


































# The Chalcidoidea bush of life: evolutionary history of a massive radiation of minute wasps

Astrid Cruaud<sup>a,†</sup> , Jean-Yves Rasplus<sup>\*a,†</sup> , Junxia Zhang<sup>b,c,†</sup> , Roger Burks<sup>c</sup> ,  
G rard Delvare<sup>a</sup> , Lucian Fusu<sup>d</sup> , Alex Gumovsky<sup>e</sup> , John T. Huber<sup>f</sup>, Petr Jan sta<sup>g,h</sup> ,  
Mircea-Dan Mitroiu<sup>d</sup> , John S. Noyes<sup>i</sup>, Simon van Noort<sup>j,k</sup> , Austin Baker<sup>c</sup> ,  
Julie B hmov a<sup>g</sup>, Hannes Baur<sup>l,m</sup> , Bonnie B. Blaimer<sup>n</sup> , Se n G. Brady<sup>o</sup> ,  
Krist yna Buben kov a<sup>g</sup>, Marguerite Chartois<sup>a</sup> , Robert S. Copeland<sup>o,p</sup>,  
Natalie Dale-Skey Papilloud<sup>i</sup> , Ana Dal Molin<sup>q</sup> , Chrysalyn Dominguez<sup>c</sup> ,  
Marco Gebiola<sup>c</sup> , Emilio Guerrieri<sup>i,r</sup> , Robert L. Kresslein<sup>c</sup> , Lars Krogmann<sup>h,s</sup> ,  
Emily Lemmon<sup>t</sup>, Elizabeth A. Murray<sup>u</sup> , Sabine Nidelet<sup>a</sup>, Jos  Luis Nieves-Aldrey<sup>v</sup> ,  
Ryan K. Perry<sup>w</sup> , Ralph S. Peters<sup>x</sup> , Andrew Polaszek<sup>i</sup> , Laure Saun e<sup>a</sup>,  
Javier Torr ens<sup>y</sup> , Serguei Triapitsyn<sup>c</sup> , Ekaterina V. Tselikh<sup>z</sup> , Matthew Yoder<sup>z</sup> ,  
Alan R. Lemmon<sup>z</sup>, James B. Woolley<sup>z,†</sup>  and John M. Heraty<sup>\*c,†</sup> 

<sup>a</sup>CBGP, INRAE, CIRAD, IRD, Montpellier SupAgro, Univ Montpellier, Montpellier, France; <sup>b</sup>Key Laboratory of Zoological Systematics and Application of Hebei Province, Institute of Life Science and Green Development, College of Life Sciences, Hebei University, Baoding, Hebei, China; <sup>c</sup>Department of Entomology, University of California Riverside, Riverside, California, USA; <sup>d</sup>Faculty of Biology, Alexandru Ioan Cuza University, Iasi, Romania; <sup>e</sup>Schmalhausen Institute of Zoology, National Academy of Sciences of Ukraine, Kiev, Ukraine; <sup>f</sup>Natural Resources Canada, c/o Canadian National Collection of Insects, Ottawa, Ontario, Canada; <sup>g</sup>Department of Zoology, Faculty of Science, Charles University, Prague, Czech Republic; <sup>h</sup>Department of Entomology, State Museum of Natural History, Stuttgart, Germany; <sup>i</sup>Insects Division, Natural History Museum, London, UK; <sup>j</sup>Research and Exhibitions Department, South African Museum, Iziko Museums of South Africa, Cape Town, South Africa; <sup>k</sup>Department of Biological Sciences, University of Cape Town, Private Bag, Rondebosch 7701, South Africa; <sup>l</sup>Department of Invertebrates, Natural History Museum Bern, Bern, Switzerland; <sup>m</sup>Institute of Ecology and Evolution, University of Bern, Bern, Switzerland; <sup>n</sup>Center for Integrative Biodiversity Discovery, Museum f r Naturkunde, Leibniz Institute for Evolution and Biodiversity Science, Berlin, Germany; <sup>o</sup>Smithsonian Institution, National Museum of Natural History, Washington, DC, USA; <sup>p</sup>International Centre of Insect Physiology and Ecology (ICIPE), Nairobi, Kenya; <sup>q</sup>Departamento de Microbiologia e Parasitologia, Universidade Federal do Rio Grande do Norte, Natal, RN, Brazil; <sup>r</sup>CNR-Institute for Sustainable Plant Protection (CNR-IPSP), National Research Council of Italy, Portici, Italy; <sup>s</sup>Institute of Zoology, University of Hohenheim, Stuttgart, Germany; <sup>t</sup>Department of Biological Science, Florida State University, Tallahassee, Florida, USA; <sup>u</sup>Department of Entomology, Washington State University, Pullman, Washington, USA; <sup>v</sup>Museo Nacional de Ciencias Naturales (CSIC), Jos  Guti rrez Abascal 2, Madrid, Spain; <sup>w</sup>Department of Plant Sciences, California Polytechnic State University, San Luis Obispo, California, USA; <sup>x</sup>Zoologisches Forschungsmuseum Alexander Koenig, Leibniz Institute for the Analysis of Biodiversity Change, Bonn, Germany; <sup>y</sup>Centro Regional de Investigaciones Cient ficas y Transferencia Tecnol gica de La Rioja (CRILAR-CONICET), Anillaco, Argentina; <sup>z</sup>Zoological Institute, Russian Academy of Sciences, St. Petersburg, Russia; <sup>1</sup>Illinois Natural History Survey, University of Illinois, Champaign, Illinois, USA; <sup>2</sup>Department of Scientific Computing, Florida State University, Dirac Science Library, Tallahassee, Florida, USA; <sup>3</sup>Department of Entomology, Texas A&M University, College Station, Texas, USA

Received 17 May 2023; Revised 6 September 2023; Accepted 12 September 2023

## Abstract

Chalcidoidea are mostly parasitoid wasps that include as many as 500 000 estimated species. Capturing phylogenetic signal from such a massive radiation can be daunting. Chalcidoidea is an excellent example of a hyperdiverse group that has remained

\*Corresponding author: E-mail address: [jean-yves.rasplus@inrae.fr](mailto:jean-yves.rasplus@inrae.fr); [john.heraty@ucr.edu](mailto:john.heraty@ucr.edu)

†These authors contributed equally to this work.

recalcitrant to phylogenetic resolution. We combined 1007 exons obtained with Anchored Hybrid Enrichment with 1048 ultra-conserved elements (UCEs) for 433 taxa including all extant families, >95% of all subfamilies, and 356 genera chosen to represent the vast diversity of the superfamily. Going back and forth between the molecular results and our collective knowledge of morphology and biology, we detected bias in the analyses that was driven by the saturation of nucleotide data. Our final results are based on a concatenated analysis of the least saturated exons and UCE datasets (2054 loci, 284 106 sites). Our analyses support an expected sister relationship with Mymarommatoidea. Seven previously recognized families were not monophyletic, so support for a new classification is discussed. Natural history in some cases would appear to be more informative than morphology, as illustrated by the elucidation of a clade of plant gall associates and a clade of taxa with planidial first-instar larvae. The phylogeny suggests a transition from smaller soft-bodied wasps to larger and more heavily sclerotized wasps, with egg parasitism as potentially ancestral for the entire superfamily. Deep divergences in Chalcidoidea coincide with an increase in insect families in the fossil record, and an early shift to phytophagy corresponds with the beginning of the “Angiosperm Terrestrial Revolution”. Our dating analyses suggest a middle Jurassic origin of 174 Ma (167.3–180.5 Ma) and a crown age of 162.2 Ma (153.9–169.8 Ma) for Chalcidoidea. During the Cretaceous, Chalcidoidea may have undergone a rapid radiation in southern Gondwana with subsequent dispersals to the Northern Hemisphere. This scenario is discussed with regard to knowledge about the host taxa of chalcid wasps, their fossil record and Earth’s palaeogeographic history.

© 2023 The Authors. *Cladistics* published by John Wiley & Sons Ltd on behalf of Willi Hennig Society.

## Introduction

Chalcidoidea (jewel wasps, chalcidoid wasps, hereafter called chalcid wasps) are among the most species-rich, ecologically important, biologically diverse and morphologically disparate groups of terrestrial organisms. These minute wasps (mostly 0.5–2 mm in size) are ubiquitous in almost every terrestrial habitat on Earth. Their diversity is staggering with 2731 genera and 27 021 species placed in 27 families and 87 subfamilies (Table S1a) with estimates of >500 000 species (Heraty, 2009; Noyes, 2019). Although most species are parasitoids, phytophagous species are known from 11 families (Böhmová et al., 2022; Burks et al., 2022). Their animal host range includes 13 insect orders, spiders, ticks, mites, pseudoscorpions and even gall-forming nematodes (Austin et al., 1998; Gibson et al., 1999). Chalcid wasps attack all life stages of their hosts from eggs to adults, as internal or external parasitoids, and they can be primary parasitoids, hyperparasitoids (parasitoids of parasitoids) or even tertiary parasitoids (parasitoids of hyperparasitoids). The economic importance of Chalcidoidea in pest management is unparalleled and they are widely used in biological control programmes against major pests throughout the world (Noyes and Hayat, 1994; Heraty, 2009).

A few studies have addressed higher-level relationships within Chalcidoidea, although with only a sparse sampling of genes (Sanger data sets) or taxa (NGS datasets). Munro et al. (2011) used 18S + 28S ribosomal DNA for 649 species of Chalcidoidea in 19 families and 343 genera; Heraty et al. (2013) used a combination of 18S + 28S + morphology for 283 species in 19 families and 268 genera; Peters et al. (2018) analysed 3239 genes from transcriptomes for 37 species in 16 families and 35 genera; and Zhang et al. (2020) extended the transcriptome dataset to 5591 genes for 55 species in 17 families and 48 genera.

Despite these efforts, the higher-level relationships of Chalcidoidea remain largely unresolved. Genome-scale data (transcriptomes) have proven particularly frustrating, presumably because of the lack of signal associated with an old, rapid radiation and the increasing probability of observing conflicting signals between markers (Peters et al., 2018; Zhang et al., 2020). When chalcid wasps are included in studies of the Hymenopteran tree-of-life, conflicts or lack of signal that are reflected in poor statistical support of (some) nodes are highlighted in all datasets: Branstetter et al. (2017) 854 ultra-conserved elements (UCEs), nine species of Chalcidoidea in nine families and nine genera; Peters et al. (2017) 3256 protein coding genes, six species in six families and six genera; and Tang et al. (2019) mitochondrial genomes, seven species in six families and seven genera. Blaimer et al. (2023) with 1100 UCEs addressed the largest sampling of Chalcidoidea for genomic data with 148 species in 23 families and 142 genera, but the results within Chalcidoidea were still poorly resolved, with several families not monophyletic. With the scarce taxonomic sampling of previously published phylogenomic datasets, it has not been possible to test the monophyly of all chalcid families or subfamilies that were questioned by Sanger datasets or morphology (Munro et al., 2011; Heraty et al., 2013). These earlier studies identified a number of *incertae sedis* taxa that can range from a single species to a clade. These lineages are never included within defined (sub)families and defy accurate placement in the resulting trees. Is the lack of definitive placement the result of limited sampling of taxa or genetic data, or could these represent the survivors of independent, possibly old evolutionary lineages, that still need to be resolved. More importantly, reduced taxonomic sampling coupled with poor resolution of phylogenetic trees limits our understanding of the drivers that fuelled this outstanding diversity of life forms and biologies across space and time.

This overview of previous works suggests that the early evolution of Chalcidoidea is likely to represent a difficult phylogenetic problem, and to date, molecular data alone have not helped to resolve many of the problems. A major collaborative effort to provide a resolved morphological tree for the superfamily (233 morphological characters scored on 283 species in 19 families; Heraty et al., 2013) did recover several family-level groups not found in the earlier analysis of ribosomal markers alone, but still failed to recover families or relationships that are considered to be well-supported or provided controversial results.

Within Chalcidoidea, when phylogenomic studies focused on smaller taxonomic units, the results were either in strong agreement with morphology, behaviour or biogeographic hypotheses (e.g. Baker et al., 2020; Rasplus et al., 2020), or strongly conflicting on some areas of the tree with intuitive and previously supported hypotheses (e.g. Cruaud et al., 2021; Zhang et al., 2022). When there is conflict, there is always the possibility that properties of the genomic data or confounding signal in morphological data may be affecting the results. Only with a thorough analysis and vetting of the data can we begin to understand the interplay between systematic bias and morphological convergence.

While genomic data offer great promise to resolve the tree of life, analytical challenges must be overcome. When more markers are analysed, the probability of observing conflicting signal among them increases (Kumar et al., 2012; Philippe et al., 2017; Zhang et al., 2020, 2022). Highly supported trees can be inferred that, without feedback from taxonomists, are considered to be accurate even though inferences may be flawed (Wiens, 2004; Zhang et al., 2022). Heterogeneity in base composition and in evolutionary rates inferred for both taxa and markers are major causes of analytical bias in phylogenomic analyses (e.g. Boussau et al., 2014; Romiguier and Roux, 2017; Borowiec et al., 2019; Rasplus et al., 2021). Recent studies have brought to light an old nemesis, mutational saturation (Philippe and Forterre, 1999), as a potentially major source of errors, especially for deep-time inferences (Borowiec et al., 2015; Borowiec, 2019; Duchêne et al., 2022). Ideally, to detect and reduce inference biases, analysed matrices should be rich both in taxa (Heath et al., 2008) and molecular characters of different origins (e.g. coding vs. noncoding; Reddy et al., 2017). They should also be mixed with or interpreted in the light of morphology, natural history and other data types (Wiens, 2004), which also can be misleading and must be interpreted cautiously. Thorough analysis of very large datasets that implement many proof-checking steps is computationally intensive. In addition, it is still impossible to perfectly describe evolutionary processes with mathematical models, which inevitably introduces bias

(Kumar et al., 2012; Reddy et al., 2017). When independent data types do not converge towards the same results, molecular trees should certainly be acknowledged as valuable contributions, but considered only as hypotheses, instead of being hailed as “the resolved tree of life”.

In this study, we brought together taxonomists and museum curators to assemble a taxon- and marker-rich dataset [exons (protein coding genes) + UCEs and their flanking regions] for Chalcidoidea. To find our way through a forest of phylogenetic trees, we evaluated topologies obtained from each genomic data type in the light of our knowledge of chalcid morphology and natural history. Taxa or groups for which we inferred unlikely relationships were used to detect and reduce potential bias in the molecular datasets. In particular, we evaluated results for five specific benchmark relationships. Departure of results from benchmarks was not necessarily regarded as “wrong” but it was an indication that analytical issues needed to be carefully examined. (i) Basal relationships of Mymaridae and Rotoitidae. This has been proposed based on morphological characters (Gibson and Huber, 2000; Heraty et al., 2013). It should be noted that during the preparation of this manuscript, it was brought to our attention that the name Baeomorphidae is a senior synonym of Rotoitidae. Here, both names are used in text and figures up to the discussion section in which only Baeomorphidae as well as new family/subfamily names resulting from this study are used. (ii) Monophyly of Eurytomidae, including Heimbrinae (Lotfalizadeh et al., 2007; Gates, 2008). (iii) Monophyly of Chalcididae including Cratocentrinae; although somewhat difficult to demonstrate with molecular data alone (Cruaud et al., 2021), this is considered to be a well-supported family based on several morphological synapomorphies (Wijesekara, 1997; Heraty et al., 2013; Cruaud et al., 2021). (iv) Monophyly of Eurytomidae + Chalcididae (Cruaud et al., 2021). (v) Monophyly of the Eupelmidae in the broad sense (including Neanastinae and Metapelmatinae), which although questioned in the past (Gibson, 1989), has been proposed based on a number of derived and correlated morphological features associated with the ability to jump (Gibson, 1986b, 1989, 2008; Heraty et al., 2013). (vi) Monophyly of Agaonidae including Sycophaginae, which is supported by a complex of morphological and life-history traits (Bouček, 1988; Heraty et al., 2013), and their association with figs. In essence, these benchmarks were used in a process of reciprocal illumination to assess the results of our current molecular and morphological hypotheses (Hennig, 1966; Mooi and Gill, 2016).

Finally, we have used the combined exons + UCEs least biased dataset to provide the foundation for a new classification of the superfamily that was recently published (Burks et al., 2022). Here we also discuss the

evolutionary history of Chalcidoidea and infer a timeline for its origin and worldwide colonization.

## Materials and methods

### *Taxonomic sampling*

Representatives of all extant families, 80 extant subfamilies (95.2%), 68 extant tribes (77.3%) and 356 genera (13%) of chalcid wasps were included. Representatives of eight *incertae sedis* taxa at the suprafamilial or tribal levels, and of one new subfamily were also included. Our combined analysis included a total of 433 taxa (414 ingroups, 19 outgroups; Table S1b), of which 414 had sequences for exons [anchor hybrid enrichment (AHE) sequences] while 407 had sequences for UCEs. Exons and UCEs were obtained from the same species in 57% of the ingroup taxa, and congeneric specimens (monophyletic genera) were used in the remaining 43%. For seven taxa (four outgroups, three ingroups), exons and UCEs were obtained from species in different genera (all very closely related for ingroups). Our outgroups include a diverse array of Proctotrupomorpha, including Platygastroidea (two genera), Cynipoidea (five genera), Proctotrupoidea (two genera), Diaprioidea (five genera) and Mymarommatoida (two genera). These outgroups form a grade in which the monophyletic Chalcidoidea was embedded in all recent analyses of Hymenoptera relationships (Heraty et al., 2011; Klopstein et al., 2013; Branstetter et al., 2017; Peters et al., 2017; Blaimer et al., 2023).

The supplementary documents also include results obtained with a larger exon dataset of 520 taxa (494 ingroups, 26 outgroups; hereafter referred to as the AHE520 dataset). AHE414 is the subset of AHE taxa that could be paired with the UCE dataset (Table S1b). As compared to the AHE414 dataset, sampling within some families in the AHE520 data was larger but the same evolutionary lineages were included. We chose not to combine all of the taxa in the AHE520 dataset with the UCE dataset to (i) avoid potential issues with adding considerably more missing data, (ii) enable better comparison between properties and phylogenetic signal brought by the exons and UCEs, and (iii) decrease computational burden. Nevertheless, trees obtained with the AHE520 dataset (exons with RY coding of the third codon position and coded as amino acids) were compared with those obtained with the AHE414, the UCE, and the combined datasets.

### *Library preparation and sequencing*

**DNA extraction.** DNA was extracted nondestructively using the DNeasy Blood and Tissue kit (Qiagen, Hilden, Germany) either at CBGP (Montferrier-sur-Lez, France); NMNH (Washington DC, USA) or UCR (Riverside, CA, USA), as outlined in previous work (Blaimer et al., 2016a, b; Cruaud et al., 2019; Baker et al., 2020; Zhang et al., 2020). DNA was quantified with a Qubit 2.0 Fluorometer (Invitrogen/ Thermo Fisher Scientific, Waltham, MA, USA). Vouchers were subsequently remounted on cards or slide mounted and deposited in either institution or returned to their owner (Table S1b). When possible, DNA extracts were shared among laboratories.

**Exons.** Exons were obtained for 520 taxa following the protocol in Zhang et al. (2022), of which 51 taxa were retrieved from previously published transcriptomes or genomes (Peters et al., 2018; Zhang et al., 2020), and 469 taxa (363 used in combined dataset) were enriched using the anchored hybrid enrichment (AHE) probe sets (*Hym\_Ich* or *Hym\_Cha*; Baker et al., 2020; Zhang et al., 2022). See Supporting information Methods S1 for further details.

**UCEs.** Ultra-conserved elements for 176 taxa were retrieved from previous studies (Cruaud et al., 2019; Rasplus et al., 2020, 2022; Rodriguez et al., 2021; Blaimer et al., 2023) and UCEs for seven taxa were extracted from published genomes (Table S1b). For the remaining 224 taxa, library preparation followed Cruaud et al. (2019). Specimens were enriched in 1432 UCEs using the 2749 probes designed by Faircloth et al. (2015; myBaits UCE Hymenoptera 1.5Kv1 kit; Arbor Biosciences). See Methods S1 for more details.

