since each of these samples contains on the order of 99% missing data and thus reduces the integrity of the whole alignment to some extent.

Additionally, repeated phylogenetic reconstruction analyses with varied parameters may be desirable to assist in identifying particularly unreliable COI sequences. Even after removing some specimens with evidently mislabeled or contaminated barcodes as detailed in the methods, a small number of barcode taxa consistently appeared in unexpected locations throughout the tree, but did not consistently show the same placements across multiple tested phylogenies. For example, *Rhinepeolus rufiventris* appeared within Ammobatoidini in our main tree as mentioned above, but in other attempts was recovered at different positions within Nomadinae or even within outgroup clades, despite the fact that this barcode sequence was complete and matched other members of Epeolini (as expected) when analyzed with BLAST. Though such problematic specimens are difficult to identify a priori, it may be necessary to exclude them from further downstream analyses, such as ancestral state reconstructions, which can be disproportionately influenced by misplaced tips.

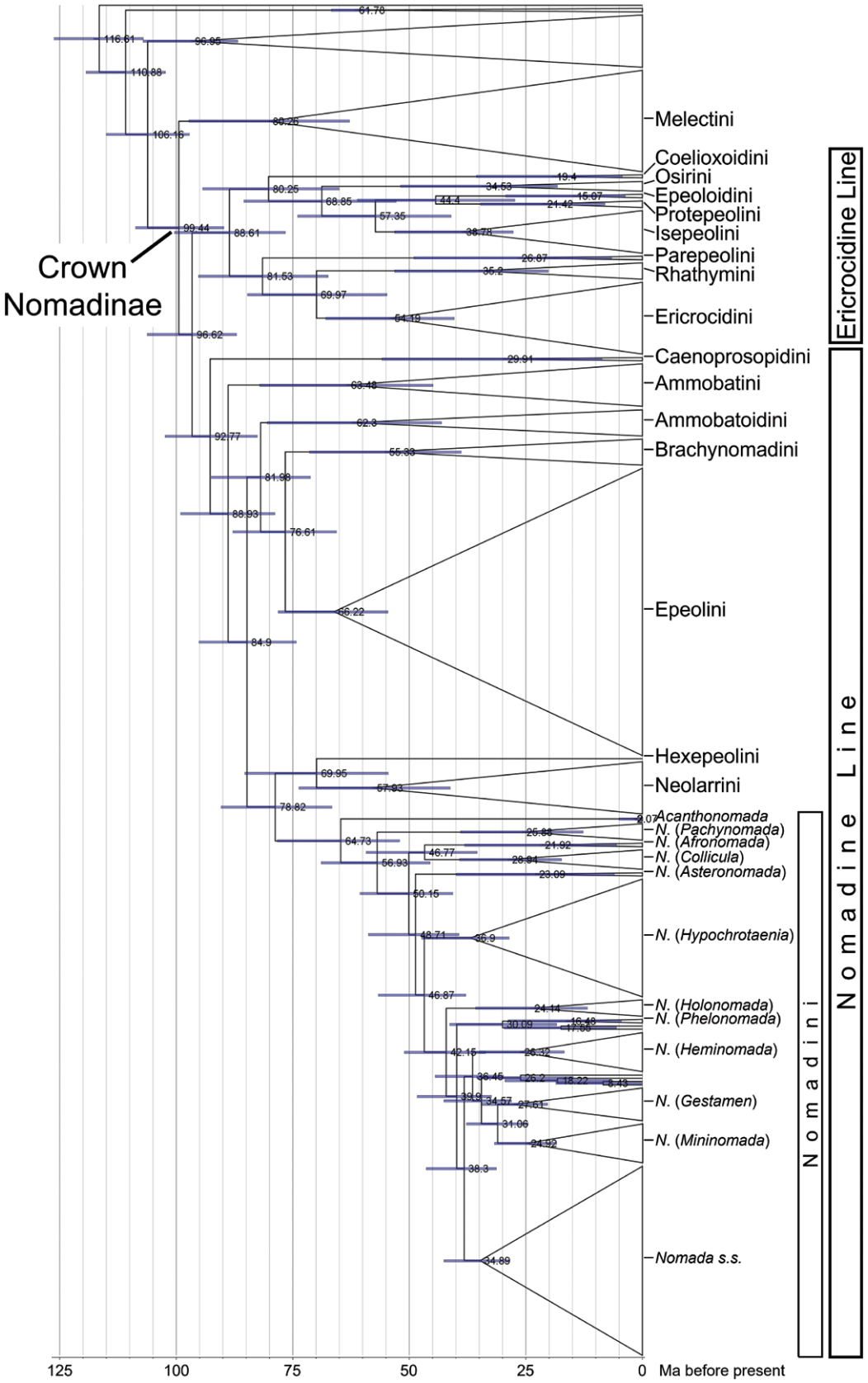

Fig. 2. Time-calibrated phylogeny of Nomadinae based on alignment of top 50 bipartition support UCE loci and COI barcode locus analyzed with MCMCtree. Nodes are collapsed for simplicity, showing tribes and subgenera of *Nomada* (following Straka et al. 2024), see Supplementary Fig. S3 for full tree. Numbers at nodes indicate median age in millions of years, and bars indicate 95% HPD intervals for estimated divergence times.

Other than these few problematic barcodes, however, our phylogeny was generally consistent with existing taxonomy—or, when it differed, was consistent with previous studies indicating taxa in need of revision. *Melecta* was rendered paraphyletic in our tree by the placement of *Brachymelecta* as reported above, but this finding is consistent with previous UCE-only phylogenies (Sless et al. 2022, Orr et al. 2024) and indicates that this genus is in need of further revision. The recovery of *Paranomada* nested within *Brachynomada* was also reported by Sless et al. (2022), and Michener (2007) previously noted that eventual synonymization of these genera would be likely. Finally, the relationships within Ericrocidini are also somewhat unclear, as evidenced by a four-way polytomy among subgroups recovered by Martins et al. (2018) and previous reports of paraphyly for *Hopliphora* (Lim et al. 2022).