### *Assembly of datasets*

**Exons.** Assembly of loci, orthology assessment and contamination checking for the AHE520 dataset followed Zhang et al. (2022). Exons were selected for 414 taxa (AHE414) that were compatible for data combination with the UCE taxa, either as the same extraction, same species, same genus, or in three cases as closely related genera (Table S1b). However, the assembly protocol used for the AHE520 dataset was overly stringent, resulting in 28.5% of the nucleotides in the AHE414 dataset being ambiguous or missing. Therefore, raw data for the anchor-enriched taxa selected to be paired with UCEs (Table S1b) were re-processed to increase matrix completeness. Quality filtering and adapter trimming were performed with Trimmomatic-0.36 (Bolger et al., 2014) (LEADING:20 TRAILING:20 SLIDINGWINDOW:4:20 MINLEN:90). Paired reads were then analysed with HybPiper (Johnson et al., 2016) to retrieve exons, introns and supercontigs (introns-exons-introns). All exons of the AHE520 dataset (without contamination) were used as the reference database. One fasta file was then created with all sequences for each exon identified by HybPiper. However, comparison with the 989 reference AHE520 exons revealed that HybPiper exons were shorter than the reference exons. Therefore, exons were instead retrieved from HybPiper supercontigs using the longest sequence of each of the 989 exons of the AHE520 dataset as a reference. Supercontigs for each locus were aligned with the reference longest exon for that locus using LASTZ release 1.02.00 (Harris, 2007). In each of the 989 alignments, nucleotides before the 5' end and after the 3' end of the reference exons were trimmed. The final 989 loci were aligned with MAFFT and translated to amino acids. This latter step revealed that 24 reference exons contained one to two small introns. In these cases, loci were cut into coding versus noncoding sequences, and noncoding sequences were discarded. We ended up with 1007 exons (final % of ambiguous/missing nucleotides = 19.1%).

**UCEs.** Assembly of UCEs followed Cruaud et al. (2019). Quality control checks were performed with FastQC v.0.11.2 (Andrews, 2010). Quality filtering and adapter trimming were performed with Trimmomatic-0.36 (LEADING:20 TRAILING:20 SLIDINGWINDOW:4:20 MINLEN:150). Overlapping reads were merged using FLASH-1.2.11 (Magoc and Salzberg, 2011) and demultiplexed with a bash custom script (Cruaud et al., 2019). Assembly of cleaned reads was performed with CAP3 (Huang and Madan, 1999) or, when the number of paired reads exceeded 300 000, with Trinity (Grabherr et al., 2011). Contigs were aligned to the set of 1432 reference UCEs that resulted from the assembly of the 2749 probes of Faircloth et al. (2015) using LASTZ. Contigs that aligned with more than one reference UCE were removed and different contigs that aligned with the same reference UCE were filtered out using Geneious R11.1.4 4 (<https://www.geneious.com>; sam files were uploaded in Geneious and sorted by the number of contigs that aligned with the reference loci to identify multiple hits). Only UCEs for which sequences were available for >50% of the taxa were kept in the next steps of the analysis ( $N = 1048$ ).

From this point, methods only refer to the AHE414 dataset that was formally compared and then combined with the UCE dataset



and to the UCE dataset. Details for the methods used for treatment of the AHE520 dataset can be found in the Methods S1.

### Quality controls of datasets

**Alignment cleaning.** Loci (exons and UCEs) were aligned with MAFFT using the `-linsi` option (Katoh and Standley, 2013). Exons (AHE) were translated to amino acids using EMBOSS (Rice et al., 2000) and sequences with stop codons were removed. Two successive rounds of TreeShrink (Mai and Mirarab, 2018) were performed on each locus (exons analysed as nucleotides) to detect and remove abnormally long branches in individual gene trees. The per-species mode was used and `b` (the percentage of tree diameter increasing from which a terminal should be removed) was set to 20. Loci were re-aligned with MAFFT after each round of TreeShrink. Gene trees were inferred with IQ-TREE v.2.0.6 (Minh et al., 2020b) with the best-fit model selected by ModelFinder (Kalyaanamoorthy et al., 2017). Positions with >50% gaps and sequences with >25% gaps were removed from the alignments of UCEs using SEQTOOLS (package PASTA; Mirarab et al., 2014) to speed up inference of individual gene trees.

**Contaminations.** A BLAST search of all DNA sequences on themselves (i.e. by using sequences of all exons/UCes for all samples as query and target sequences) was performed (`blastn` with `-evalue 1e-20` and `-max_target_seqs 2`). Only hits for which the same locus was identified as both target and query sequence, and for which samples were different for target and query sequences were kept for downstream analysis. Putative contaminations were identified using a script that scored hits according to four criteria appropriate for our data (see <https://github.com/mjy/cgq> for details): (i) *taxon difference*: either subfamily or family was different between target and query sequences; (ii) *proportional difference*: the percentage of similarity between target and query was >99.95; (iii) *proportional length difference*: length of match divided by the smaller of length of the target and query was >0.95; (iv) *plate similarity*: target and query sequences were obtained from specimens processed on the same plate for DNA extraction/library preparation. When a criterion was met, a score of 1 was attributed, otherwise the score was 0. A composite score was calculated as the sum of criteria 1 to 4. When the composite score equalled 4 both target and query loci were considered as potentially contaminated. The ratio of query species DNA concentration to target species DNA concentration was calculated. When ratio  $\leq 0.3$  the sequence with the smaller qubit

concentration was excluded from the dataset, otherwise both sequences were excluded.

### Properties of taxa, loci, trees and exploration of bias

**Workflow to detect and decrease bias.** We focused on saturation, a property that could artificially distort relationships of an old group such as Chalcidoidea. We also explored whether properties of taxa (GC content and heterogeneity of evolutionary rates between taxa) could explain placements that were unexpected based on our six benchmark criteria (described in the introduction) or life-history traits.

We analysed data subsets that were less and less saturated, and analysed whether resulting topologies had a better fit to our benchmark criteria, contemporary classifications or biological (natural history) data. Reciprocally, morphological/biological data were re-examined to highlight convergences and assess support for unexpected relationships. Thus, to assess fit to morphological data we for example examined whether currently recognized (sub)family-level taxa were monophyletic in resulting trees. To assess fit to biological data, we determined whether clades supported by biological properties were monophyletic in resulting trees.

In order to incrementally decrease saturation, exons were analysed as three datasets: first as nucleotide sequences (exons), then with RY coding of the third codon position of each amino acid (exonsRY), and finally as amino acids (exonsAA; Table 1). RY coding of the third codon position of each amino acid was performed with the script `RYplace.py` (Ballesteros and Hormiga, 2016) and translation to amino acids was performed with EMBOSS.

In order to reduce saturation in the UCE dataset, nucleotide positions in each locus were kept only when they were present in  $\geq 50\%$ ,  $70\%$  or  $90\%$  of the taxa. This resulted in UCE datasets with levels of saturation comparable to that of the concatenated exonAA dataset (Fig. 1). In addition, to avoid bias resulting from misalignment of extremities of short sequences, sequences with >25% gaps (i.e. shorter/incomplete sequences) were removed from each UCE (all trimmings with SEQTOOLS). Three corresponding UCE datasets were thus built, hereafter referred to as UCEs50-25, UCEs70-25 and UCEs90-25 (Table 1).

**Calculation and analysis of properties.** The  $R^2$  of the linear regression of uncorrected p-distances against inferred distances in individual gene trees was used as a proxy for saturation, and

Table 1  
Description of the exons (AHE414) and UCE datasets.

Datasets	Description	Ntaxa	Nloci	Length of concatenated dataset (bp or AA)
exons	Exons as nucleotide sequences	414	1007	310 185
exonsRY	Exons with RY coding of 3rd codon positions	414	1007	310 185
exonsAA	Exons as amino acid sequences	414	1007	103 395
UCEs50-25	UCes with alignment positions kept only when they are present in $\geq 50\%$ of the taxa + sequences with >25% gaps removed.	407	1048	479 872
UCEs70-25	UCes with alignment positions kept only when they are present in $\geq 70\%$ of the taxa + sequences with >25% gaps removed.	407	1048	331 574
UCEs90-25	UCes with alignment positions kept only when they are present in $\geq 90\%$ of the taxa + sequences with >25% gaps removed.	407	1048	180 870
combined	exonsAA + UCEs90-25	433	2054*	103 395 AA +180 711 bp*

Detailed properties are given in Table S2b.

\*Note that the locus shared among the exons and the UCE dataset was removed from the UCE dataset before running the combined analysis.

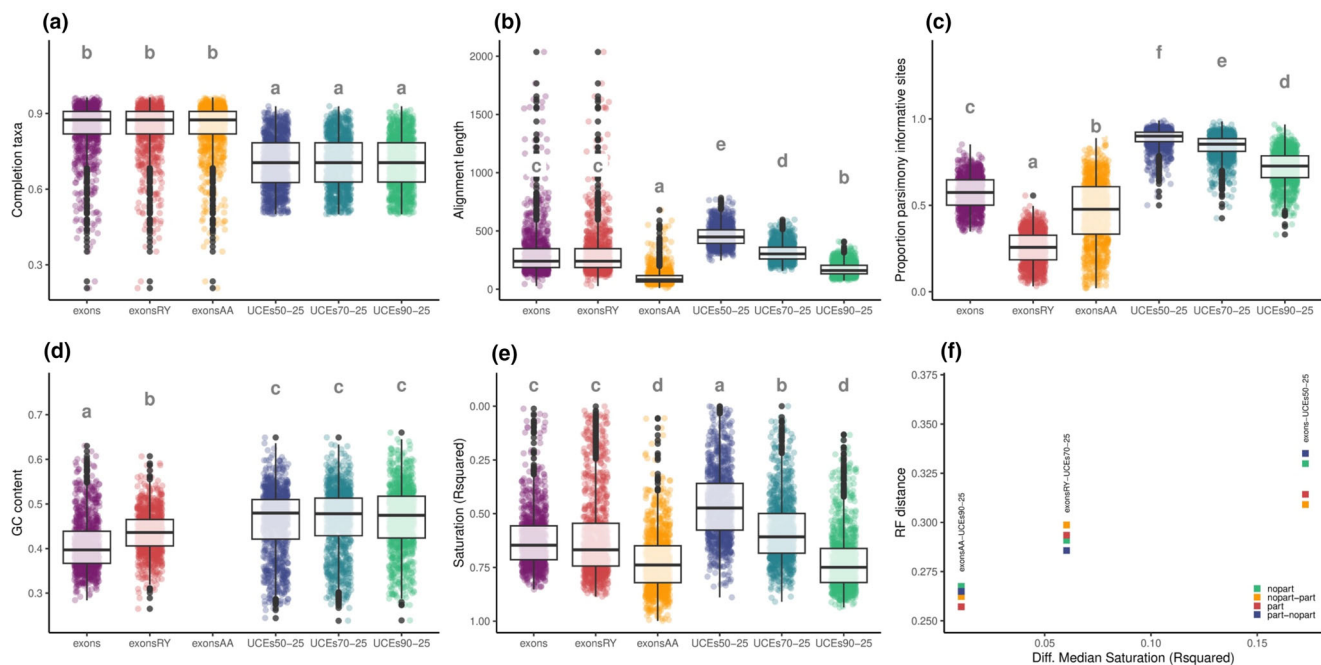


Fig. 1. Comparison of properties of the analysed datasets. Datasets (AHE414 and UCEs) are described in Table 1. For each panel, letters above box plots reflect pairwise comparisons of marginal means estimated from the best-fit models; distributions sharing a letter do not differ significantly. Points: raw data (Table S2a). In (e), saturation was assessed by calculating the  $R^2$  of the linear regression of uncorrected p-distances against inferred distances in individual gene trees. Highest  $R^2$  are for least saturated loci. The scale of the Y axis is reversed to better show decrease in saturation. (f) The convergence of trees as saturation decreases. The Y axis shows the relative RF distance between pairs of trees obtained with either the combined exons or the combined UCEs. The X axis show the absolute value of the difference between the medians of the  $R^2$  of the linear regression of uncorrected p-distances against inferred distances in gene trees that were combined to get the compared trees [cf. (e)]. Four comparisons were performed in each case as datasets were analysed with and without partitioning.

saturation of loci was calculated as in Borowiec et al. (2015). Other properties of loci, properties of concatenated datasets and GC content of taxa (Table S2a) were calculated with AMAS (Borowiec, 2016). Long branch (LB) score heterogeneity for taxa in trees (taxon's percentage deviation from the average pairwise distance between taxa on a given tree) was used as a proxy of evolutionary rate of taxa and was calculated with TreSpEx (Struck, 2014). Statistical analyses of loci, tree and taxa properties were performed in R (R Core Team, 2018). Hierarchical clustering of taxa based on GC content and LB scores was performed with the package *cluster* (Maechler et al., 2018). Strength and direction of association between variables were assessed (i) with Spearman's rank correlation using *PerformanceAnalytics* (Peterson and Carl, 2018) or (ii) by fitting linear models (with log-transformation of variables when relevant (Ives, 2015)). Significant deviations from model assumptions (normality of residuals, homoscedasticity) and absence of highly influential data points were detected with *DHARMA* (Hartig, 2022) and *performance* (Lüdecke et al., 2021). A likelihood ratio test was used to test the significance of fixed factors. A Tukey *post hoc* test was used when more than two groups were compared [packages *emmeans* (Lenth, 2021) and *multcomp* (Hothorn et al., 2008)]. Graphs were generated with *ggplot2* (Wickham, 2016).

### Combined dataset

The exonsAA and UCEs90-25 datasets that were considered as the best hypotheses (see "Results" section) were combined (Table 1). Before combination, overlap between exons and UCEs was tested with reciprocal BLAST (UCEs not trimmed with SEQTOOLS; blastn with -evalue 1e-20). Only a single locus was shared between

datasets and it was removed from the UCE dataset before combination.

### Phylogenetic inference

Datasets (Table 1) were analysed with maximum likelihood (using IQ-TREE 2.0.6, Minh et al., 2020b) and parsimony (using TNT; Goloboff et al., 2008) approaches.

For IQ-TREE analyses, loci were merged and the resulting dataset was analysed (i) without partitioning, (ii) with one partition for each locus and (iii) with one partition for each data type (for the combined dataset only; one partition for the exons another for the UCEs). Best-fit models for each partition were selected with the Bayesian Information Criterion as implemented in ModelFinder. FreeRate models with up to ten categories of rates were included in tests for the unpartitioned exon and UCE datasets, but only common substitution models were tested when datasets were partitioned by locus. The candidate tree set for all tree searches was composed of 98 parsimony trees +1 BIONJ tree and only the 20 best initial trees were retained for NNI search. Statistical support of nodes was assessed with ultrafast bootstrap (UFBoot, Minh et al., 2013) with a minimum correlation coefficient set to 0.99 and 1000 replicates of SH-aLRT tests (Guindon et al., 2010). Gene (gCF) and site (sCF) concordance factors (Minh et al., 2020a) also were calculated.

In order to account for possible heterotachy, the combined exonsAA + UCEs90-25 dataset was also analysed with a GHOST model (Crotty et al., 2020) in IQ-TREE with four mixture classes in conjunction with the best-fit model chosen by ModelFinder that was fitted to each of the two partitions (i.e. one partition for each type of marker). Exploration of tree space and inference of statistical

support for nodes followed those used with mixed models. Analyses with the CAT-GTR model to deal with substitutional heterogeneity were attempted with PhyloBayes-MPI v.1.9 (Lartillot et al., 2013) but were computationally not tractable as expected from previous reports (PhyloBayes-MPI 1.9 manual: <https://github.com/bayesiancook/pbmpi> and Whelan and Halanach (2017); memory issues with all tests using  $\leq 96$  CPUs and 192Go for a single chain).

For parsimony analyses with TNT, we used New Technology Search with default Sectorial Search, Ratchet, Drifting and Tree Fusing search criteria, and finding the minimum length ten times. We only analysed the final datasets: combined exonsAA+UCEs90-25 [433 taxa], UCEs90-25 [407 taxa], exonsAA [414 taxa] using the appropriate amino acid (prot) or DNA partition. Bootstrap support was assessed using absolute frequencies and 1000 replicates.

Finally, distance between a node and its parent node on trees was calculated with the R package *ape* (Paradis and Schliep, 2018). RF distances (Robinson and Foulds, 1981) among recovered trees were calculated with RAXML-NG\_v.0.9.0 (Kozlov et al., 2019).

### Divergence time estimates

Time-calibrated trees were generated with MCMCtree (Yang and Rannala, 2006). Twenty-one fossils (Table S3a) were used as calibration priors (see for rationale for selecting calibrations). Using log-normal or even normal priors may excessively restrict the breadth of posteriors (e.g. Brown and Smith, 2018). Following previous studies (e.g. Cruaud et al., 2012a; Tong et al., 2015; Blaimer et al., 2023) and given the large uncertainty for the Chalcidoidea fossil record, uniform distributions were used as calibration densities. Analyses were run with uncorrelated relaxed clock models. The combined IQ-TREE tree (partitioning by data type) was used as the input tree. Five datasets, each composed of 10 000 amino acid sites randomly selected (custom script in [https://github.com/acruaud/saturniidae\\_phylogenomics\\_2022](https://github.com/acruaud/saturniidae_phylogenomics_2022); Rougerie et al., 2022) from the exonsAA partition +10 000 nucleotide sites randomly selected from the UCEs90-25 partition, were used as sequence data to make computation tractable. Each dataset was partitioned into two partitions, exonsAA (WAG + G model) and UCEs (GTR + G model). Four chains were run for each dataset; 20 000 generations were discarded as burnin and chains were run for 2 million generations with sampling every 100 generations. Convergence was assessed in Tracer (Rambaut et al., 2018). Possible conflicts between priors and data were assessed by running MCMCtree without sequence data. Posterior estimates obtained with the different datasets were compared and combined with LogCombiner 2.6.0 (Bouckaert et al., 2019).

### Historical biogeography

Distributions of species in each genus were mined from the Universal Chalcidoidea Database (Noyes, 2019). Occurrences were double-checked by the authors who are experts in each group. Species used as biocontrol agents or accidentally introduced with their host (plant or insect; Rasplus et al., 2010; Noyes, 2019) were removed. Genera were scored as present/absent in the six following biogeographical areas: Neotropical, Nearctic, Afrotropical, Palaearctic, Oriental and Australasian (Table S4a). The chronogram built with MCMCtree was used as input, but only one specimen per genus was included and outgroups were pruned to avoid artefacts. We acknowledge that a species-level analysis of ancestral areas would have been more appropriate than at the genus-level, because by using the genus-level the results could be biased towards more widespread ancestors (BioGeoBEARS manual). However, in old, hyperdiverse and widely distributed groups, covering the entire taxonomic and geographical diversity is impossible. Instead of keeping single or a couple of species that were randomly sampled and could bias

results towards too restricted areas we chose to use generic distribution as input. Ancestral area estimations were performed in the R package *BioGeoBEARS 1.1.1* (Matzke, 2014). Dispersal-Extinction-Cladogenesis (DEC; Ree and Smith, 2008), BAYAREALIKE and DIVALIKE (Ronquist, 1997) models were used with and without considering the jump parameter for founder events (+J; Matzke, 2014). Model selection was performed based on statistical (AICc; Matzke, 2022) and nonstatistical (i.e. biological and geographical) considerations (Ree and Sanmartin, 2018). The maximum number of areas that a taxon could occupy was set to 6. To account for the main geological events that occurred during the diversification of Chalcidoidea, we defined five time periods with different dispersal rate scalars: (i) from their mean crown age to 145 Ma (Jurassic); (ii) 145 to 100 Ma (Early Cretaceous); (iii) 100 to 66 Ma (Late Cretaceous); (iv) 66 to 23 Ma (Palaeogene); and (v) 23 Ma to present (Table S4b).

## Results

### Exploration of biases

The initial set of exons (AHE414 exons as nucleotides; Table 1) had a high number of loci recovered across taxa and was less saturated and less GC rich than the initial set of UCEs (UCEs50-25) that, in comparison, contained longer loci and more parsimony informative sites (Fig. 1; Table S2a). All of our phylogenetic analyses from either exons or UCEs returned somewhat similar results (Figs 2, 3 and S1; Appendix S1). Chalcidoidea was always monophyletic, as were many of the currently recognized families and subfamilies.

However, in the most saturated datasets (exons as nucleotides and UCEs50-25), there was overall lack of support for many of our benchmark relationships, with discordance in (i) the placement of Myrmaridae or Rotoitidae (=Baeomorphidae) as sister to the remaining Chalcidoidea; (ii–iv) Eurytomidae and Chalcididae were neither monophyletic or grouped together; (v) Eupelmidae were not monophyletic; and (vi) Sycophaginae did not group with Agaonidae (Fig. 2). Clustering analyses of taxa properties (GC content and LB scores) did not provide evidence that these initial unexpected placements were driven by compositional bias or long branch attraction (Fig. S2; Table S5). Indeed, monophyletic groups with different GC content as well as polyphyletic groups with similar LB scores were recovered in the different trees.

As we reduced saturation (Fig. 2), we began to see some of our benchmarks stabilize in the results, such that (i) Myrmaridae were sister to the remaining Chalcidoidea, (ii) Eurytomidae were monophyletic; (iii–iv) Chalcididae + Eurytomidae were monophyletic, but Chalcididae were paraphyletic with Cratocentrinae sister to the remaining Chalcididae + Eurytomidae. Other topological changes that may be attributed to reduction of mutational saturation are listed in Table S6.



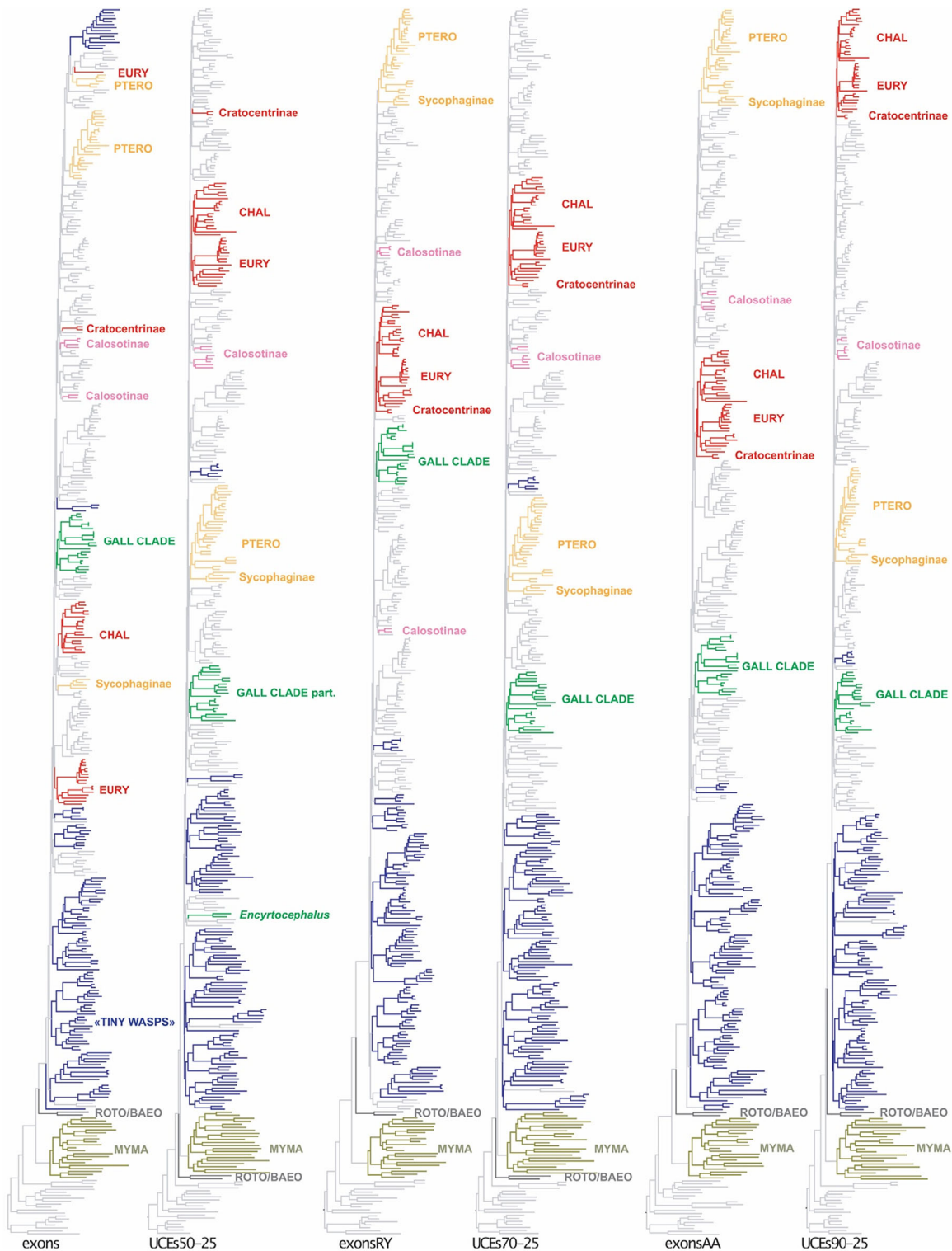


Fig. 2. Overview of the topologies obtained with the different datasets as saturation decreases. Datasets (AHE414 and UCes) are described in Table 1, and trees are available in Fig. S1 and Appendix S1. Groups that are discussed in text are highlighted. Only IQ-TREE trees are shown. ROTO/BAEO = Rotoitidae (Baeomorphidae); CHAL = Chalcididae; EURY = Eurytomidae; GALL = gall clade (see text); MYMA = Mymaridae; PTERO = group of Pteromalid wasps (Austroterobiinae; part Colotrechninae; Miscogastrinae; part Ormocerinae; Oti-tesellinae; Pteromalinae; Sycoecinae; Sycoryctinae); “Tiny Wasp clade” (see text).



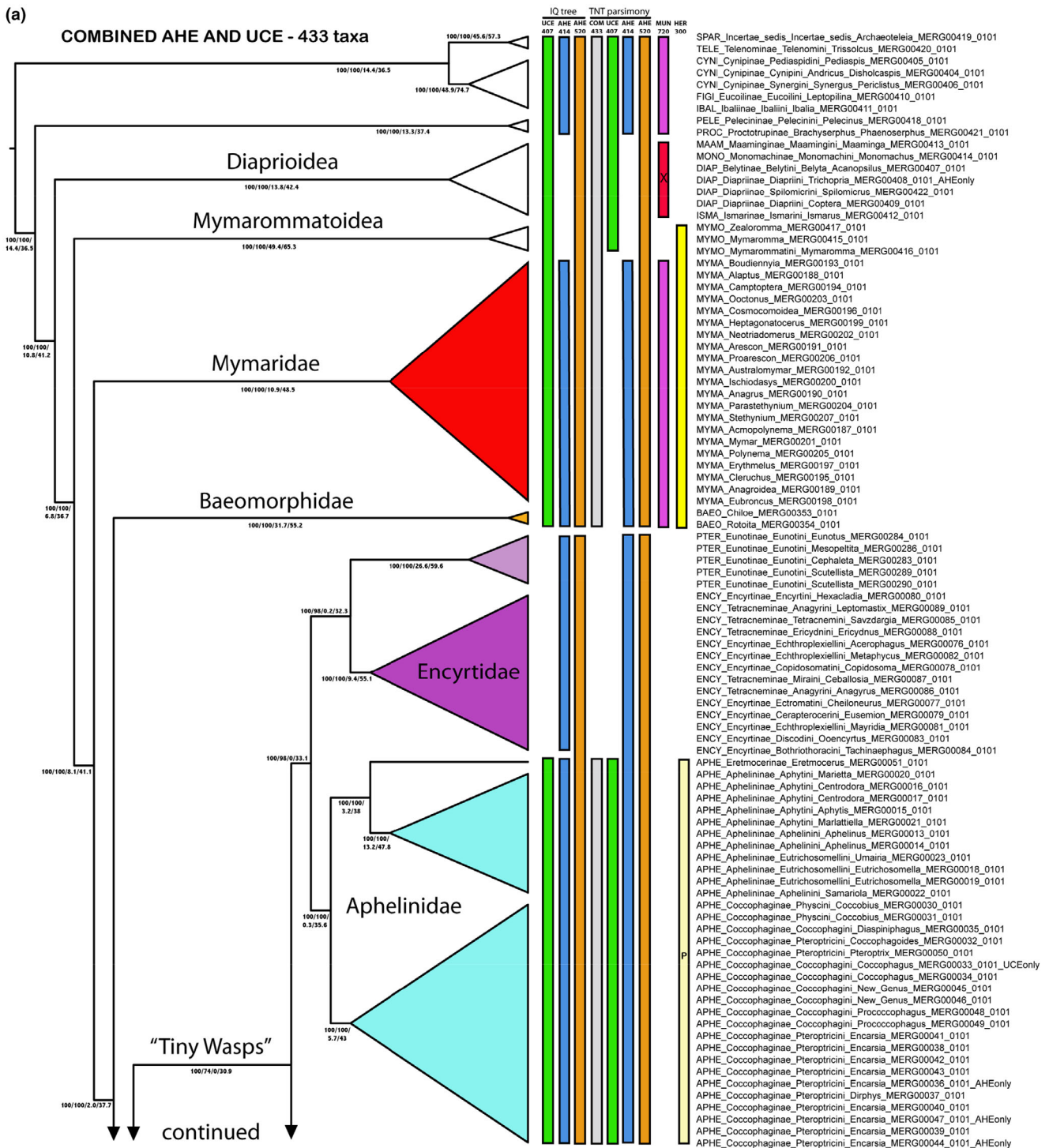


Fig. 3. Collapsed summary cladograms. Combined AHE (exonsAA) and UCE (UCEs90-25) results for 433 taxa, 2054 loci including 103 395 AA +180 711 nucleotide sites. Results of IQ-TREE concatenated analysis with one partition for each type of data (exonsAA vs. UCEs90-25). SHaLRT/UFBoot/gCF/sCF are indicated at nodes. Clades were collapsed to higher level groups (family, subfamily, tribe). Colours are meant to allow for comparisons between trees. Vertical bars represent similar clade relationships for the analyses of IQ-TREE: UCEs90-25 (UCE407), exonsAA (AHE414), AHE520AA; parsimony: combined (COM433), UCEs90-25 (UCE407), exonsAA (AHE414), AHE520AA; Munro et al. (2011; 720 taxa); Heraty et al. (2013; 300 taxa). Vertical red bars with an X were not recovered as monophyletic in that analysis. Faded colour bars represent that the clade was included but relationships alternated. P indicates paraphyletic lineages. Clades without an X or bar were supported; the lack of a bar indicates the clade was supported but the deeper relationships were not. Higher group names refer to the classification before Burks et al. (2022). Family abbreviations expanded in Table S1.

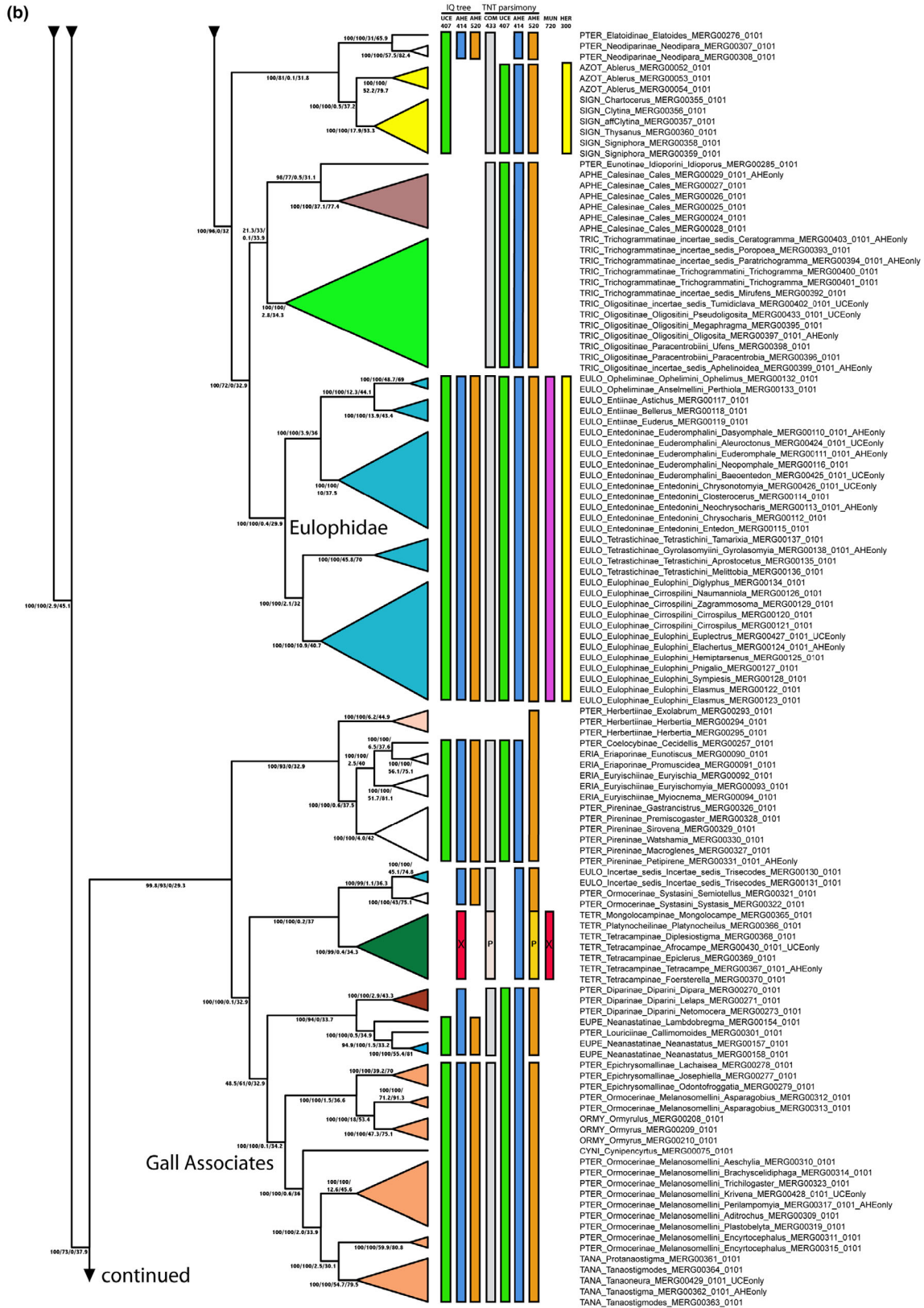


Fig. 3. Continued.

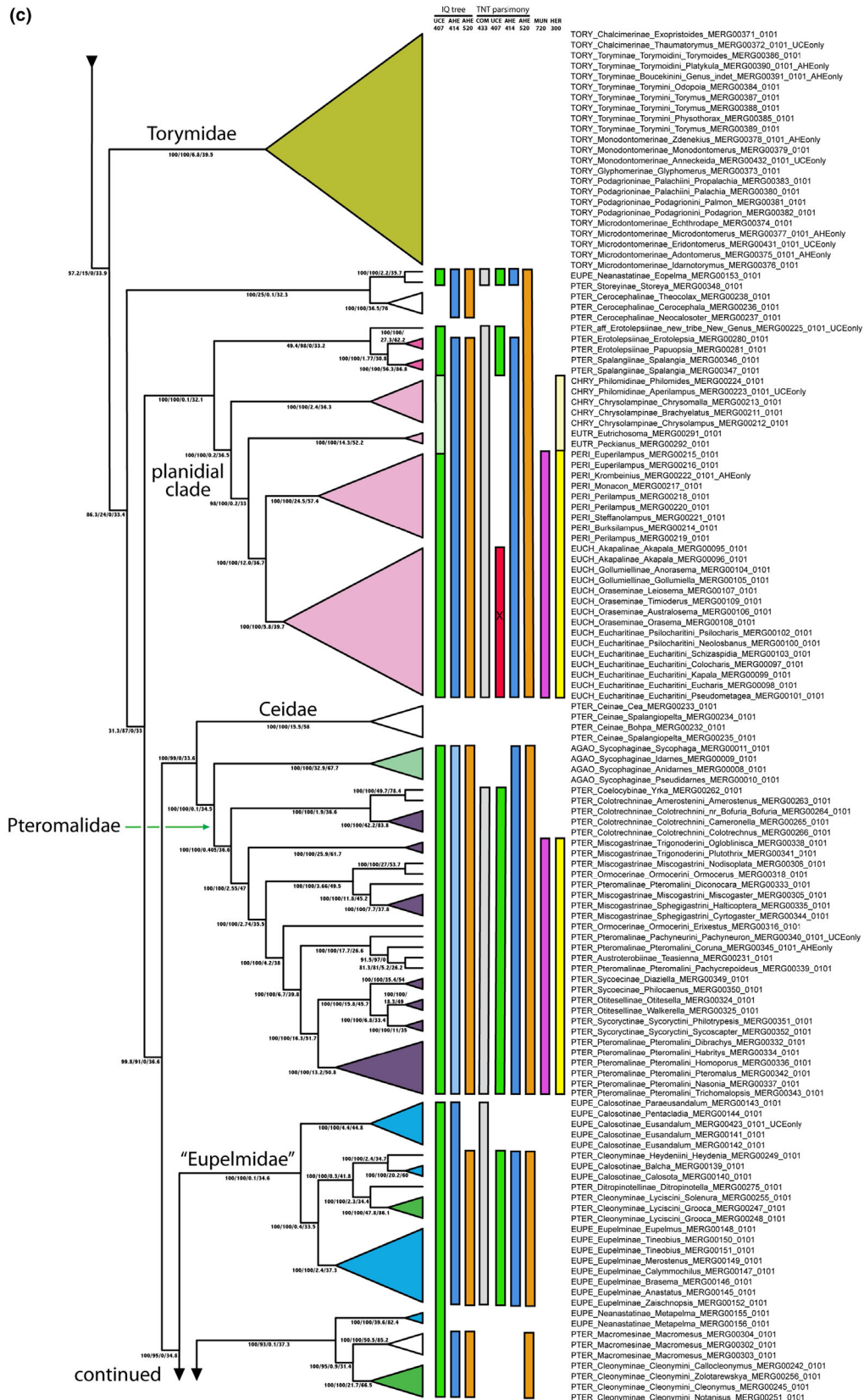


Fig. 3. Continued.



(d)

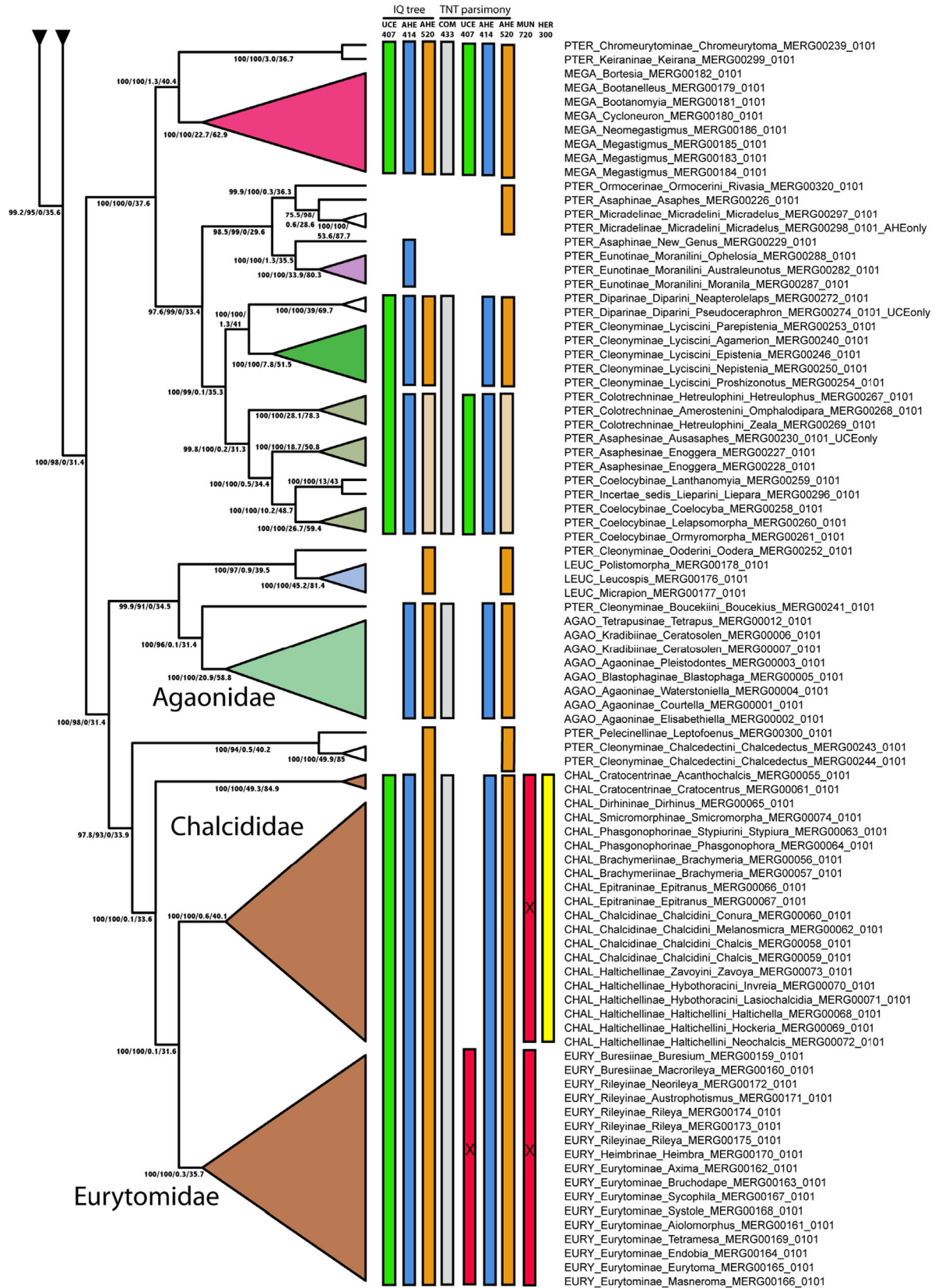


Fig. 3. Continued.