The consistent recovery of these particular genera as paraphyletic across multiple studies would suggest that this is actually reflective of their taxonomic validity, rather than due to any artifact of barcode-only taxa included herein. Ultimately, many of the fine-scale relationships among barcode and UCE species as depicted in our tree cannot be directly evaluated by comparison to previous studies due to the greater taxonomic resolution included herein. The placement of some taxa currently represented only by barcodes may change in light of new data, but nevertheless our topology should be considered and evaluated as the most taxonomically complete hypothesis of nomadine relationships to date.

Divergence Time Estimation with Combined Datasets

Our analyses of divergence time estimation on this dataset further suggest that this method can be used to generate trees that are suitable for phylogenetic dating but again highlight some points of caution. Overall, the full and UCE-only datasets showed close agreement with respect to the ages inferred for the majority of basal nodes in the tree, as well as similar widths for 95% HPD intervals. This finding is encouraging and suggests that the inclusion of taxa represented by barcode sequences alone does not strongly shift divergence time estimates for deep nodes, nor does it noticeably increase uncertainty for these estimates. Simultaneously, however, the inclusion of such taxa clearly pushes back the ages of shallower nodes and does so in a manner approximately consistent with the proportion of barcode-only taxa represented in a given clade.

In direct comparison with previous studies, the estimated ages of divergence we recover for Nomadinae and its subgroups are somewhat older than average, though generally still within estimated probability distributions. Our estimated crown age of approximately 100 Ma for the subfamily is 15 million years more ancient than the dates recovered by Almeida et al. (2023) and Cardinal et al. (2018) respectively, though is in closer agreement with some older studies (eg, 95 Ma in Cardinal et al. 2010, 111 Ma in Litman et al. 2013). As previously discussed, the crown age we recovered for Epeolini was strongly influenced by the presence or absence of barcode taxa. Interestingly, the ~38 Ma age of this tribe reported by (Onuferko et al. 2019) is much closer to the age recovered in our dating analysis based on UCE taxa alone (47 Ma) than the age for our full taxon set (65 Ma), despite the fact that most of the *Epeolus* barcode sequences included in the latter analysis came from this publication. With respect to the tribe Nomadini, our recovered age of 60-65 Ma is again somewhat older than the ~48 Ma crown age of Straka et al. (2024), though is in close agreement with the 65 Ma date of Odanaka et al. (2022).