Visual comparison of trees and RF distances (Figs 1, 2 and S1; Table S2c) showed that trees obtained from the likelihood analyses of the exons and UCEs were the most similar when saturation was the lowest (i.e. with exonsAA and UCEs90-25). They also were more concordant with most, but not all, of our benchmark criteria. The exonsAA and UCEs90-25 datasets were thus combined to produce a phylogenetic tree on which phylogenetic relationships were discussed and dating analyses were performed.

### Phylogenetic relationships

Differences were observed when the combined dataset was partitioned by data type (AA vs. UCEs) or by locus in IQ-TREE, but they were all observed in poorly supported sections of the topology (Fig. S1). The GHOST model did not improve the topology and one higher-level grouping disappeared (the “Tiny Wasp”; Fig. S1). Given that short loci generated numerical instability for the estimation of model parameters as reported by IQ-TREE (partitioning by locus), we chose to rely on the IQ-TREE mixed model combined tree (Figs 4a and 5) for discussion of phylogenetic relationships.

More UCEs than exons supported the IQ-TREE combined tree (gCF; Fig. 4b), but a significant component of discordance was due to the lack of resolution of gene trees (gDFP). Distributions of sCF for nodes were identical for the two types of markers (Fig. 4b). Statistical support was higher for longer branches (Fig. S3A,B). Absolute RF distances between the exonsAA and the combined trees or the UCEs90-25 and the combined trees were close (ten more branches shared between the UCEs90-25 and the combined trees; Table S2c). The combined tree was thus considered as an acceptable synthesis of the two types of markers.

Nearly identical results were recovered from our combined IQ-TREE and TNT analyses (Figs 2, 3 and S1, RF between IQ-TREE mixed model and TNT trees = 98). Of the 25 extant families (Table S1a), 18 were recovered as monophyletic with strong support, while seven were recovered as paraphyletic or polyphyletic (Figs 3, 4a and S1). The worst case was Pteromalidae, which were spread across the entire tree (PTER; Fig. 5). Aphelinidae and Eulophidae were both monophyletic, but only if one genus was excluded from each family, *Cales* and *Trisecodes*, respectively. Finally, Agaonidae (two lineages), Chalcididae (two lineages) and Eupelmidae (five lineages) were polyphyletic.

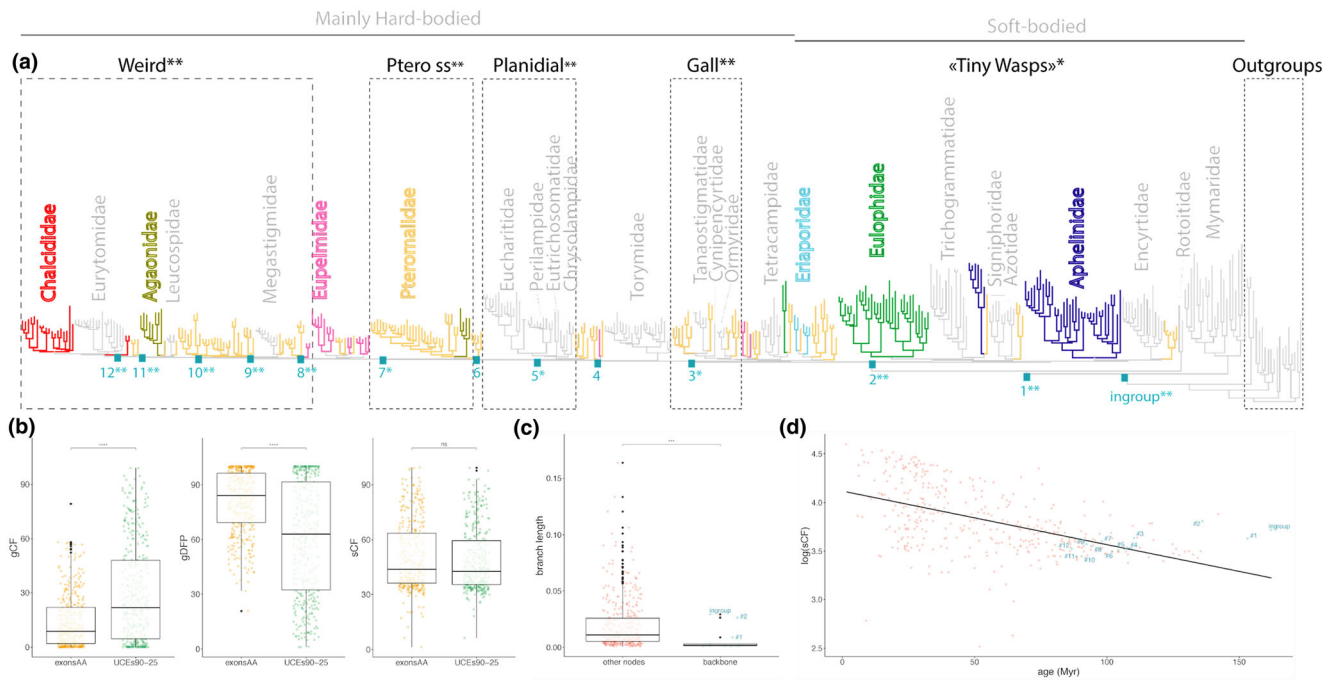


Fig. 4. The Chalcidoidea bush of life. (a) IQ-TREE tree obtained from the combined exonsAA+UCEs90-25 datasets (see also Fig. S1). Monophyletic families are in grey, para- or polyphyletic families are in colour. Higher level groups/clades discussed in text are highlighted with boxes. Statistical support for backbone nodes are shown with single (SH-aLRT  $\geq 80\%$  or UFboot  $\geq 95\%$ ) or double stars (SH-aLRT  $\geq 80\%$  and UFboot  $\geq 95\%$ ). (b) Contribution of the exonsAA and UCEs90-25 datasets to the combined tree. Gene concordance factor (gCF); gene discordance factor due to polyphyly (gDFP); site concordance factor averaged over 100 quartets (sCF). Points: raw data (Table S2d). (c) Comparison of branch length for the backbone nodes and other ingroup nodes. Points: raw data (Table S2c). For (b) and (c), stars above box plots indicate statistical significance: ns,  $p > 0.05$ ; \*\*\*,  $p \leq 0.001$ ; \*\*\*\*,  $p \leq 0.0001$ . (d) Correlation between node age and sCF (outgroups excluded). Points: raw data (Table S2e); line: regression curve for the best-fit model (log linear model;  $p < 2.2e-16$ ).

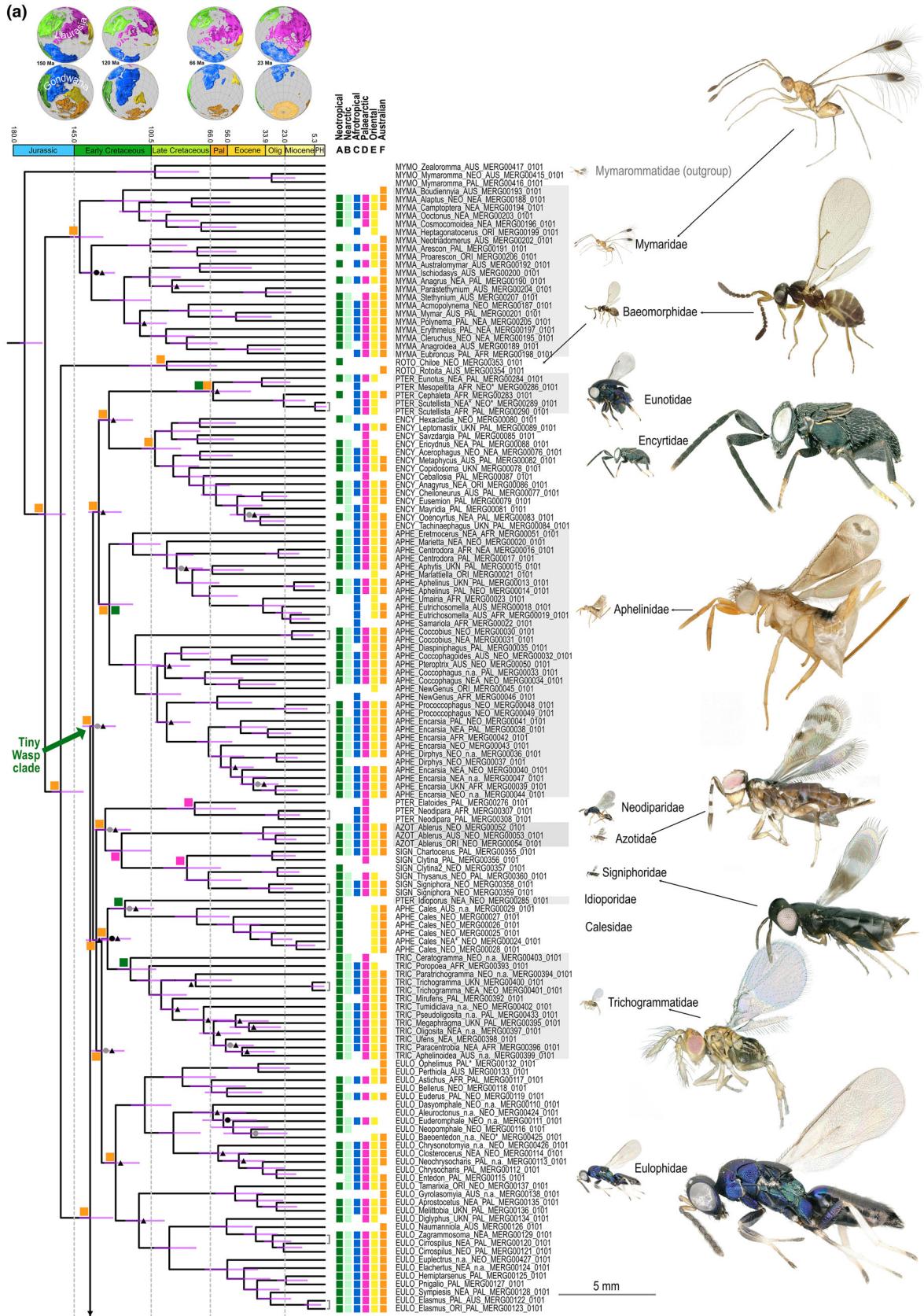


Fig. 5. Global historical biogeography of Chalcidoidea and new classification. The chronogram obtained from the complete set of ingroup taxa is illustrated. The previous classification is used to annotate tips (four letter prefixes; see also Table S1 for complete information on sampling) with successive grey and white boxes grouping the tip labels. The new familial classification from Burks et al. (2022) is shown to the right. For clarity, ancestral ranges are given only up to family level and only for the BAYEAREALIKE + *J* model (which was selected by AICc). All inferences of ancestral ranges are provided in Fig. S5. Inferences of ancestral ranges were conducted with only one specimen per genus as shown with brackets that connect tips. Current distribution of genera is shown with coloured boxes at tips. Sampling area of specimens is indicated in tip labels. NEO = Neotropical; NEA = Nearctic; AFR = Afrotropical; PAL = Palaearctic; ORI = Oriental; AUS = Australasian. UKN = Unknown when collection data are unavailable. Stars indicate that specimens were sampled in areas where species was introduced or not yet cited. Sampling area for the specimen used for sequencing exons is listed first, sampling area for the specimen used for sequencing UCEs is listed second; n.a. is used when no specimen was sequenced and only one sampling area is reported when exons and UCEs were obtained from specimens sampled in the same areas (or from the same specimen). Unless specified, nodes are supported by SHaLRT  $\geq 80\%$ , UFBoot  $\geq 95\%$  and sCF  $\geq 34.3$  (minimum support for a family that is well defined morphologically, Trichogrammatidae). Nodes with a grey circle are supported by SHaLRT  $< 80\%$  or UFBoot  $< 95\%$ ; nodes with a black circle are supported by SHaLRT  $< 80\%$  and UFBoot  $< 95\%$ ; nodes with a black triangle are supported with sCF  $< 34.3$ . Images on the left of tentative family names are all at the same scale. Images on the right of tentative family names have been magnified. Photos ©K. Bolte (Baeomorphiidae); ©J.-Y. Rasplus (all others).

There were some higher-level relationships inferred that may reflect biology more so than morphology (Figs 4a and 5). A clade of gall-associates emerged (hereafter referred to as the “Gall Clade”), which had not been proposed previously based on morphology. A clade of chalcids with planidial larvae as already found by Zhang et al. (2022) also was recovered. A group of “Tiny Wasps” (usually  $< 4$  mm in length and soft-bodied) mostly associated with Hemiptera was inferred as a monophyletic group with IQ-TREE. However, most analyses supported this group as a grade of small soft-bodied chalcids leading to a clade of larger more robustly sclerotized wasps that include Chalcididae and Eurytomidae. This latter higher level grouping (hereafter referred to as the “Weird clade”) was unexpected based on either morphology or natural history.

All but two nodes (#1 and #2; Fig. 4a) in the backbone (i.e. the first 12 bifurcations) of the combined IQ-TREE were closer to each other than most ingroup nodes, suggesting near simultaneous old divergences for most nodes (Fig. 4c). Only these two backbone nodes were recovered in the set of gene trees. Sixty percent of the backbone nodes have SH-aLRT and UFBoot higher than the suggested cut-off for validity ( $\geq 80\%$  and  $\geq 95\%$ , respectively; IQ-TREE manual; Figs 4a, 5 and S1). A generic cut-off for sCF across a tree that spans such a long time makes little sense. Hence, we used the lowest sCF obtained for families that are well-defined morphologically and biologically (Heraty et al., 2013): 34.3 for Trichogrammatidae. Using this value as a cut-off, 50% of the backbone nodes are supported (Figs 4a, 5 and S1). We observed a significant overall decrease in all statistical support for nodes (gCF, sCF, UFBoot, SH-aLRT) with increasing age (Figs 4d and S3C).

#### *Divergence time estimates and historical biogeography*

The chronogram obtained from the combined tree is shown in Figs 5 and S4. Divergence time estimates

and confidence intervals for all nodes are given in Table S3b. Estimates of divergence time indicate that Chalcidoidea diverged from their sister group (Mymarommatoidea) 174.0 Ma [95% Equal-tail Confidence Interval (95% CI) 167.3–180.5 Ma]. Crown Chalcidoidea is dated at 162.2 Ma (153.9–169.8 Ma). The four first splits on the backbone [Mymaridae; Baeomorphiidae (= Rotoitidae); “Tiny Wasp clade”; all other chalcid wasps] occur over a timespan of  $\sim 53$  Myr. From  $\sim 110$  Ma, divergences are closer in time with the remaining eight splits on the backbone spanning  $\sim 24$  Myr. Ancestral range estimations for all biogeographical models are provided in Fig. S5. AICc favoured the BAYAREALIKE+*J* model (Table S4c; Fig. 5). A South Gondwanan origin of Chalcidoidea [Australasian (DEC, DEC + *J*, BAYAREALIKE + *J*; DIVALIKE + *J*) or Australasian + Neotropical (BAYAREALIKE; DIVALIKE)] is suggested by all models, with colonization of the rest of the world more or less delayed in time depending on the model used (Fig. S5).

## Discussion

### *Mutational saturation disturbs phylogenomic inferences*

This study expands upon earlier investigations (Munro et al., 2011; Heraty et al., 2013; Peters et al., 2018; Zhang et al., 2020; Blaimer et al., 2023) to yield a more comprehensive phylogenetic framework for higher relationships within Chalcidoidea. We used a comparison of phylogenomic inferences from molecular datasets (exons for 414 taxa and UCEs for 407 taxa) and morphological/biological/ecological knowledge to enable us to detect systematic bias attributable to mutational saturation, and, conversely, morphological convergence. With decreasing saturation, exons and UCEs topologies tended to become more similar (Figs 1, 2 and S1; Appendix S1) and relationships of groups for which placement could be evaluated based



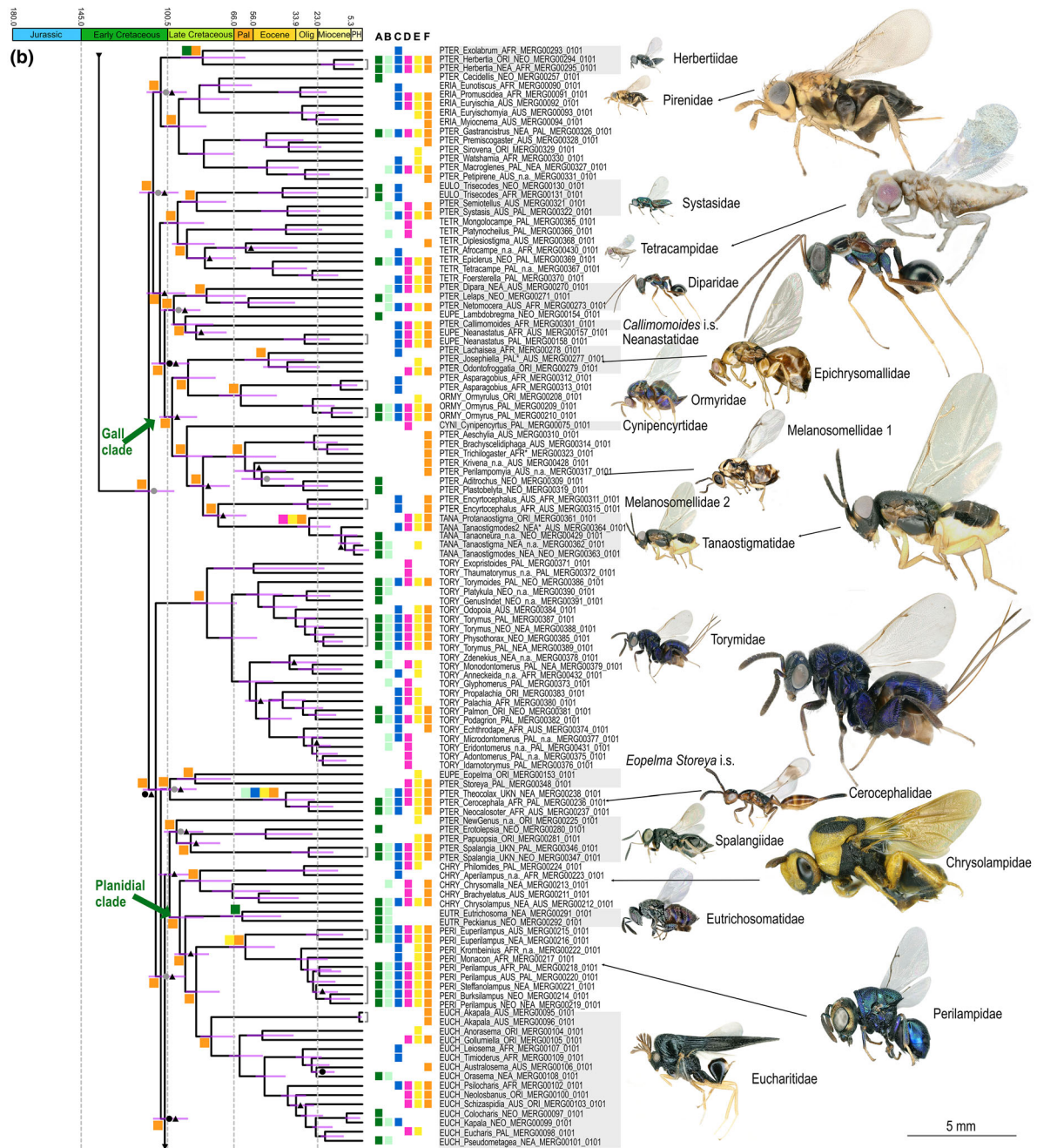


Fig. 5. Continued.

on morphological data tended to become more concordant with morphology or biology (Table S6), though not always. Thus, we confirm that mutational saturation is an important source of error in phylogenomics, especially for deep-time inferences.

We comment below on the results with respect to selected groups of taxa, including our benchmarks, and what they reveal about inference bias due to saturation.