These disparities may be explained in part by different methodologies used for the divergence time estimation, which was based on BEAST and MrBayes in the above studies rather than MCMCtree as in our case. The absence of fossil calibrations as priors in Onuferko et al. (2019) may also explain some of the difference. Conversely, it is encouraging that our results for the age of Nomadini (both with and without the inclusion of barcode taxa) match well with those of Odanaka et al. (2022), who also used MCMCtree. However, there is substantial overlap in datasets between this study and ours, making it difficult to isolate the effects of specific software on divergence time estimates. The mitochondrial origins of the COI barcode sequence in contrast to nuclear DNA may also contribute to the difference in ages between our two analyses. Our findings that the inclusion of barcode sequences and barcode-only taxa inflated certain node ages echo those of van Tuinen and Torres (2015), who also identified that mitochondrial data resulted in older age estimates than nuclear DNA and that this effect was particularly pronounced at shallower nodes. However, Talavera et al. (2022) did not identify any such bias in divergence times when comparing their backbone and combined datasets.

Conclusion

In summary, we present a new and expanded reconstruction of the phylogeny of the subfamily Nomadinae focused on including the maximum possible taxonomic breadth of sampling based on available sequence data from a multitude of sources. The evolutionary relationships among this group's diverse species are here presented in more detail than previously possible, and our divergence time estimates further provide insights into the ancient origins of this group, the oldest and most diverse clade of brood parasitic bees. The broadly successful integration of COI barcodes with phylogenomic data based on UCEs is encouraging and highlights the potential for such data to retain their usefulness in the modern era of phylogenetics. However, as with any such approach there are caveats that must be considered, including the need for a robust and well-sampled "backbone" of more fully sequenced taxa to ensure a well-supported tree as well as the potential for the inclusion of barcode taxa to influence divergence time estimation analyses.

Acknowledgements

We thank Laurence Packer for the use of bee images included in Fig. 1. We are grateful to the reviewers and editors for providing constructive feedback and suggestions.

Author contributions

Trevor Sless (Conceptualization [lead], Data curation [lead], Formal analysis [lead], Investigation [lead], Methodology [lead], Writing—original draft [lead], Writing—review & editing [lead]), Katherine Odanaka (Data curation [supporting], Formal analysis [supporting], Writing—review & editing [supporting]), L. Ximena Alva-Caballero (Data curation [supporting], Formal analysis [supporting], Writing—review & editing [supporting]), Jeremy Searle (Conceptualization [supporting], Supervision [equal], Writing—review & editing [supporting]), Bryan Danforth (Conceptualization [supporting], Funding acquisition [equal], Supervision [equal], Writing—review & editing [supporting]), and Sandra Rehan (Conceptualization [supporting], Funding acquisition [equal], Supervision [equal], Writing—review & editing [supporting])

Supplementary material

Supplementary material is available at *Insect Systematics and Diversity* online.

Funding

While at Cornell University, TJLS was supported by an NSERC PGS-D fellowship. KAO was supported by an Ontario Graduate Scholarship, Mitacs training fellowship, Carswell scholarship, and Susan Mann Dissertation Fellowship. We would like to thank the NGO Research Experience for Peruvian Undergraduates (REPU) and FONDECYT for financing the internship of LXA-C at Cornell University. We also acknowledge the support of the Government of Canada's New Frontiers in Research Fund (NFRF) [NFRFT-2020-00073]. Additional funding was provided by Foundation for Food and Agriculture Research grant # 549038, NSERC Discovery Grants, Supplements and an EWR Steacie Memorial Fellowship to SMR. In addition, this work was partially supported by a U.S. National Science Foundation grant (DEB-1555905) to BND.

Conflicts of interest. None declared.

Data availability

Alignment files, phylogenetic trees, and other supplementary material produced for this study can be accessed from the following Data Dryad repository: https://doi.org/10.5061/dryad.12jm63z5d.