*Mymaridae* and *Baeomorphidae* (= *Rotoitidae*). *Mymaridae* was recovered as sister to all other Chalcidoidea with strong support (Benchmark 1) in all but the most saturated UCE datasets. With this last dataset, *Baeomorphidae* was inferred as sister to the rest of chalcid wasps including *Mymaridae*. With decreasing saturation, the SHaLRT and UFBoot support for *Mymaridae* as sister to other Chalcidoidea increased in the UCE trees. *Mymaridae* and



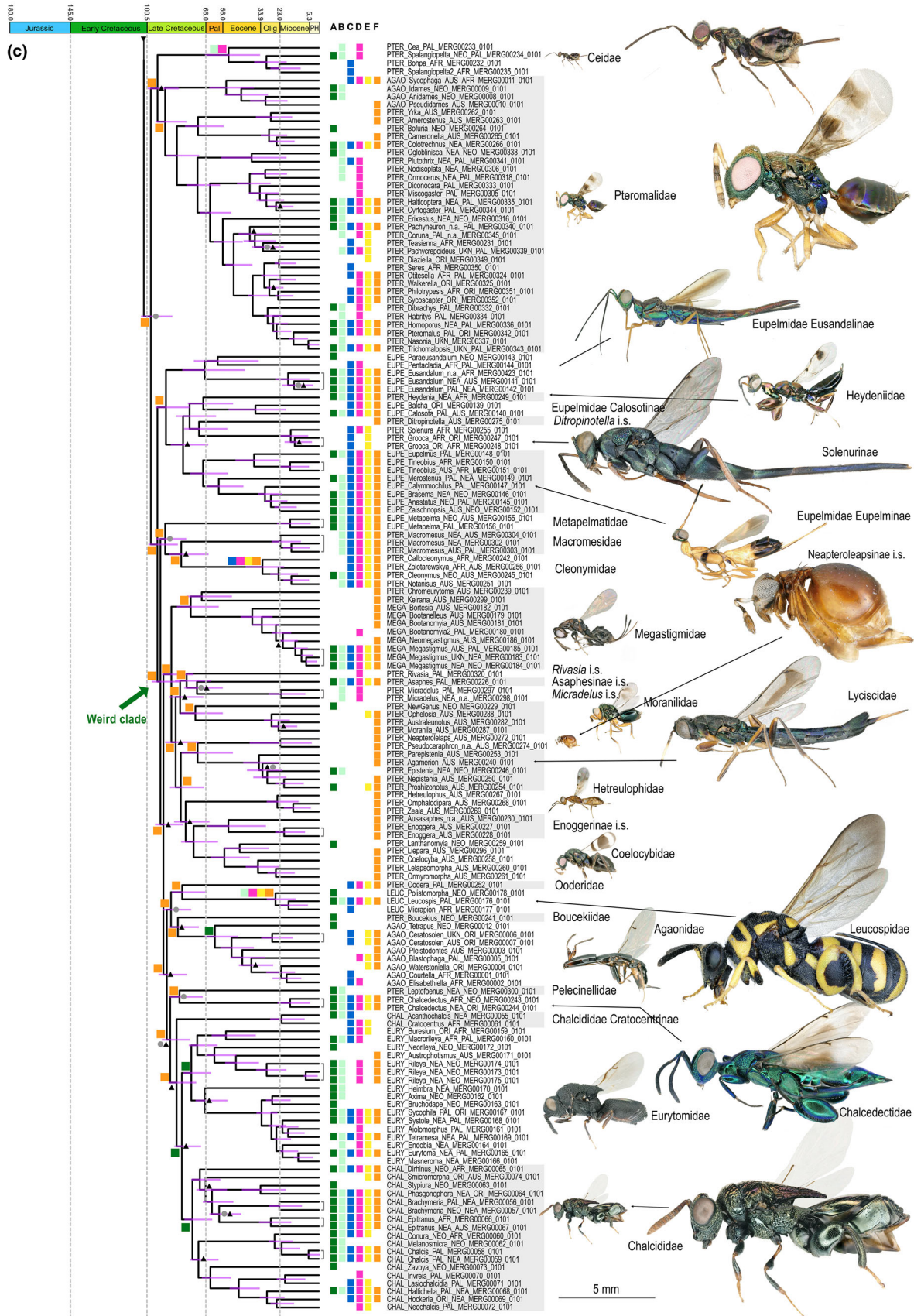


Fig. 5. Continued.

Baeomorphidae have long been hypothesized as the first and second lineages of Chalcidoidea to diverge from their common ancestor (Gibson and Huber, 2000; Munro et al., 2011; Heraty et al., 2013). Blaimer et al. (2023) recovered the same relationships with strong support, but here we confirm this with a greater sampling of taxa.

*Chalcididae and Eurytomidae.* The polyphyly of Eurytomidae (Benchmark 2) as well as the nonmonophyly of Chalcididae + Eurytomidae (Benchmark 4) in the most saturated exon and UCE datasets can be attributed to mutational saturation. Indeed, Eurytomidae is a well-supported family (Heraty et al., 2013). Cratocentrinae is considered to be the sister group of other Chalcididae (Cruaud et al., 2021). However, a monophyletic Chalcididae (Benchmark 3) was only recovered in the parsimony analyses of the AHE 414 and 520 datasets (Figs 3 and S1). A mesothoracic spiracle that is hidden in Eurytomidae and all Chalcididae except for Cratocentrinae gives support to Cratocentrinae being sister to the two other taxa, but 15 synapomorphies support Chalcididae as monophyletic (Cruaud et al., 2021). This suggests that despite high statistical support (SHaLRT = 94/UFBoot = 97 for exonsAA; 100/100 for UCEs90-25 and combined), IQ-TREE inferences may be incorrect but this result requires more investigation. Benchmark 2 (monophyly of Eurytomidae) was only supported by the prot-AA dataset of Blaimer et al. (2023); in other cases, Heimbrinae was placed outside of this clade, similar to our most saturated analyses (Fig. 2). With regard to our Benchmark 3 (monophyly of Chalcididae), they did not include Cratocentrinae, so this latter hypothesis was not tested by them.

*Calosotinae (Eupelmidae).* Although morphology weakly supports Calosotinae as monophyletic (Gibson, 1989; Heraty et al., 2013), it was always recovered as polyphyletic in our analyses. Indeed, a group of Calosotinae (*Eusandalum*, *Pentacladia* and *Paraeusandalum*), which exhibit V-shaped notauli, never clustered with other Calosotinae that show paramedially parallel notauli (Gibson, 1989). This group is instead more closely related to several Pteromalidae genera (*Heydenia*, *Ditropinotella*, *Grooca* and *Solemura*), a result that is somewhat corroborated by morphology. With decreasing saturation, species belonging to Calosotinae were less scattered across the trees, and Calosotinae with V-shaped notauli became sister to the clade composed of other Calosotinae, some Pteromalidae genera and Eupelminae, with the result that a core group of Eupelmidae (Calosotinae and Eupelminae) was not monophyletic. This result was likewise supported by all of our preferred phylogenomic datasets (Fig. 3). Convergent modifications of mesosomal structure in “jumping” taxa, including an enlarged acropleuron (actually the external manifestation of profound changes in internal

skeletomusculature; (Gibson, 1989) and associated pegs or spines on the mesotarsal segments linked to the ability to jump could have misled morphological studies (Peters et al., 2018; Zhang et al., 2020). Notably, none of the studies based on molecular data alone supported monophyly of a clade with jumping abilities that includes Eupelmidae, Cynipencyrtidae, Encyrtidae, Tanaostigmatidae and some Aphelinidae. Monophyly of Calosotinae, Eupelmidae and a clade that included Cynipencyrtidae, Tanaostigmatidae and Encyrtidae was only found in the combined morphological and molecular analysis of Heraty et al. (2013). Modifications linked to the ability to jump may be at the origin of one of the two characteristics considered as apomorphies defining the clade (mesoscutal lateral lobes “shoulder-like” on either side of pronotum). The presence of parapsidal lines in Calosotinae and in *Solemura* and *Grooca* (Gibson et al., 1999) is a potential argument to redefine Calosotinae, with not all taxa having an enlarged acropleuron. Despite all of our efforts, our fifth benchmark criterion, monophyly of Eupelmidae in the broad sense (including Neanastinae and Metapelmatinae) or even monophyly of Eupelminae and Calosotinae (the narrow sense), was never achieved and so was also the case in Blaimer et al. (2023).

*Agaonidae (Benchmark 6).* Sycophaginae and all of the other subfamilies of Agaonidae are associated with *Ficus* (Moraceae). They have been considered as part of the same family based on morphological data (Heraty et al., 2013), but this result was not supported by molecular data (Munro et al., 2011; Blaimer et al., 2023). In our results, Sycophaginae consistently clustered away from other Agaonidae. With decreasing saturation in the exon dataset and with UCEs, Sycophaginae were consistently recovered in the same group of pteromalids [now Pteromalidae sensu stricto (s.s.); Burks et al., 2022]. While several morphological characters group Sycophaginae and other Agaonidae, several others are shared with lineages of Pteromalidae with which Sycophaginae clustered on our trees. Characters supporting a close relationship between Pteromalidae and Sycophaginae include the separated postgenae with an interceding lower tentorial bridge impressed relative to the postgena, the structure of the antenna, being 14-segmented with a terminal button, and presence of an axillular sulcus in all but a few highly derived species (Heraty et al., 2013). In addition, the gall-associated biology of Sycophaginae is similar to that of Colotrechninae (Pteromalidae), another lineage with which Sycophaginae cluster. This case is another example of a result not concordant with a benchmark criterion that is therefore rejected.

*Encyrtacephalus (Melanosomellidae).* As saturation decreased, the genus *Encyrtacephalus* grouped within the Gall clade (Fig. 2). Morphologically,

*Encyrtcephalus* shares characters with Melanosomellini and only differs from a majority of them by the large supracoxal flange of propodeum and a curved stigmal vein. However, in all our analyses, including parsimony, *Encyrtcephalus* never grouped within Melanosomellini or as its sister group, but instead was mostly recovered as sister to Tanaostigmatidae, a result that requires more investigation. This was not among our benchmark criteria, but an example of where better taxon sampling may help to resolve conflicting relationships.

*Other bias and possible options to improve inferences.* No objective criterion has been proposed so far to determine what fraction of genes/sites should be removed from a dataset to converge to a “correct” topology and it is unlikely that such a criterion will emerge in the future. Therefore, we advocate that the IQ-TREE combined tree (Fig. 5) is the best compromise we could achieve with this dataset, current evolutionary models and inference methods. In addition, this tree is close to the topology inferred with parsimony (Fig. 3). Results may be improved in the future either with increasing taxonomic sampling and/or better evolutionary models, assuming that analyses are computationally tractable (it was, for example, not feasible to use a CAT-GTR model on our dataset). However, the increased taxonomic sampling of the AHE520 dataset yielded nearly identical results both with maximum-likelihood and parsimony approaches (Figs 3 and S1). Bias will be difficult to track and alternative relationships hard to evaluate given the extreme plasticity of morphological characters in chalcid wasps. Indeed, the astounding diversity of morphologies that evolved in *c.* 160 Myr will continue to complicate the finding of strong synapomorphies to support many of the groups. To the best of our knowledge, no morphological analysis provides convincing evidence to reject the global topology inferred with the molecular dataset, but there are data suggesting that alternative placements of a few groups are as plausible as those recovered here.

#### *Validity of current families*

The complete revision, that relies on the present phylogenetic hypothesis and in which 50 extant families are now recognized has been published elsewhere (Burks et al., 2022). The main changes to the familial classification that resulted from this study are discussed below. To help readers, new family names are mapped on Fig. 5, and stem and crown ages are listed in Table S3c. Of the 25 previously (before Burks et al., 2022) recognized (extant) families, seven were recovered as paraphyletic or polyphyletic in the least biased molecular datasets, which confirms both rampant morphological convergence within Chalcidoidea and the

problem of inadequate diagnoses of families such as the former characterization of Pteromalidae.

*Aphelinidae.* Calesinae clustered away from other aphelinids in all molecular analyses despite morphological affinities (Heraty et al., 2013). Therefore, Aphelinidae should be restricted to the four subfamilies Aphelininae, Coccophaginae, Eretmocerinae and Eriaphytinae, whereas Calesinae is upgraded to family rank (Burks et al., 2022).

*Agaonidae.* Following discussion in the previous section, Sycophaginae is removed from Agaonidae and is placed into Pteromalidae sensu lato (s.l.) (Burks et al., 2022).

*Chalcididae.* Until proven otherwise and following the discussion in the previous section, Cratocentrinae is maintained within Chalcididae.

*Eulophidae.* As previously reported with molecular data (Burks et al., 2011; Munro et al., 2011; Heraty et al., 2013; Rasplus et al., 2020), *Trisecodes* (*incertae sedis* within Eulophidae) never clustered with other eulophids, but its placement remains ambiguous. *Trisecodes* is either recovered as sister to Systasini (exonsAA 433/520, combined) or Trichogrammatidae (UCEs90-25). *Trisecodes* exhibits the 3-segmented tarsi of Trichogrammatidae, but this is a characteristic that has occurred several times independently across Chalcidoidea. *Trisecodes* shares with Systasini the presence of a mesofurcal pit on the mesotrochantal plate between the mesocoxal insertions, which suggests a closer relationship between the two groups. Defining a family grouping for *Trisecodes* and Systasini (Systasidae) seemed the best solution, even though *Trisecodes* differs from Systasini in tarsomere and flagellomere count (Burks et al., 2022).

*Eupelmidae.* This family was never recovered as monophyletic in our analyses. No single morphological feature is unique to Eupelmidae (Gibson, 1989; Heraty et al., 2013), casting doubt on its validity. In all molecular trees, Neanastatinae (including *Metapelma*) and Calosotinae are polyphyletic, while Eupelminae are recovered as monophyletic. Furthermore, *Eopelma* never groups with other eupelmid clades and is instead consistently recovered as sister to *Storeya*, the unique genus of Storeyinae (Pteromalidae). The status and placement of the current genera and subfamilies of Eupelmidae are thoroughly discussed by Burks et al. (2022), but Metapelmidae and Neanastatidae were removed from Eupelmidae and treated as distinct families, and *Eopelma* was treated as *incertae sedis* in Chalcidoidea.

*Eriaporidae.* In all topologies, *Cecidellis* (Pteromalidae) renders Eriaporidae paraphyletic. However, although the



genus is well-defined by a short lamina covering the posterior propodeal margin in the female, it resembles some Eriaporidae in coloration and eye divergence, and Pireninae in body shape features. In addition, Eriaporidae was recovered as sister to Pireninae (Pteromalidae) in all our reconstructions, a group with which it shares several morphological characters. Therefore, a single family, Pirenidae containing Eriaporinae and *Cecidellis* in Cecidellinae was proposed by Burks et al. (2022).

**Pteromalidae.** Pteromalidae s.l. is scattered across all inferred trees. This hyperdiverse family that contains 33 subfamilies and nearly 650 genera has long been considered as a “taxonomic garbage can” for taxa that could not be easily included in other chalcid families (Burks et al., 2022). Pteromalidae has long been shown to be polyphyletic (Munro et al., 2011; Heraty et al., 2013; Peters et al., 2018; Zhang et al., 2020), but the lack of robust support for previous phylogenies has precluded any taxonomic rearrangement. Although the backbone of our Chalcidoidea tree remains unresolved in some places, shallower clades are strongly supported which has enabled a first revision of the former Pteromalidae (Burks et al., 2022). Our results do support a Pteromalidae sensu stricto (s.s.) that includes a number of odd fig-wasp parasitoids and the Sycophaginae that were previously treated as Agaonidae. Twenty families, that mostly correspond to the current subfamilies or tribes of pteromalids apart from Pteromalidae s.s., have thus been erected and their circumscription redefined by Burks et al. (2022). Most of these new families are ancient lineages (100–80 Ma; Table S3c) that are likely to have been grouped together within Pteromalidae based on symplesiomorphies, or because they had no apparent affinities with any other family. Other subfamilies have been included in other existing families that were redefined (e.g. Keiraninae and Chromeurytominae within Megastigmidae) or remained *incertae sedis* (e.g. Storeyinae) (Burks et al., 2022).

#### Higher level relationships

From here on, family names refer to the new classification (Burks et al., 2022).

Relationships in some places are still unresolved, although it is difficult to say whether it is to the consequence of a lack of signal or noise creating conflicting signal across the genome as previously suggested (Zhang et al., 2020). Depending on which measure of statistical support is used, the backbone is either a rake or moderately supported (Table S2e; Figs 4, 5 and S1). Only two backbone nodes are recovered in the set of gene trees (#1 and #2; Fig. 4), which is likely because of a lack of signal linked with the short length of loci that results in unresolved gene trees (Fig. 4b). Three backbone nodes (#4–6; Fig. 4) do not receive

any significant statistical support (gCF, sCF, UFBoot and SH-aLRT) and the corresponding part of the topology should be better regarded as a polytomy. As with the early evolution of birds (Suh, 2016), phylogenomics may not be able to resolve these difficult nodes because of near-simultaneous speciation. Indeed, from backbone node #2 (“Tiny Wasp clade”), one lineage appears, on average, every 3 Myr (Table S3b; Fig. 5) and the branching pattern is characterized by very short branches that are likely to lead to gene tree incongruence.

Below we discuss some of the higher-level relationships inferred with decreasing saturation in the individual datasets or emerged in the combined tree in agreement with hidden support in the individual datasets (Fig. 4a).

“Planidial clade” (*Eucharitidae* + *Perilampidae* + *Eutrichosomatidae* + *Chrysolampidae*). We confirm that families with planidial larvae (Zhang et al., 2022) form a monophyletic group that is recovered in all trees. Notably, with the inclusion of Eutrichosomatidae, planidia are defined by their behaviour (parasitoids that move across different larval instars of the host and then finish development externally on the host pupa) and not as a clade by their morphology (Zhang et al., 2022). There is no morphological support for this clade based on immatures or adults. Statistical support for this clade in the combined tree is high (SHaLRT = 100/UFBoot = 100/gCF = 0.197/sCF = 36.5) with, again, the exception of gCF. Sister to the “planidial clade” we recovered a clade of two old pteromalid lineages that split between 89 and 96 Ma that are now considered as one family (Spalangiidae; Burks et al., 2022). Only larvae of Spalangiinae have been described so far (Tormos et al., 2009) and, interestingly, *Spalangia* larvae appear to be mobile (Gerling and Legner, 1968), with a series of tubercles across the ventral region of body segments II–XII (Tormos et al., 2009). This latter feature also is documented in several lineages of the planidial clade [Chrysolampidae: Chrysolampinae (Askew, 1980; Darling and Miller, 1991); Philomidinae (Darling, 1992); and Eutrichosomatidae (Baker and Heraty, 2020)], which may corroborate the close relationship of these taxa and potentially expand the definition of the planidial clade.

“Gall clade” (*Cynipencyrtidae* + *Epichrysomallidae* + *Melanosomellidae* + *Ormyridae* + *Tanaostigmatidae*). This higher-level grouping is recovered here for the first time. Again, with the exception of gCF, statistical support in the combined tree is high (100/100/0.049/34.2). On the morphological side, there is limited evidence to support this clade that groups wasp lineages previously classified in four different families. However, from the perspective of life-history strategy all lineages



are gall-associated wasps. Melanosomellidae, Epichrysomallidae and Tanaostigmatidae are gall-makers associated with several groups of angiosperms (e.g. *Nothofagus*, *Casuarina*, *Eucalyptus*, *Ficus* and legumes, among others; LaSalle, 1987; Bouček, 1988; Beardsley and Rasplus, 2001; LaSalle, 2005). *Asparagobius* is recovered as sister to *Ormyrus* and also is a gall-maker on *Asparagus* in Africa. *Ormyrus* has been demonstrated to be a parasitoid of gall-makers (Gomez et al., 2017), while *Cynipencyrtus* is a parasitoid of either gall-forming Cynipidae or their inquiline (Ito and Hiji, 2000).

“Tiny wasps” (*Aphelinidae* + *Azotidae* + *Calesidae* + *Neodiparidae* + *Encyrtidae* + *Eulophidae* + *Eunotidae* + *Idioporidae* + *Signiphoridae* + *Trichogrammatidae*). This higher-level grouping is highlighted for the first time although it was recovered as monophyletic only in the IQ-TREE combined tree with moderate support (100/74/0/30.9). However, the grouping of these taxa near the base of Chalcidoidea, either as a monophyletic or polyphyletic group, with the exclusion of Mymaridae and Baeomorphae, is maintained across all of our analyses (Fig. S1). The backbone node (#3) that splits Mymaridae/Baeomorphae/“Tiny Wasp clade” from the other Chalcidoidea is moderately supported (100/73/0/37.9) and receives hidden support from the exonsAA (sCF = 38.6) and the UCEs90-25 (sCF = 40.4) datasets (Fig. S1). Statistical support for the group is possibly weakened by the ambiguous placement of the rogue *Trisecodes* (cf. previous section) that is either nested within this clade (UCEs) or that clusters with Systasidae and Tetracampidae (exonsAA). Tiny wasps form a grade in the TNT combined tree (Fig. 3). From a morphological point of view, the “Tiny Wasp clade” has hardly any support other than usually being small and mostly soft-bodied (except some Eulophidae). The group comprises several lineages characterized by a reduction in the number of flagellomeres or tarsomeres, and the frequent presence of a mesophragma that extends into the metasoma through a broad union with the mesosoma. Biologies are diverse in this group and there are no clear trends. With the exception of Eulophidae, which parasitize nearly all insect orders, and Trichogrammatidae, which are egg parasitoids, lineages of the “Tiny Wasp clade” are more frequently associated with Hemiptera. They appear to be mainly endoparasitoids of exophytic hosts (such as mealybugs), while other chalcid wasps are more frequently ectoparasitoids of endophytic hosts. In the same manner, species of this clade also appear to be more frequently oophagous than other chalcid wasps. Confirmation of this clade as a monophyletic lineage will require increasing taxonomic sampling, which may help to stabilize the placement of *Trisecodes* and would

greatly facilitate a formal study of the evolution of host-associations within Chalcidoidea.