References

Adamowicz SJ, Boatwright JS, Chain F, et al. 2019. Trends in DNA barcoding and metabarcoding. Genome 62:5–8. https://doi.org/10.1139/gen-2019-0054

Alexander BA. 1994. Species-groups and cladistic analysis of the cleptoparasitic bee genus *Nomada* (Hymenoptera: Apoidea). Univ. Kans. Sci. Bull. 55:175–238.

Allio R, Schomaker-Bastos A, Romiguer J, et al. 2020. MitoFinder: efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics. Mol. Ecol. Resour. 20:892–905. https://doi.org/10.1111/1755-0998.13160

Almeida EAB, Bossert S, Danforth BN, et al. 2023. The evolutionary history of bees in time and space. Curr. Biol. 33:3409–3422.e6. https://doi.org/10.1016/j.cub.2023.07.005

Ascher JS, Pickering J. 2020. Discover life bee species guide and world checklist (Hymenoptera: Apoidea: Anthophila). http://www.discoverlife.org/mp/20q?guide=Apoidea_species

Boratyn GM, Camacho C, Cooper PS, et al. 2013. BLAST: a more efficient report with usability improvements. Nucleic Acids Res. 41:W29–W33. https://doi.org/10.1093/nar/gkt282

Borowiec ML. 2016. AMAS: a fast tool for alignment manipulation and computing of summary statistics. PeerJ 4:e1660. https://doi.org/10.7717/peerj.1660

Borowiec ML. 2019. Spruceup: fast and flexible identification, visualization, and removal of outliers from large multiple sequence alignments. J. Open Source Softw. 4:1635. https://doi.org/10.21105/joss.01635

Bossert S, Murray EA, Almeida EAB, et al. 2019. Combining transcriptomes and ultraconserved elements to illuminate the phylogeny of Apidae. Mol. Phylogenet. Evol. 130:121–131. https://doi.org/10.1016/j.ympev.2018.10.012

Bossert S, Copeland RS, Sless TJL, et al. 2020. Phylogenomic and morphological reevaluation of the bee tribes Biastini, Neolarrini, and Townsendiellini (Hymenoptera: Apidae) with description of three new species of *Schwarzia*. Insect Syst. Div. 4:1. https://doi.org/10.1093/isd/ixaa013

Branstetter MG, Danforth BN, Pitts JP, et al. 2017. Phylogenomic insights into the evolution of stinging wasps and the origins of ants and bees. Curr. Biol. 27:1019–1025. https://doi.org/10.1016/j.cub.2017.03.027

Brown JW, Walker JF, Smith SA. 2017. Phyx: phylogenetic tools for Unix. Bioinformatics 33:1886–1888. https://doi.org/10.1093/bioinformatics/btx063

Camacho C, Coulouris G, Avagyan V, et al. 2009. BLAST+: architecture and applications. BMC Bioinf. 10:421. https://doi.org/10.1186/1471-2105-10-421

Capella-Gutiérrez S, Silla-Martinez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Cardinal S, Straka J, Danforth BN. 2010. Comprehensive phylogeny of apid bees reveals the evolutionary origins and antiquity of cleptoparasitism. Proc. Natl. Acad. Sci. USA 107:16207–16211. https://doi.org/10.1073/pnas.1006299107

Cardinal S, Buchmann SL, Russell AL. 2018. The evolution of floral sonication, a pollen foraging behavior used by bees (Anthophila). Evolution 72:590–600. https://doi.org/10.1111/evo.13446

Chen D, Hosner PA, Dittmann DL, et al. 2021. Divergence time estimation of Galliformes based on the best gene shopping scheme of ultraconserved elements. BMC Ecol. Evol. 21:209. https://doi.org/10.1186/s12862-021-01935-1

dos Reis M, Yang Z. 2011. Approximate likelihood calculation on a phylogeny for Bayesian estimation of divergence times. Mol. Biol. Evol. 28:2161–2172. https://doi.org/10.1093/molbev/msr045

Faircloth BC. 2016. PHYLUCE is a software package for the analysis of conserved genomic loci. Bioinformatics 32:786–788. https://doi.org/10.1093/bioinformatics/btv646

Faircloth BC, Branstetter MG, White ND, et al. 2015. Target enrichment of ultraconserved elements from arthropods provides a genomic perspective on relationships among Hymenoptera. Mol. Ecol. Res. 15:489–501. https://doi.org/10.1111/1755-0998.12328

Freitas FV, Branstetter MG, Griswold T, et al. 2021. Partitioned gene-tree analyses and gene-based topology testing help resolve incongruence in a phylogenomic study of host-specialist bees (Apidae: Eucerinae). Mol. Biol. Evol. 38:1090–1100. https://doi.org/10.1093/molbev/msaa277

Freitas FV, Branstetter MG, Casali DM, et al. 2022. Phylogenomic dating and Bayesian biogeography illuminate an antitropical pattern for eucerine bees. J. Biogeogr. 49:1034–1047. https://doi.org/10.1111/jbi.14359

Guindon S, Dufayard J-F, Lefort V, et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59:307–321. https://doi.org/10.1093/sysbio/syq010

Hebert PDN, Cywinska A, Ball SL, et al. 2003. Biological identifications through DNA barcodes. Proc. Biol. Sci. 270:313–321. https://doi.org/10.1098/rspb.2002.2218

Hoang DT, Chernomor O, von Haeseler A, et al. 2018. UFBoot2: improving the ultrafast bootstrap approximation. Mol. Biol. Evol. 35:518–522. https://doi.org/10.1093/molbev/msx281

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30:772–780. https://doi.org/10.1093/molbev/mst010

Koch NM. 2021. Phylogenomic subsampling and the search for phylogenetically reliable loci. Mol. Biol. Evol. 38:4025–4038. https://doi.org/10.1093/molbev/msab151

Lim K, Lee S. 2023. New suggestion of the species group reconstruction of genus Nomada Scopoli, 1770 (Hymenoptera, Apidae) from Korea. J. Hymenopt. Res. 96:805–816. https://doi.org/10.3897/jhr.96.106452

Lim K, Lee S, Orr M, et al. 2022. Harrison's rule corroborated for the body size of cleptoparasitic cuckoo bees (Hymenoptera: Apidae: Nomadinae) and their hosts. Sci. Rep. 12:10984. https://doi.org/10.1038/s41598-022-14938-9

Litman JR, Praz CJ, Danforth BN, et al. 2013. Origins, evolution, and diversification of cleptoparasitic lineages in long-tongued bees. Evolution 67:2982–2998. https://doi.org/10.1111/evo.12161

Lukhtanov VA, Shapoval NA, Dantchenko AV, et al. 2023. Phylogenetic structure revealed through combining DNA barcodes with multi-gene data for *Agrodiaetus* blue butterflies (Lepidoptera, Lycaenidae). Insects 14:769. https://doi.org/10.3390/insects14090769

Martins AC, Luz DR, Melo GAR. 2018. Palaeocene origin of the Neotropical lineage of cleptoparasitic bees Ericrocidini-Rhathymini (Hymenoptera, Apidae). Syst. Entomol. 43:510–521. https://doi.org/10.1111/syen.12286

Michener CD. 2007. The bees of the world. Johns Hopkins University Press.

Minh BQ, Schmidt HA, Chernomor O, et al. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 37:1530–1534. https://doi.org/10.1093/molbev/msaa015

Montgelard C, Muller T, Arnal V, et al. 2023. Diversification and evolutionary history of the African laminated-toothed rats (Rodentia, Otomyini). Mol. Phylogenet. Evol. 183:107779. https://doi.org/10.1016/j.ympev.2023.107779

Odanaka KA, Branstetter MG, Tobin KB, et al. 2022. Phylogenomics and historical biogeography of the cleptoparasitic bee genus *Nomada* (Hymenoptera: Apidae) using ultraconserved elements. Mol. Phylogenet. Evol. 170:107453. https://doi.org/10.1016/j.ympev.2022.107453

Onuferko TM, Bogusch P, Ferrari RR, et al. 2019. Phylogeny and biogeography of the cleptoparasitic bee genus Epeolus (Hymenoptera: Apidae) and cophylogenetic analysis with its host bee genus Colletes (Hymenoptera: Colletidae). Mol. Phylogenet. Evol. 141:106603. https://doi.org/10.1016/j.ympev.2019.106603