“Weird clade” (*Megastigmidae* + *Leucospidae* + *Agaonidae* + *Metapelmatidae* + *Eurytomidae* + *Chalcididae* + several poorly diversified families). This higher-level grouping was unexpected based on morphology or biology. Yet, this “Weird clade” is statistically slightly better supported than the “Tiny Wasp clade” in the combined tree (99.2/95/0/35.6). Hidden sCF support for this clade is 35.9 for UCEs and 34.3 for exonsAA. From a morphological point of view the “Weird clade”, as its name indicates, groups disparate lineages of chalcid wasps that exhibit contrasting morphologies that could be correlated with their diverse biologies (Figs 4 and 5). Many families, such as Boucekiidae, Chalcedectidae, Cleonymidae, Lyciscidae, Metapelmatidae and Pelecinnellidae, but also several genera of Eurytomidae and a few Chalcididae, are parasitoids of xylophagous insects (mostly Coleoptera). Leucospidae have shifted to solitary bees and wasps that nest in wood or in mud nests, but *L. dorsigera* also has been reported as a hyperparasitoid of xylophagous beetles (Hesami et al., 2005). Agaonidae, Megastigmidae and Eurytomidae are mostly phytophagous, but the last two families also comprise parasitoid species. Agaonidae enter figs (inflorescences of *Ficus*, Moraceae) through a small aperture called the ostiole and exhibit strong morphological adaptations (e.g. mandibular appendage, anelli fused into a hook-like process, short protibia with spurs; Cruaud et al., 2010). Some lineages have more specialized biology such as Moranilidae that are predators of mealybug eggs, Asaphesinae (*incertae sedis*) that are hyperparasitoids of aphids through Braconidae wasps, and Enoggerinae (*incertae sedis*) that are oophagous parasitoids of Coleoptera. Coelocybidae are gall-associated wasps. Finally, Chalcididae attack nearly all insect orders, but mostly Lepidoptera.

*Sister group to Chalcidoidea.* Finally, with the exception of the exonsAA and AHE520AA trees (weak support for Diapriidae as sister to Chalcidoidea), all of our results strongly support the sister-group relationship between Mymarommatoidea and Chalcidoidea. Gibson (1986a) was the first to propose a sister-group relationship between these two superfamilies. Munro et al. (2011) recovered Diapriidae + Mymarommatoidea as sister to Chalcidoidea, although when sequence data were combined with morphology (Heraty et al., 2013), mymarommatoidea were sister to Chalcidoidea. A sister relationship with mymarommatoidea also was recovered in the UCE results of Blaimer et al. (2023).

*Soft- and hard-bodied chalcid wasps.* Morphology is rather uninformative in its support for most higher-level relationships. Nevertheless, we find a dichotomy between early diverging lineages that are small “soft-

bodied” chalcid wasps prone to shrivelling when air-dried (Mymaridae, Baeomorphidae, “Tiny Wasp clade”, Pirenidae) and “hard-bodied” chalcid wasps that diversified later (Figs 4 and 5).

*Morphological convergences.* Higher level morphological convergences are confirmed as exemplified by the evolution of an enlarged acropleuron and correlative transformation of legs linked to the ability to jump that may have happened at least seven times independently during the evolutionary history of chalcid wasps in (i) Encyrtidae, (ii) *Lambrodegma* and *Neanastatus*, (iii) *Eopelma*, (iv) *Metapelma*, (v) Tanaostigmatidae, (vi) *Cynipencyrtus* and (vii) Calosotinae; all relationships except *Metapelma* with maximum support.

#### *Time line and historical biogeography*

An important result of our study is a revision of the temporal scale over which Chalcidoidea have evolved and dispersed throughout the world (Figs 5 and S4; Table S3). Likely because our taxonomic sampling is one or two orders of magnitude higher than most previous phylogenomic studies (Branstetter et al., 2017; Peters et al., 2017, 2018; Tang et al., 2019), we infer an older crown age for Chalcidoidea: 162.2 (154.0–170.0) Ma. Nevertheless, this age is close to the recent estimates made from the most-representative phylogeny of Hymenoptera [168.6–166.0 Ma (151.6–183.3), Blaimer et al., 2023].

*Compatibility of ages with the fossil record, Earth’s palaeogeological history and previous works.* Importantly, our estimates are compatible not only with the meagre fossil records of chalcid wasps, but also with Earth’s palaeo-geological history (Fig. 5). The oldest fossils of chalcid wasps belong to the so-called “soft-bodied” chalcid wasps (Mymaridae, Baeomorphidae and “Tiny Wasp clade”; Haas et al., 2018). The oldest putative chalcid fossil is *Minutoma yathribi* Kaddumi, 2005 from Jordanian amber (Albian, ~113.0–100.5 Ma; Kaddumi, 2005). But the uncertain affinities of this fossil precluded its use in our analyses. The oldest unambiguous chalcid fossils are a Mymaridae [*Myanmymar aresconoides*; Poinar Jr. and Huber, 2011] and a Baeomorphidae (*Baeomorpha liorum*; Huber et al., 2019) from Myanmar (Burmese) amber (minimum age 98.2 Ma; see rationale for calibration priors in the Supporting information Materials and Methods). Species of Baeomorphidae are also frequent in Cretaceous ambers of the Northern Hemisphere [in the retinites of Baikura (minimum age 94.3 Ma) and of Yantardakh (minimum age 83.6 Ma; Gumovsky et al., 2018), and Canadian ambers of Cedar Lake and Grassy Lake, which are Campanian in age (83.6–72.1 Ma; McKellar et al., 2008)].

Only a dozen fossils of “hard-bodied” Chalcidoidea are known from Cretaceous formations. Among them, three were not used as calibrations because they have uncertain affinities that prevent them from being assigned to a clade. Nevertheless, all of them fit relatively well within the proposed time frame for Chalcidoidea. Diversinitidae (Myanmar amber) possibly belongs to the group of “hard bodied” taxa (Haas et al., 2018); however, it has uncertain relationships with extant chalcid wasps. Two other undescribed fossils clearly belong to this large clade: (i) a few specimens with uncertain morphological affinities (Pirenidae or Micradelinae) (A. Gumovsky and M.-D. Mitroiu, pers. comm.) from Taimyr amber (86.3–83.6 Ma), and (ii) an unidentified “torymid” specimen from Canadian amber (83.6–72.1 Ma) [fig. 4c in McKellar and Engel, 2012, that probably belongs instead to an extinct lineage].

Splits between Neotropical and Australian lineages also corroborated our new time frame for chalcid wasps. Indeed, several clades whose ancestor probably dispersed through the Antarctic land bridge pre-date connection break-ups and temperature decreases (i.e. 45 Ma; van den Ende et al., 2017). Thus, the clade grouping the Australian *Aeschylia* with the Neotropical *Aditrochus* and *Plastobelyta* (Melanosomellidae), all of which are gallers on *Nothofagus*, is dated at 60.1 Ma (Fig. S4; Table S3b), while the split between the Australian *Liepara* and the south Andean *Lanthanomyia* (Coelocybidae) that is, among others, a parasite of *Aditrochus* species is dated at 46.2 Ma. The split between the Neotropical *Erotolepsia* and its Australian/Oriental sister *Papuopsia* (Spalangiidae) is estimated at 49.9 Ma. The stem age of Neotropical Lyciscidae that are nested within an Australian clade is estimated at 33 Ma, which is a very late time to cross Antarctica.

Only a few dating analyses have been performed for groups of chalcid wasps. Variability in mean age estimates were noted depending on the study, datasets and methods used (differences up to 12 Myr). Our estimates for the age of the planidial clade [94.3 Ma (85.6–103.6)] and Eucharitidae [78.0 Ma (69.2–87.4)] are close to previous estimates (Murray et al., 2013; Zhang et al., 2022). The mean age for crown Agaonidae at 60.7 Ma (50.7–70.9) is 15 Myr younger than previous estimates of 75 Ma (94.9–56.2; Cruaud et al., 2012b) but this difference is likely to have been a consequence of the reduced set of taxa used in our study. Nevertheless, ages of all other groups of wasps that are strictly associated with *Ficus* (Epichrysomallidae, Sycophaginae, and other pteromaline fig wasps) post-date the age of Agaonidae.

*Insights on the origin of Baeomorphidae.* Recent findings by Heraty et al. (2018), strongly suggest that Baeomorphidae may be egg parasitoids of Peloridiidae

(moss-feeding bugs) and/or of Myerslopiidae (tree-hoppers), two ancient families of Hemiptera, that are (respectively) sister to all other Auchenorrhyncha and to all other Membracoidea (Johnson et al., 2018). Interestingly, both peloridiids and myerslopiids also have an amphi-Pacific distribution and Myerslopiidae fossils were only found at Crato in Brazil (122.5–112.6 Ma). In addition, the split between peloridiids from Chile and New Zealand [98 Ma (46–155 Ma); Ye et al., 2019] is close to our estimate for Baeomorphidae and corroborates our time-scale for Chalcidoidea. The fossil record shows that Baeomorphidae were largely distributed not only in Laurasia, but also on the Myanmar terrane, between the lower Cenomanian and the Campanian (98.2–72.1 Ma; Gumovsky et al., 2018; Huber et al., 2019), which suggests, in the framework of our scenario, a long northward dispersal. Our results contradict the hypothesis of a Laurasian origin for Baeomorphidae (Gumovsky et al., 2018) that has been made outside of a formal phylogenetic framework. However, given the area of origin for Mymaridae, and the current distribution of *Rotoita* and *Chiloe*, whatever the position of *Baeomorpha* + *Taimyromorpha* (fossil taxa) within Baeomorphidae, it would lead to a South Gondwanan origin for Baeomorphidae and a northward dispersal between 150 and 100 Ma. Their distribution fits well with that of fossils of Coleorrhyncha, the lineage to which Peloridiidae belongs and is now the only extant representative. Coleorrhyncha are divided into two large groups: the Progonocimicidoidea and the Peloridioidea. All extinct families of Peloridioidea occurred in Laurasia (between 201.6 and 112.6 Ma). The oldest fossil of Progonocimicidae occurred in Australia and was dated back to the Changhsingian, the uppermost stage of Permian (254.1–251.9 Ma). Progonocimicidae are recorded from Gondwanan and Laurasian Triassic formations (Evans, 1956) but became mostly confined to Laurasia later on (Jurassic and Cretaceous) with the notable exceptions of two Gondwanan fossils found in Cretaceous ambers (Lebanese and Myanmar; Szewo et al., 2011; Jiang et al., 2019). Most Coleorrhyncha groups declined during the mid-Cretaceous biotic crisis when vegetation was replaced by modern angiosperms and did not survive the Chicxulub impact (Dong et al., 2014). Our scenario suggests that baeomorphids expanded northward and experienced the same extinction event. We recognize that based on the ancient age and presence of numerous Laurasian fossils, their area of origin may be challenged but the retraction to the current relictual distribution is certain. Furthermore, the scarce Gondwanan deposits prevent to ascertain their absence in the Southern Hemisphere during the Early Cretaceous.

*Divergences that coincide with increases in insect fossil diversity and diversification of flowering plants.* The

origin of Chalcidoidea and divergence of the two first lineages (Baeomorphidae 153.1 Ma; “Tiny Wasp clade” 136.1 Ma) coincide with a rise in insect fossil families in the late Jurassic to the Hauterivian–Barremian (Schachat et al., 2019). The next splits on the backbone (from 110.3 Ma) coincide with the second sharp increase in fossil diversity through the Albian and Cenomanian (Schachat et al., 2019). This rapid radiation of “hard-bodied” Chalcidoidea, between 110 and 80 Ma, coincides with the onset and diversification of flowering plants and holometabolan insects. Although this hypothesis should be formally tested through a thorough compilation of host associations and reconstruction of ancestral life histories, the first lineages to diverge (Mymaridae, Baeomorphidae, “Tiny Wasp clade”) are first likely to have been oophagous and subsequently mostly associated with hemipteran hosts (Aphelinidae, Azotidae, Calesidae, Encyrtidae, Eunotidae, Signiphoridae). Recently it was discovered that the sister group to Chalcidoidea, Mymarommatoidae, are parasitoids of the eggs of Lepidopsocidae (Psocoptera; Honsberger et al., 2022), thus adding further support to an ancestral habit of egg parasitism for Chalcidoidea. Subsequently, chalcid wasps switched to virtually all orders and life stages of Holometabola. Interestingly, the first shifts to phytophagy (e.g. stem-crown gall clade: 102.1–98.2 Ma) correspond to the beginning of the “Angiosperm Terrestrial Revolution” c. 100 Ma (Benton et al., 2022).

*A first hypothesis on the global historical biogeography of chalcid wasps.* Our biogeographical scenario for Chalcidoidea should be regarded as a first hypothesis. Indeed, although our sampling is highly representative of the main lineages occurring on Earth (few suprageneric taxa are missing) and representative of the distribution of described genera (Noyes, 2019), several hyperdiverse clades have been scarcely sampled (e.g. Trichogrammatidae, Encyrtidae). Finally, neither model of biogeographical inference appears fully realistic to cover the entire range of biological traits that may regulate dispersal and therefore impact biogeographical processes. The size of chalcid wasps varies from 0.15 to >25 mm long, and the tiniest ones can easily be blown by the wind (Glick, 1939) or disperse actively over long distances ( $\leq 100$  km; (Grillenberger et al., 2009; Lander et al., 2014). Life-history characteristics also may be important in regulating dispersal; for example, idiobiont parasitoids are usually generalists, a characteristic that may facilitate establishment in newly colonized habitats. However, the success of long-distance dispersal remains to be demonstrated and relies mostly on host availability and quality.

Vicariance and therefore the support for the DEC/DIVA models, can be suspected when species of



chalcid wasps are subdivided geographically together with their ancestral host species. Such allopatric speciation appears to be a frequent scenario in chalcid wasps (e.g. Nicholls et al., 2010). Conversely, chalcid wasps can undergo rapid speciation by the emergence of ecotypes (Malec et al., 2021) or cytoplasmic incompatibility (Gebiola et al., 2016) and thus may be prone to sympatric speciation (König et al., 2019). Logically, the range of a parasitoid species will be limited by the range of its host(s). The higher statistical support for the BAYAREALIKE + *J* model suggests an important role for sympatric speciation, a rapid colonization of the widespread host range and long-distance dispersal. Because the relative contribution of the different speciation modes in chalcid wasps is unknown, we took a conservative position and considered an East-Gondwanan origin for chalcid wasps that is inferred by most models (DEC, DEC + *J*, BAYAREALIKE + *J*; DIVALIKE + *J*) as the most likely.

Although there is no formal biogeographical analysis of Mymaridae published yet, their East-Gondwanan origin (inferred by all models) is reasonable. Indeed, the Australian fauna contains the earliest diverging lineages of Mymaridae as well as a large diversity of genera that encompasses all familial subgroups (Lin et al., 2007). Extant genera of Baeomorphidae occur in southern temperate rainforests of Chile (*Chiloe*) and New Zealand (*Rotoita*) and show a disjunct amph-Pacific distribution (van den Ende et al., 2017). The split between these genera (91.7 Ma) appears contemporaneous with the breakup of southeast Gondwana and with the drift of Zealandia away from Antarctica (between 95 and 84 Ma) (Schellart et al., 2006; Mortimer et al., 2017). The conflicting hypothesis of the Laurasian origin of Baeomorphidae is discussed in the Supporting information.

Colonization of the rest of the world by the ancestors of other lineages is more or less delayed in time depending on the model used. Models that suppose identical inheritance of ancestral distributions by the two descendants immediately after speciation (BAYAREALIKE) suggest that the Cretaceous and rapid radiation of Chalcidoidea occurred in southern Gondwana and favour a subsequent "multiple dispersal out of Gondwana" hypothesis for younger lineages. Models that suppose allopatric speciation and vicariance (DEC, DIVA) infer a rapid colonization of the Northern Hemisphere (ancestors of all Chalcidoidea but Mymaridae and Baeomorphidae) with subsequent multiple recolonizations of southern Gondwana (Australia and/or Neotropics) during the radiation of the group. As detailed above, assumptions made by biogeographical models do not fully describe speciation modes in chalcid wasps. Therefore, and with regard to elements discussed below, we consider the scenario inferred by

BAYAREALIKE + *J*, which furthermore was selected by statistical tests, as plausible.

The widespread distribution of most groups included in the "Tiny Wasp clade" makes corroboration of a southern Gondwana origin (inferred by DEC + *J*, DIVALIKE + *J*, and BAYAREALIKE without or +*J*) difficult. Diversity should not be considered as a clue for origin. However, the diversity and placement of some Australian or Neotropical endemic taxa in the topology support a southern Gondwana origin (Fig. S5). Within Eulophidae, *Ophelimus* and *Perthiola* (mainly Australian) are sister to all Entiinae, while *Aleuroctonus* and *Dasyomphale* (Neotropical) are the first lineages to diverge within Euderomphalini. In addition, *Cales* is subdivided in two species groups, one occurring mostly in Australasia, the other in the Neotropics (Mottern et al., 2011), therefore it possibly originates in southern Gondwana. Conversely, a Palaeartic origin for the "Tiny Wasp clade" (inferred by DEC and DIVALIKE) supposes a fast recolonization of the Neotropics through a pathway that is difficult to identify; possibly through Africa but no members of early diverging lineages within clades of tiny wasps currently in Africa.

A Gondwanan origin of the "hard-bodied" chalcid wasps (inferred by BAYAREALIKE without or +*J*) is well-supported by (i) the cosmopolitan lineages that have their early diverging taxa occurring in Australia, such as Eucharitidae with Akapalinae (Murray et al., 2013), Perilampidae with *Euperilampus* (Zhang et al., 2022), Pteromalidae s.s. with Sycophaginae (Cruaud et al., 2011), and Colotrechninae as well as Megastigmidae with Keiraninae and Chromeurytominae; and (ii) lineages that originated in southern Gondwana and are mostly Australian and Neotropical (Melanosomellidae, Lyciscidae, Coelocybidae). Additionally, two other families, Boucekiidae (represented by *Boucekius*) and Pelecinellidae (*Leptofoenus*) could be included here as they have Australian representatives (*Chalcidiscelis* and *Doddifoenus*, respectively), that, unfortunately, could not be included in our sampling. Conversely, scenarios that infer a widespread ancestral area for the "hard-bodied" chalcid wasps (all areas for DEC without or +*J* or Afrotropical + Oriental + Palaeartic for DIVALIKE without or +*J*) suggest successive long distance dispersals and cladogenetic events between Laurasia and either the Neotropics or Australasia that seem less plausible given the distance between these areas, the specificity of the chalcid-host association, the larger size of these wasps, and the absence of African lineages in early diverging clades.

*On long distance dispersals and pathways of colonization of the northern hemisphere during the Cretaceous.* Scenarios that favour a South Gondwanan radiation (BAYAREALIKE without or +*J*)



identify several subfamilial or familial lineages that colonized the northern hemisphere (Palaeartic or Oriental region) during the Cretaceous, at a time when intercontinental dispersals were difficult. Four of them may derive from ancient dispersal events between Antarctica + Australia and Africa (where they also occur) or from a colonization of the Neotropics followed by dispersal to Africa and recent colonization of the Palaeartic region: Eunotidae between 125 and 65 Ma; the ancestor of Neodiparidae + Signiphoridae + Azotidae between 133 and 127 Ma; Ceidae between 102 and 53 Ma and Cleonymidae between 80 and 33 Ma. Two other apparent long-distance dispersal events are difficult to explain as no members of these clades are presently known outside the Palaeartic region (*Rivasia* and *Micradelus* between 81 and 77 Ma, and *Cynipencyrtus* at 90.6 Ma). Some other lineages possibly colonized the Northern Hemisphere through the Neotropics (*Exolabrum* and *Herbertia* between 104 and 92 Ma; *Trisecodes* between 86 and 41 Ma).