Orr MC, Chesters D, Williams PH, et al. 2024. Integrative taxonomy of a new species of a bumble bee-mimicking brood parasitic bee, *Tetralonioidella mimetica* (Hymenoptera, Apoidea, Apidae), investigated through phylogenomics. J. Hymenopt. Res. 97:755–780. https://doi.org/10.3897/jhr.97.129470

Paradis E, Schliep K. 2019. Ape 5.0: an environment for modern phylogenetics and evolutionary analysis in R. Bioinformatics 35:526–528. https://doi.org/10.1093/bioinformatics/bty633

Portik DM, Streicher JW, Wiens JJ. 2023. Frog phylogeny: a time-calibrated, species-level tree based on hundreds of loci and 5,242 species. Mol. Phylogenet. Evol. 188:107907. https://doi.org/10.1016/j.ympev.2023.107907

Puttick MN. 2019. MCMCtreeR: functions to prepare MCMCtree analyses and visualize posterior ages on trees. Bioinformatics 35:5321–5322. https://doi.org/10.1093/bioinformatics/btz554

Rambaut A. 2018. FigTree v1.4.4. http://tree.bio.ed.ac.uk/software/figtree/.

Rambaut A, Drummund AJ, Xie D, et al. 2018. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. 67:901–904. https://doi.org/10.1093/sysbio/syy032

Ratnasingham S, Hebert PDN. 2007. BOLD: the barcode of life data system. Mol. Ecol. Notes 7:355–364. https://doi.org/10.1111/j.1471-8286.2007.01678.x

Revell LJ. 2012. Phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. 3:217–223. https://doi.org/10.1111/j.2041-210x.2011.00169.x

Sayers EW, Bolton EE, Brister JR, et al. 2022. Database resources of the national center for biotechnology information. Nucleic Acids Res. 50:D20–D26. https://doi.org/10.1093/nar/gkab1112

Sless TJL, Branstetter MG, Gillung JP, et al. 2022. Phylogenetic relationships and the evolution of host preferences in the largest clade of brood parasitic bees (Apidae: Nomadinae). Mol. Phylogenet. Evol. 166:107326. https://doi.org/10.1016/j.ympev.2021.107326

Sless TJL, Danforth BN, Searle JB. 2023. Evolutionary origins and patterns of diversification in animal brood parasitism. Am. Nat. 202:107–121. https://doi.org/10.1086/724839

Smith SA, Brown JW, Walker JF. 2018. So many genes, so little time: a practical approach to divergence-time estimation in the genomic era. PLoS One 13:e0197433. https://doi.org/10.1371/journal.pone.0197433

Straka J, Benda D, Policarová J, et al. 2024. A phylogenomic monograph of West-Palearctic *Nomada* (Hymenoptera: Apidae). Insect Syst. Divers 8:1. https://doi.org/10.1093/isd/ixad024

Tagliacollo VA, Lanfear R. 2018. Estimating improved partitioning schemes for ultraconserved elements. Mol. Biol. Evol. 35:1798–1811. https://doi.org/10.1093/molbev/msy069

Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst. Biol. 56:564–577. https://doi.org/10.1080/10635150701472164

Talavera G, Lukhtanov V, Pierce NE, et al. 2022. DNA barcodes combined with multilocus data of representative taxa can generate reliable higher-level phylogenies. Syst. Biol. 71:382–395. https://doi.org/10.1093/sysbio/syab038

Trunz V, Packer L, Vieu J, et al. 2016. Comprehensive phylogeny, biogeography, and new classification of the diverse bee tribe Megachilini: can we use DNA barcodes in phylogenies of large genera? Mol. Phylogenet. Evol. 103:245–259. https://doi.org/10.1016/j.ympev.2016.07.004

van Tuinen M, Torres CR. 2015. Potential for bias and low precision in molecular divergence time estimation of the Canopy of Life: an example from aquatic bird families. Front. Genet. 6:203. https://doi.org/10.3389/fgene.2015.00203

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24:1586–1591. https://doi.org/10.1093/molbev/msm088

Zahiri R, Holloway JD, Rota J, et al. 2023. Evolutionary history of Euteliidae (Lepidoptera, Noctuoidea). Syst. Entomol. 48:445–462. https://doi.org/10.1111/syen.12587