Finally, Torymidae was hypothesized to have originated in the Palaeartic region (Janšta et al., 2018). Scenarios that favour a South Gondwanan radiation suggest a dispersal from southern Gondwana to Laurasia between 106 and 80 Ma, possibly through South America and Africa. In this case, the first lineages of Torymidae have either disappeared from South America or have not yet been sampled there. Another possibility, that also may be the route used by *Eopelma* (c. 86.6 Ma) and a new subfamily close to Erotolepsiinae (c. 96 Ma) to reach the Northern Hemisphere (Sunda), would be through drifting India or through ancient terranes (Hall, 2012) or the Trans-Tethyan island arc that separated from northern Australia c. 120 Ma (Westerweel et al., 2019), bringing away the Gondwanan fauna (Poinar, 2018). Conversely, the Trans-Tethyan island arc might have been used to colonize southern continents for scenarios that infer a Gondwanan + Laurasian radiation, though as previously emphasized, multiple recolonizations of southern Gondwana seem less likely.

## Conclusion

Chalcidoidea has deep origins in the middle Jurassic (95% CI for stem age = 167.3–180.5 Ma) followed by a tremendous diversification in the Palaeogene concomitant with the radiation of plants and the insects that feed upon them. The combined analysis of the exonsAA and UCEs90-25 datasets (2054 loci and 284 106 sites) produced a generally well-supported hypothesis. In some cases, we found clades that matched more with natural history over their morphological support and in others we found some historical morphological groups (i.e. Eupelmidae, Cleonyminae)

that were never recovered in our phylogenomic results, which confirm previous analyses (Heraty et al., 2013). In either case, we need to further explore these groups to find the basis of disagreement. Notably, we found a clade of gall-associated wasps that was not previously envisioned, but is indeed a good example of molecules suggesting a clade that is agreeable with re-examined morphology (Burks et al., 2022). Our biogeographical inference hypothesizes a general dispersal of taxa from a southern Gondwanan origin; however, a much larger sampling is required to assess this. Importantly, even with large and independent datasets, the importance of taxonomic evaluation and attempts to reduce saturation and homoplasy were all important factors in developing a concrete phylogenetic hypothesis for this massive radiation.

## Acknowledgements

JYR and AC are grateful to Audrey Weber (INRAE, France) for sequencing of the UCE libraries and to the Genotoul bioinformatics platform Toulouse Midi-Pyrénées, France for providing computing resources. We thank Gary Gibson (Agriculture and Agri-Food, Canada), Paul Hanson (Univ. de Costa Rica), Christopher Darling (Univ. of Toronto, Canada), Nicole Fisher (CSIRO, Australia), Michael Gates (USDA, USA), Michael Haas (Univ. of Marburg, Germany), Christer Hansson (Museum of Biology, Sweden), Jason Mottern (USDA, USA), John D. Pinto (UCR, USA), Stefan Schmidt (ZSM, Germany), Christine Lambkin, Chris Burwell and Susan Wright (QM, Australia) for providing specimens and for helpful discussion. We thank Matt Buffington and two anonymous reviewers for their comments on the final manuscript. AC and JYR acknowledge the Queensland parks and wildlife services for collecting permits (WITK18278817-WIF418664617). We dedicate this work to the memory of our dear friend and colleague John LaSalle, specialist of Eulophidae, who was an enthusiastic member of this project.

## Funding

This work was supported by the NSF DEB-1555808 to JMH, JBW and MY; the ANR projects TRIPTIC (ANR-14-CE18-0002), BIDIME (ANR-19-ECOM-0010) and recurring funding of the INRAE to AC and JYR.

## Conflict of interest

None declared.

## Data availability statement

Raw paired reads were uploaded as NCBI Sequence Read Archives (PRJNA884376 for AHE and PRJNA1017994 for UCEs). Data matrices are available on Zenodo (<https://doi.org/10.5281/zenodo.8362523>).

## REFERENCES

- Andrews, S., 2010. FastQC—A quality control application for FastQ files. Available from <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- Askew, R.R., 1980. The biology and larval morphology of *Chrysolampus thenae* (Walker) (Hymenoptera, Pteromalidae). *Entomol. Mon. Mag.* 115, 155–159.
- Austin, A.D., Gibson, G.A.P. and Harvey, M.S., 1998. Synopsis of Australian *Calymnophilus* Masi (Hymenoptera: Eupelmidae), description of a new Western Australian species associated with a pseudoscorpion, and review of pseudoscorpions parasites. *J. Nat. Hist.* 32, 329–350.
- Baker, A.J. and Heraty, J.M., 2020. Larval morphology and life history of *Eutrichosoma mirabile* Ashmead and description of a new species of *Eutrichosoma* (Hymenoptera, Chalcidoidea). *J. Hymenopt. Res.* 75, 67–85. <https://doi.org/10.3897/jhr.75.47880>.
- Baker, A.J., Heraty, J.M., Mottern, J., Zhang, J., Hines, H.M., Lemmon, A.R. and Lemmon, E.M., 2020. Inverse dispersal patterns in a group of ant parasitoids (Hymenoptera: Eucharitidae: Oraseminae) and their ant hosts. *Syst. Entomol.* 45, 1–19. <https://doi.org/10.1111/syen.12371>.
- Ballesteros, J.A. and Hormiga, G., 2016. A new orthology assessment method for phylogenomic data: unrooted phylogenetic orthology. *Mol. Biol. Evol.* 33, 2117–2134.
- Beardsley, J.W. and Rasplus, J.Y., 2001. A new species of *Josephiella* (Hymenoptera: Agaonidae) forming leaf galls on *Ficus microcarpa* L. (Moraceae). *J. Nat. Hist.* 35, 33–40.
- Benton, M.J., Wilf, P. and Sauquet, H., 2022. The angiosperm terrestrial revolution and the origins of modern biodiversity. *New Phytol.* 233, 2017–2035.
- Blaimer, B.B., LaPolla, J.S., Branstetter, M.G., Lloyd, M.W. and Brady, S.G., 2016a. Phylogenomics, biogeography and diversification of obligate mealybug-tending ants in the genus *Acropyga*. *Mol. Phylogenet. Evol.* 102, 20–29.
- Blaimer, B.B., Lloyd, M.W., Guillory, W.X. and Brady, S.G., 2016b. Sequence capture and phylogenetic utility of genomic ultraconserved elements obtained from pinned insect specimens. *PLoS ONE* 11, e0161531.
- Blaimer, B.B., Santos, B.F., Cruaud, A., Gates, M.W., Kula, R.R., Mikó, I., Rasplus, J.-Y., Smith, D.R., Talamas, E.J., Brady, S.G. and Buffington, M.L., 2023. Key innovations and the diversification of Hymenoptera. *Nat. Commun.* 14, 1212.
- Böhmová, J., Rasplus, J.-Y., Taylor, G.S. and Janšta, P., 2022. Description of two new Australian genera of Megastigmidae (Hymenoptera: Chalcidoidea) with notes on the biology of the genus *Bortesia*. *J. Hymenopt. Res.* 90, 75–99.
- Bolger, A.M., Lohse, M. and Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120.
- Borowiec, M.L., 2016. AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ* 4, e1660.
- Borowiec, M.L., 2019. Convergent evolution of the army ant syndrome and congruence in big-data phylogenetics. *Syst. Biol.* 68, 642–656.
- Borowiec, M.L., Lee, E.K., Chiu, J.C. and Plachetzki, D.C., 2015. Extracting phylogenetic signal and accounting for bias in whole-genome data sets supports the Ctenophora as sister to remaining Metazoa. *BMC Genomics* 16, 987.
- Borowiec, M.L., Rabeling, C., Brady, S.G., Fisher, B.L., Schultz, T.R. and Ward, P.S., 2019. Compositional heterogeneity and outgroup choice influence the internal phylogeny of the ants. *Mol. Phylogenet. Evol.* 134, 111–121.
- Bouček, Z., 1988. Australasian Chalcidoidea (Hymenoptera): A Biosystematic Revision of Genera of Fourteen Families, With a Reclassification of Species. CAB International, Wallingford, Oxon, UK.
- Bouckaert, R., Vaughan, T.G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., Matschiner, M., Mendes, F.K., Müller, N.F., Ogilvie, H.A., du Plessis, L., Poppinga, A., Rambault, A., Rasmussen, D., Siveroni, I., Suchard, M.A., Wu, C.-H., Xie, D., Zhang, C., Stadler, T. and Drummond, A., 2019. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* 15, e1006650.
- Boussau, B., Walton, Z., Delgado, J.A., Collantes, F., Beani, L., Stewart, I.J., Cameron, S.A., Whitfield, J.B., Johnston, J.S., Holland, P.W.H., Bachtrog, D., Kathirithamby, J. and Huelsenbeck, J.P., 2014. Strepsiptera, phylogenomics and the Long Branch attraction problem. *PLoS ONE* 9, e107709.
- Branstetter, M.G., Danforth, B.N., Pitts, J.P., Faircloth, B.C., Ward, P.S., Buffington, M.L., Gates, M.W., Kula, R.R. and Brady, S.G., 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, J.W. and Smith, S.A., 2018. The past sure is tense: on interpreting phylogenetic divergence time estimates. *Syst. Biol.* 67, 340–353.
- Burks, R.A., Heraty, J.M., Gebiola, M. and Hansson, C., 2011. Combined molecular and morphological phylogeny of Eulophidae (Hymenoptera: Chalcidoidea), with focus on the subfamily Entedoninae. *Cladistics* 27, 1–25.
- Burks, R., Mitroiu, M.D., Fusu, L., Heraty, J.M., Janšta, P., Heydon, S., Dale-Skey Papilloud, N., Peters, R.S., Woolley, J.B., van Noort, S., Tselikh, E., Baur, A., Cruaud, A., Darling, C., Haas, M., Hanson, P., Krogmann, L. and Rasplus, J.-Y., 2022. From hell's heart I stab at thee! A determined approach towards a monophyletic Pteromalidae and reclassification of Chalcidoidea (Hymenoptera). *J. Hymenopt. Res.* 94, 13–88.
- Crotty, S., Minh, B., Bean, N., Holland, B., Tuke, J., Jermin, L. and von Haeseler, A., 2020. GHOST: Recovering historical signal from heterotachously-evolved sequence alignments. *Syst. Biol.* 69, 249–264.
- Cruaud, A., Jabbour-Zahab, R., Genson, G., Cruaud, C., Couloux, A., Kjellberg, F., van Noort, S. and Rasplus, J.Y., 2010. Laying the foundations for a new classification of Agaonidae (Hymenoptera: Chalcidoidea), a multilocus phylogenetic approach. *Cladistics* 26, 359–387.
- Cruaud, A., Jabbour-Zahab, R., Genson, G., Kjellberg, F., Kobmoo, N., van Noort, S., Yang, D.R., Peng, Y.Q., Ubaidillah, R., Hanson, P.E., Santos-Mattos, O., Farache, F.H.A., Pereira, R.A.S., Kerdelhue, C. and Rasplus, J.Y., 2011. Phylogeny and evolution of life-history strategies in the Sycophaginae non-pollinating fig wasps (Hymenoptera, Chalcidoidea). *BMC Evol. Biol.* 11, 15. <https://doi.org/10.1186/1471-2148-11-178>.
- Cruaud, A., Jabbour-Zahab, R., Genson, G., Ungricht, S. and Rasplus, J.Y., 2012a. Testing the emergence of New Caledonia: Fig wasp mutualism as a case study and a review of evidence. *PLoS ONE* 7, e30941.
- Cruaud, A., Rønsted, N., Chantarasuwan, B., Chou, L.S., Clement, W., Couloux, A., Cousins, B., Forest, F., Genson, G., Harrison, R.D., Hossaert-McKey, M., Jabbour-Zahab, R., Jousset, E., Kerdelhue, C., Kjellberg, F., Lopez-Vaamonde, C., Peebles, J., Pereira, R.A.S., Schramm, T., Ubaidillah, R., van Noort, S., Weiblen, G.D., Yang, D.R., Yan-Qiong, P., Yodpinyanee, A., Libeskind-Hadas, R., Cook, J.M., Rasplus, J.Y. and Savolainen, V., 2012b. An extreme case of plant-insect co-diversification: figs and fig-pollinating wasps. *Syst. Biol.* 61, 1029–1047.

- Cruaud, A., Nidelet, S., Arnal, P., Weber, A., Fusu, L., Gumovsky, A., Huber, J., Polaszek, A. and Rasplus, J.-Y., 2019. Optimised DNA extraction and library preparation for small arthropods: application to target enrichment in chalcid wasps used for biocontrol. *Mol. Ecol. Resour.* 19, 702–710.
- Cruaud, A., Delvare, G., Nidelet, S., Sauné, L., Ratnasingham, S., Chartois, M., Blaimer, B.B., Gates, M., Brady, S.G., Faure, S., van Noort, S., Rossi, J.-P. and Rasplus, J.-Y., 2021. Ultra-conserved elements and morphology reciprocally illuminate conflicting phylogenetic hypotheses in Chalcidoidea (Hymenoptera, Chalcidoidea). *Cladistics* 37, 1–35.
- Darling, D.C., 1992. The life history and larval morphology of *Aperilampus* (Hymenoptera: Chalcidoidea) with a discussion of the phylogenetic affinities of the Philomidinae. *Syst. Entomol.* 17, 331–339.
- Darling, D.C. and Miller, T.D., 1991. Life history and larval morphology of *Chrysolampus* (Hymenoptera: Chalcidoidea: Chrysolampinae) in western North America. *Can. J. Zool.* 69, 2168–2177.
- Dong, Q.P., Yao, Y.Z. and Ren, D., 2014. New fossil Progonocimicidae (Hemiptera: Coleorrhyncha: Progonocimicoidea) from the upper Mesozoic of northeastern China, with a phylogeny of Coleorrhyncha. *Syst. Entomol.* 39, 773–782.
- Duchêne, D.A., Mather, N., Van Der Wal, C. and Ho, S.Y.W., 2022. Excluding loci with substitution saturation improves inferences from phylogenomic data. *Syst. Biol.* 71, 676–689.
- van den Ende, C., White, L.T. and van Welzen, P.C., 2017. The existence and break-up of the Antarctic land bridge as indicated by both amphi-Pacific distributions and tectonics. *Gondwana Res.* 44, 219–227.
- Evans, J.W., 1956. Palaeozoic and Mesozoic Hemiptera (Insecta). *Aust. J. Zool.* 4, 164–258.
- Faircloth, B.C., Branstetter, M.G., White, N.D. and Brady, S.G., 2015. Target enrichment of ultraconserved elements from arthropods provides a genomic perspective on relationships among Hymenoptera. *Mol. Ecol. Resour.* 15, 489–501.
- Gates, M.W., 2008. Species revision and generic systematics of World Rileyinae (Hymenoptera: Eurytomidae). *Univ. Calif. Publ. Ent.* 127, 1–332.
- Gebiola, M., Kelly, S.E., Hammerstein, P., Giorgini, M. and Hunter, M.S., 2016. “Darwin’s corollary” and cytoplasmic incompatibility induced by *Cardinium* may contribute to speciation in *Encarsia* wasps (Hymenoptera: Aphelinidae). *Evolution* 70, 2447–2458.
- Gerling, D. and Legner, E.F., 1968. Developmental history and reproduction of *Spalangia cameroni*, a parasite of synanthropic flies. *Ann. Entomol. Soc. Am.* 61, 1436–1443.
- Gibson, G.A.P., 1986a. Evidence for monophyly and relationships of Chalcidoidea, Mymaridae, and Mymaromatidae (Hymenoptera, Terebrantes). *Can. Entomol.* 118, 205–240.
- Gibson, G.A.P., 1986b. Mesothoracic skeletomusculature and mechanics of flight and jumping in Eupelminae (Hymenoptera, Chalcidoidea, Eupelmidae). *Can. Entomol.* 118, 691–728.
- Gibson, G.A.P., 1989. Phylogeny and classification of Eupelmidae, with a revision of the world genera of Calosotinae and Metapelmatinae (Hymenoptera: Chalcidoidea). *Memoirs Entomol. Soc. Canada* 149, 1–121.
- Gibson, G.A.P., 2008. Description of *Leptoomus janzeni*, n. gen. and n. sp. (Hymenoptera: Chalcidoidea) from Baltic amber, and discussion of its relationships and classification relative to Eupelmidae, Tanaostigmatidae and Encyrtidae. *Zootaxa* 1730, 1–26.
- Gibson, G.A.P. and Huber, J.T., 2000. Review of the family Rotoitidae (Hymenoptera: Chalcidoidea), with description of a new genus and species from Chile. *J. Nat. Hist.* 34, 2293–2314.
- Gibson, G.A.P., Heraty, J.M. and Woolley, J.B., 1999. Phylogenetics and classification of Chalcidoidea and Mymarommatidae—a review of current concepts (Hymenoptera, Apocrita). *Zool. Scripta* 28, 87–124.
- Glick, P.A., 1939. The distribution of insects, spiders, and mites in the air. *Technical Bull.* 673, 1–150.
- Goloboff, P.A., Farris, J.S. and Nixon, K.C., 2008. TNT, a free program for phylogenetic analysis. *Cladistics* 24, 1–13.
- Gomez, J.F., Nieves, M.H., Gayubo, S.F. and Nieves-Aldrey, J.L., 2017. Terminal-instar larval systematics and biology of west European species of Ormyridae associated with insect galls (Hymenoptera, Chalcidoidea). *Zookeys* 644, 51–88.
- Grabherr, M., Haas, B., Yassour, M., Levin, J., Thompson, D., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B., Nusbaum, C., Lindblad-Toh, K., Friedman, N. and Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29, 644–652.
- Grillenberger, B.K., Gadau, J., Bijlsma, R., van de Zande, L. and Beukeboom, L.W., 2009. Female dispersal and isolation-by-distance of *Nasonia vitripennis* populations in a local mate competition context. *Entomol. Exp. Appl.* 132, 147–154.
- Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W. and Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321.
- Gumovsky, A., Perkovsky, E. and Rasnitsyn, A., 2018. Laurasian ancestors and “Gondwanan” descendants of Rotoitidae (Hymenoptera: Chalcidoidea): what a review of late cretaceous *Baeomorpha* revealed. *Cretaceous Res.* 84, 286–322.
- Hall, R., 2012. Late Jurassic-Cenozoic reconstructions of the Indonesian region and the Indian Ocean. *Tectonophysics* 570–571, 1–41. <https://doi.org/10.1016/j.tecto.2012.04.021>.
- Haas, M., Burks, R.A. and Krogmann, L., 2018. A new lineage of cretaceous jewel wasps (Chalcidoidea: Diversinitidae). *PeerJ* 6, 30.
- Harris, R.S., 2007. Improved pairwise alignment of genomic DNA. Ph.D. Thesis. The Pennsylvania State University.
- Hartig, F., 2022. DHARMA: Residual diagnostics for hierarchical (multi-level/mixed) regression models. Available from: <https://CRAN.R-project.org/package=DHARMA>.
- Heath, T.A., Hedtke, S.M. and Hillis, D.M., 2008. Taxon sampling and the accuracy of phylogenetic analyses. *J. Syst. Evol.* 46, 239–257.
- Hennig, W., 1966. *Phylogenetic Systematics*. Chicago Illinois: University of Illinois Press.
- Heraty, J.M., 2009. Parasitoid biodiversity and insect pest management. In: Footitt, B. and Adler, P. (Eds.), *Insect Biodiversity: Science and Society*. Springer-Verlag Press, Netherlands, pp. 445–462.
- Heraty, J., Ronquist, F., Carpenter, J.M., Hawks, D., Schulmeister, S., Dowling, A.P., Murray, D., Munro, J., Wheeler, W.C., Schiff, N. and Sharkey, M., 2011. Evolution of the hymenopteran megaradiation. *Mol. Phylogenet. Evol.* 60, 73–88.
- Heraty, J.M., Burks, R.A., Cruaud, A., Gibson, G.A.P., Liljebblad, J., Munro, J., Rasplus, J.-Y., Delvare, G., Janšta, P., Gumovsky, A., Huber, J., Woolley, J.B., Krogmann, L., Heydon, S., Polaszek, A., Schmidt, S., Darling, D.C., Gates, M.W., Mottern, J., Murray, E., Dal Molin, A., Triapitsyn, S., Baur, H., Pinto, J.D., van Noort, S., George, J. and Yoder, M., 2013. A phylogenetic analysis of the megadiverse Chalcidoidea (Hymenoptera). *Cladistics* 29, 466–542.
- Heraty, J.M., Dominguez, C., Baker, A.J., Woolley, J.B., Krogmann, L. and Peters, R.S., 2018. Collecting “Moss wasps” (Rotoitidae) at the southern ends of the world. *Hamuli* 9, 1–4.
- Hesami, S.A., Akrami, M.A. and Baur, H., 2005. *Leucospis dorsigera* Fabricius (Hymenoptera, Leucospidae) as a hyperparasitoid of Cerambycidae (Coleoptera) through Xoridinae (Hymenoptera: Ichneumonidae) in Iran. *J. Hymenopt. Res.* 14, 66–68.
- Honsberger, D.N., Huber, J.T. and Wright, M.G., 2022. A new *Mymaromma* sp. (Mymarommatodea, Mymaromatidae) in Hawai‘i and first host record for the superfamily. *J. Hymenopt. Res.* 89, 73–87.
- Hothorn, T., Bretz, F. and Westfall, P., 2008. Simultaneous inference in general parametric models. *Biometr. J.* 50, 346–363.
- Huang, X. and Madan, A., 1999. CAP3: A DNA sequence assembly program. *Genome Res.* 9, 868–877.



- Huber, J.T., Shih, C. and Ren, D., 2019. A new species of *Baeomorpha* (Hymenoptera, Rotoitidae) from mid-cretaceous Burmese amber. *J. Hymenopt. Res.* 72, 1–10.
- Ito, M. and Hijii, N., 2000. Life-history traits in the parasitoid complex associated with cynipid galls on three species of Fagaceae. *Entomol. Sci.* 3, 471–479.
- Ives, A.R., 2015. For testing the significance of regression coefficients, go ahead and log-transform count data. *Methods Ecol. Evol.* 6, 828–835.
- Janšta, P., Cruaud, A., Delvare, G., Genson, G., Heraty, J., Křížková, B. and Rasplus, J.-Y., 2018. Torymidae (Hymenoptera, Chalcidoidea) revised: molecular phylogeny, circumscription and reclassification of the family with discussion of its biogeography and evolution of life-history traits. *Cladistics* 34, 627–651.
- Jiang, T., Wang, B. and Szwedo, J., 2019. The first representative of Progonocimicidae (Hemiptera: Coleorrhyncha) from mid-cretaceous Burmese amber. *Cretaceous Res.* 93, 346–359.
- Johnson, M.G., Gardner, E.M., Liu, Y., Medina, R., Goffinet, B., Shaw, A.J., Zerega, N.J.C. and Wickett, N.J., 2016. HybPiper: extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Appl. Plant Sci.* 4, 1600016.
- Johnson, K.P., Dietrich, C.H., Friedrich, F., Beutel, R.G., Wipfler, B., Peters, R.S., Allen, J.M., Petersen, M., Donath, A., Walden, K.K.O., Kozlov, A.M., Podsiadlowski, L., Mayer, C., Meusemann, K., Vasilikopoulos, A., Waterhouse, R.M., Cameron, S.L., Weirauch, C., Swanson, D.R., Percy, D.M., Hardy, N.B., Terry, I., Liu, S., Zhou, X., Misof, B., Robertson, H.M. and Yoshizawa, K., 2018. Phylogenomics and the evolution of hemipteroid insects. *Proc. Natl Acad. Sci. USA* 115, 12775–12780.
- Kaddumi, H.F., 2005. Amber of Jordan the oldest prehistoric insects in Fossilized Resin. 2nd edition. Amman: Eternal River Museum of Natural History, p. 224.
- Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A. and Jermini, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589.
- Katoh, K. and Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780.
- Klopfstein, S., Vilhelmsen, L., Heraty, J.M., Sharkey, M. and Ronquist, F., 2013. The hymenopteran tree of life: Evidence from protein-coding genes and objectively aligned ribosomal data. *PLoS ONE* 8, e69344.
- König, K., Zundel, P., Krimmer, E., König, C., Pollmann, M., Gottlieb, Y. and Steidle, J.L.M., 2019. Reproductive isolation due to prezygotic isolation and postzygotic cytoplasmic incompatibility in parasitoid wasps. *Ecol. Evol.* 9, 10694–10706.
- Kozlov, A.M., Darriba, D., Flouri, T., Morel, B. and Stamatakis, A., 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 35, 4453–4455.
- Kumar, S., Filipski, A.J., Battistuzzi, F.U., Pond, S.L.K. and Tamura, K., 2012. Statistics and truth in phylogenomics. *Mol. Biol. Evol.* 29, 457–472.
- Lander, T.A., Klein, E.K., Oddou-Muratario, S., Candau, J.N., Gido, C., Chalon, A., Roig, A., Fallour, D., Auger-Rozenberg, M.A. and Boivin, T., 2014. Reconstruction of a windborne insect invasion using a particle dispersal model, historical wind data, and Bayesian analysis of genetic data. *Ecol. Evol.* 4, 4609–4625.
- Lartillot, N., Rodrigue, N., Stubbs, D. and Richer, J., 2013. PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Syst. Biol.* 62, 611–615.
- LaSalle, J., 1987. New World Tanaostigmatidae (Hymenoptera, Chalcidoidea). *Biology* 23, 1–181.
- LaSalle, J., 2005. Biology of gall inducers and evolution of gall induction in Chalcidoidea (Hymenoptera: Eulophidae, Eurytomidae, Pteromalidae, Tanaostigmatidae, Torymidae). In: Raman, A., Schaefer, C.W. and Withers, T.M. (Eds.), *Biology, Ecology, and Evolution of Gall-Inducing Arthropods*. Sciences Publishers, Inc., Enfield, New Hampshire, pp. 503–533.
- Lenth, R.V., 2021. Emmeans: estimated marginal means, aka least-squares means. Available from: <https://CRAN.R-project.org/package=emmeans>.
- Lin, N.Q., Huber, J.T. and LaSalle, J., 2007. The Australian genera of Mymaridae (Hymenoptera: Chalcidoidea). *Zootaxa* 1596, 1–111.
- Lotfalizadeh, H., Delvare, G. and Rasplus, J.-Y., 2007. Phylogenetic analysis of Eurytominae (Chalcidoidea: Eurytomidae) based on morphological characters. *Zool. J. Linnean Soc.* 151, 441–510.
- Lüdecke, D., Ben-Shachar, M.S., Patil, I., Waggoner, P. and Makowski, D., 2021. Performance: an R package for assessment, comparison and testing of statistical models. *J. Open Sour. Softw.* 6, 3139.
- Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M. and Hornik, K., 2018. Cluster: Cluster analysis basics and extensions. R package version 2.0.7-1.
- Magoc, T. and Salzberg, S.L., 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27, 2957–2963.
- Mai, U. and Mirarab, S., 2018. TreeShrink: fast and accurate detection of outlier long branches in collections of phylogenetic trees. *BMC Genomics* 19, 272.
- Malec, P., Weber, J., Böhmer, R., Fiebig, M., Meinert, D., Rein, C., Reinisch, R., Henrich, M., Polyvas, V., Pollmann, M., von Berg, L., König, C. and Steidle, J., 2021. The emergence of ecotypes in a parasitoid wasp: a case of incipient sympatric speciation in Hymenoptera? *BMC Ecol. Evol.* 21, 204.
- 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.
- Matzke, N.J., 2022. Statistical comparison of DEC and DEC + J is identical to comparison of two ClaSSE submodels, and is therefore valid. *J. Biogeogr.* 49, 1805–1824.
- McKellar, R.C. and Engel, M.S., 2012. Hymenoptera in Canadian cretaceous amber (Insecta). *Cretaceous Res.* 35, 258–279.
- McKellar, R.C., Wolfe, A.P., Tappert, R. and Muehlenbachs, K., 2008. Correlation between grassy Lake and cedar Lake ambers using infrared spectroscopy, stable isotopes, and palaeoentomology. *Can. J. Earth Sci.* 45, 1061–1082.
- Minh, B.Q., Nguyen, M.A.T. and von Haeseler, A., 2013. Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* 30, 1188–1195.
- Minh, B.Q., Hahn, M.W. and Lanfear, R., 2020a. New methods to calculate concordance factors for phylogenomic datasets. *Mol. Biol. Evol.* 37, 2727–2733.
- Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A. and Lanfear, R., 2020b. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534.
- Mirarab, S., Nguyen, N. and Warnow, T., 2014. PASTA: ultra-large multiple sequence alignment. *Res. Comput. Mol. Biol.* 22, 177–191.
- Mooi, R. and Gill, A., 2016. Hennig's auxiliary principle and reciprocal illumination revisited. In: Williams, D., Schmitt, M. and Wheeler, Q. (Eds.), *The Future of Phylogenetic Systematics: The Legacy of Willi Hennig*. Systematics Association Special Volume Series. Cambridge University Press, Cambridge, pp. 258–285.
- Mortimer, N., Campbell, H.J., Tulloch, A.J., King, P.R., Stagpoole, V.M., Wood, R.A., Rattenbury, M.S., Sutherland, R., Adams, C.J., Collot, J. and Seton, M., 2017. Zealandia: Earth's hidden continent. *GSA Today* 27, 27–35.
- Mottern, J.L., Heraty, J.M. and Hartop, E., 2011. *Cales* (Hymenoptera: Chalcidoidea): morphology of an enigmatic taxon with a review of species. *Syst. Entomol.* 36, 267–284.
- Munro, J.B., Heraty, J.M., Burks, R.A., Hawks, D., Mottern, J., Cruaud, A., Rasplus, J.Y. and Janšta, P., 2011. A molecular phylogeny of the Chalcidoidea (Hymenoptera). *PLoS ONE* 6, e27023.

- Murray, E.A., Carmichael, A.E. and Heraty, J.M., 2013. Ancient host shifts followed by host conservatism in a group of ant parasitoids. *Proc. R. Soc. B Biol. Sci.* 280, 20130495.
- Nicholls, J.A., Preuss, S., Hayward, A., Melika, G., Csoka, G., Nieves-Aldrey, J.-L., Askew, R.R., Tavakoli, M., Schonrogge, K. and Stone, G.N., 2010. Concordant phylogeography and cryptic speciation in two Western Palaearctic oak gall parasitoid species complexes. *Mol. Ecol.* 19, 592–609.
- Noyes, J.S., 2019. Universal Chalcidoidea Database. World Wide Web Electronic Publication. <http://www.nhm.ac.uk/chalcidoids> (accessed September 2021).
- Noyes, J.S. and Hayat, M., 1994. *Oriental Mealybug Parasitoids of the Anagryini (Hymenoptera: Encyrtidae)*. CAB International, Oxon, UK. viii+554 pp.
- Paradis, E. and Schliep, K., 2018. Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35, 526–528.
- Peters, R.S., Krogmann, L., Mayer, C., Donath, A., Gunkel, S., Meusemann, K., Kozlov, A., Podsiadlowski, L., Petersen, M., Lanfear, R., Diez, P.A., Heraty, J., Kjer, K.M., Klopstein, S., Meier, R., Polidori, C., Schmitt, T., Liu, S., Zhou, X., Wappler, T., Rust, J., Misof, B. and Niehuis, O., 2017. Evolutionary history of the Hymenoptera. *Curr. Biol.* 27, 1013–1018.
- Peters, R.S., Niehuis, O., Gunkel, S., Bläser, M., Mayer, C., Podsiadlowski, L., Kozlov, A., Donath, A., van Noort, S., Liu, S., Zhou, X., Misof, B., Heraty, J. and Krogmann, L., 2018. Transcriptome sequence-based phylogeny of chalcidoid wasps (Hymenoptera: Chalcidoidea) reveals a history of rapid radiations, convergence, and evolutionary success. *Mol. Phylogenet. Evol.* 120, 286–296.
- Peterson, B.G. and Carl, P., 2018. PerformanceAnalytics: econometric tools for performance and risk analysis. R package version 12. <https://CRAN.R-project.org/package=PerformanceAnalytics>.
- Philippe, H. and Forterre, P., 1999. The rooting of the universal tree of life is not reliable. *J. Mol. Evol.* 49, 509–523.
- Philippe, H., de Vienne, D.M., Ranwez, V., Roure, B., Baurain, D. and Delsuc, F., 2017. Pitfalls in supermatrix phylogenomics. *Eur. J. Taxonomy* 283, 1–25.
- Poinar, G., 2018. Burmese amber: evidence of Gondwanan origin and cretaceous dispersion. *Hist. Biol.* 2963, 1–6.
- Poinar, G., Jr. and Huber, J.T., 2011. A new genus of fossil Mymaridae (Hymenoptera) from cretaceous amber and key to cretaceous mymarid genera. *Zookeys* 130, 461–472.
- R Core Team, 2018. R version 3.5.1 (Feather Spray): A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Rambaut, A., Drummond, A.J., Xie, D., Baele, G. and Suchard, M.A., 2018. Posterior summarization in Bayesian phylogenetics using tracer 1.7. *Syst. Biol.* 67, 901–904.
- Rasplus, J.-Y., Villemant, C., Paiva, M.R., Delvare, G. and Roques, A., 2010. Chapter 12. Hymenoptera. In: Roques, A., Kenis, M., Lees, D., Lopez-Vaamonde, C., Rabitsch, W., Rasplus, J.-Y. and Roy, D.B. (Eds.), *Alien Terrestrial Arthropods of Europe*. *BioRisk*. Pensoft Edition. Bulgaria: Sofia, pp. 669–766.
- Rasplus, J.-Y., Blaimer, B.B., Brady, S.G., Burks, R.A., Delvare, G., Fisher, N., Gates, M., Gauthier, N., Gumovsky, A.V., Hansson, C., Heraty, J.M., Fusu, L., Nidelet, S., Pereira, R.A.S., Sauné, L., Ubaidillah, R. and Cruaud, A., 2020. A first phylogenomic hypothesis for Eulophidae (Hymenoptera, Chalcidoidea). *J. Nat. Hist.* 54, 597–609. <https://doi.org/10.1080/00222933.2020.1762941>.
- Rasplus, J.-Y., Rodriguez, L.J., Sauné, L., Peng, Y.-Q., Bain, A., Kjellberg, F., Harrison, R.D., Pereira, R.A.S., Ubaidillah, R., Tollon-Cordet, C., Gautier, M., Rossi, J.-P. and Cruaud, A., 2021. Exploring systematic biases, rooting methods and morphological evidence to unravel the evolutionary history of the genus *Ficus* (Moraceae). *Cladistics* 37, 402–422.
- Rasplus, J.-Y., Nieves-Aldrey, J.-L. and Cruaud, A., 2022. *Cecinothofagus* Nieves-Aldrey & Liljebäck (Hymenoptera: Cynipidae) is likely an endoparasitoid of the gall-maker genus *Aditrochus* Rübsaamen (Hymenoptera: Pteromalidae). *J. Hymenopt. Res.* 93, 33–42.
- Reddy, S., Kimball, R.T., Pandey, A., Hosner, P.A., Braun, M.J., Hackett, S.J., Han, K.-L., Harshman, J., Huddleston, C., Kingston, S., Marks, B.D., Miglia, K.J., Moore, W.S., Sheldon, F.H., Witt, C.C., Yuri, T. and Braun, E.L., 2017. Why do phylogenomic data sets yield conflicting trees? Data type influences the avian tree of life more than taxon sampling. *Syst. Biol.* 66, 1063–1157.
- Ree, R.H. and Sanmartin, I., 2018. Conceptual and statistical problems with the DEC + J model of founder-event speciation and its comparison with DEC via model selection. *J. Biogeogr.* 45, 741–749.
- Ree, R.H. and Smith, S.A., 2008. Maximum-likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Syst. Biol.* 57, 4–14.
- Rice, P., Longden, I. and Bleasby, A., 2000. EMBOSS: the European molecular biology open software suite. *Trends Genet.* 16, 276–277.
- Robinson, D. and Foulds, L., 1981. Comparison of phylogenetic trees. *Math. Biosci.* 53, 131–147.
- Rodriguez, L.J., Cruaud, A. and Rasplus, J.-Y., 2021. Low sampling effort and high genetic isolation contribute to underdocumented diversity in Philippine fig wasps. *Philippine J. Sci.* 150, 173–180.
- Romiguier, J. and Roux, C., 2017. Analytical biases associated with GC-content in molecular evolution. *Front. Genet.* 8, 16.
- Ronquist, F., 1997. Dispersal-variance analysis: a new approach to the quantification of historical biogeography. *Syst. Biol.* 46, 195–203.
- Rougerie, R., Cruaud, A., Arnal, P., Ballesteros-Mejia, L., Condamine, F.L., Decaëns, T., Elias, M., Gey, D., Hebert, P.D.N., Kitching, I.J., Lavergne, S., Lopez-Vaamonde, C., Murienne, J., Cuenot, Y., Nidelet, S. and Rasplus, J.-Y., 2022. Phylogenomics illuminates the evolutionary history of wild silkmoths in space and time (Lepidoptera: Saturniidae). <https://doi.org/10.1101/2022.03.29.486224>.
- Schachat, S.R., Labandeira, C.C., Clapham, M.E. and Payne, J.L., 2019. A cretaceous peak in family-level insect diversity estimated with mark-recapture methodology. *Proc. R. Soc. B Biol. Sci.* 286, 20192054.
- Schellart, W.P., Lister, G.S. and Toy, V.G., 2006. A late cretaceous and Cenozoic reconstruction of the Southwest Pacific region: tectonics controlled by subduction and slab rollback processes. *Earth-Sci. Rev.* 76, 191–233.
- Struck, T.H., 2014. TreSpEx—detection of misleading signal in phylogenetic reconstructions based on tree information. *Evol. Bioinform.* 10, 51–67.
- Suh, A., 2016. The phylogenomic forest of bird trees contains a hard polytomy at the root of Neoaves. *Zool. Scripta* 45, 50–62.
- Szwedo, J., Azar, D. and Ziadé, K., 2011. The first Progonocimicidae (Insecta: Hemiptera: Coleorrhyncha) from lower cretaceous Lebanese amber. *Insect Syst. Evol.* 42, 161–177.
- Tang, P., Zhu, J.C., Zheng, B.Y., Wei, S.J., Sharkey, M., Chen, X.X. and Vogler, A.P., 2019. Mitochondrial phylogenomics of the Hymenoptera. *Mol. Phylogenet. Evol.* 131, 8–18.
- Tong, K.J., Duchêne, S., Ho, S.Y. and Lo, N., 2015. Comment on phylogenomics resolves the timing and pattern of insect evolution. *Science* 349, 487.
- Tormos, J., Beitia, F., Bockmann, E.A. and Asis, J.D., 2009. The preimaginal stages and development of *Spalangia cameroni* Perkins (Hymenoptera: Pteromalidae) on *Ceratitidis capitata* (Wiedemann) (Diptera: Tephritidae). *Micron* 40, 646–658.
- Westerweel, J., Roperch, P., Licht, A., Dupont-Nivet, G., Win, Z., Poblete, F., Ruffet, G., Swe, H.H., Thi, M.K. and Aung, D.W., 2019. Burma terrane part of the trans-Tethyan arc during collision with India according to palaeomagnetic data. *Nat. Geosci.* 12, 863–868.
- Whelan, N. and Halaných, K., 2017. Who let the CAT out of the bag? Accurately dealing with substitutional heterogeneity in phylogenomic analyses. *Syst. Biol.* 66, 232–255.
- Wickham, H., 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York.

- Wiens, J.J., 2004. The role of morphological data in phylogeny reconstruction. *Syst. Biol.* 53, 653–661.
- Wijesekara, G.A.W., 1997. Generic relationships within the tribes Cratocentrini and Phasganophorini (Hymenoptera: Chalcididae). *J. Hymenopt. Res.* 6, 297–335.
- Yang, Z. and Rannala, B., 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol. Biol. Evol.* 23, 212–226.
- Ye, Z., Damgaard, J., Burckhardt, D., Gibbs, G., Yuan, J., Yang, H. and Bu, W., 2019. Phylogeny and historical biogeography of Gondwanan moss-bugs (Insecta: Hemiptera: Coleorrhyncha: Peloridiidae). *Cladistics* 35, 135–149.
- Zhang, J., Lindsey, A.R.I., Peters, R.S., Heraty, J.M., Hopper, K.R., Werren, J.H., Martinson, E.O., Woolley, J.B., Yoder, M.J. and Krogman, L., 2020. Conflicting signal in transcriptomic markers leads to a poorly resolved backbone phylogeny of chalcidoid wasps. *Syst. Entomol.* 45, 783–802.
- Zhang, J., Heraty, J.M., Darling, C., Kresslein, R.L., Baker, A.J., Torr ns, J., Rasplus, J.-Y., Lemmon, A.R. and Lemmon, E.M., 2022. Anchored phylogenomics and a revised classification of the planidial larva clade of jewel wasps (Hymenoptera: Chalcidoidea). *Syst. Entomol.* 47, 329–353. <https://doi.org/10.1111/syen.12533>.

### Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

- Appendix S1** Trees  
**Figure S1** Phylogenetic trees  
**Figure S2** Dendrograms based on taxa properties  
**Figure S3** Spearman’s rank correlations between dataset/ tree properties  
**Figure S4** Time tree of Chalcidoidea  
**Figure S5** Historical biogeography of Chalcidoidea  
**Table S1** Classification and sampling  
**Table S2** Properties: datasets and trees  
**Table S3** Dating  
**Table S4** Biogeography  
**Table S5** Properties: taxa  
**Table S6** Phylogenetic relationships  
**Data S1** Materials and Methods