Received: 31 March 2025 | Accepted: 15 September 2025

DOI: 10.1111/syen.70009

ORIGINAL ARTICLE

Systematic Entomology | Royal Entomological Society

Integrating Sanger and next-generation sequencing data sheds light on phylogenetic relationships among gelechioid moths (Lepidoptera: Gelechioidea)

Etka Yapar[1] | Andrea Chiocchio[1,2] | Maria Heikkilä[3] | Jadranka Rota[1,4] | Lauri Kaila[3] | Niklas Wahlberg[1]

[1]Department of Biology, Lund University, Lund, Sweden

[2]Department of Biological and Ecological Sciences, Tuscia University, Viterbo, Italy

[3]Finnish Museum of Natural History Luomus, University of Helsinki, Helsinki, Finland

[4]Biological Museum, Lund University, Lund, Sweden

Correspondence
Etka Yapar, Department of Biology, Lund University, Kontaktvägen 10, 22362 Lund, Sweden.
Email: etka.yapar@biol.lu.se

Funding information
Crafoord Foundation (Crafoordska Stiftelsen), Grant/Award Number: 20180794; ELLIIT

Editor: Emmanuel Toussaint

Abstract

Accounting for the estimated number of undescribed species, Gelechioidea are thought to be among the most species-rich superfamilies of Lepidoptera, with 18,500 described species, including numerous pests of economically significant crops such as cotton, tomato and wheat. Family-level topology of the superfamily and the extent and the number of accepted families have received important revisions throughout the previous phylogenetic work on the group. Here we extracted 1767 nuclear protein-coding genes from genomic and transcriptomic data from 57 ingroup taxa, including *Haplochrois buvati* Baldizzone and *Urodeta hibernella* Staudinger, for which the whole-genome sequences were generated in this study. We first analyse this phylogenomic dataset within a maximum likelihood framework to revisit the interfamilial relationships within Gelechioidea and then integrate it with the existing taxon-rich, Sanger-sequenced data from up to 24 genes to re-evaluate the extent of the 20 currently accepted families by analysing this integrated dataset encompassing 381 ingroup taxa. Although we also recover some of the previously suggested multifamilial clades, the backbone topology we infer presents novel arrangements of families compared to previously published work: overall, we observe that Gelechioidea have diversified in four main lineages and find Stenomatinae (Depressariidae) to be sister to the rest of Gelechioidea. We therefore elevate it to family, Stenomatidae **stat. nov**. Moreover, we find the current circumscription of Elachistidae to be non-monophyletic and propose a new delimitation to include the subfamilies Elachistinae, Parametriotinae, Cacochroinae (Depressariidae) and Ethmiinae **stat. nov**.

KEYWORDS
classification, curved-horn moths, Elachistidae, phylogenomics, Stenomatidae, transcriptomics

INTRODUCTION

Gelechioidea are a highly diverse superfamily of apoditrysian Lepidoptera with around 18,500 described species (van Nieukerken et al., 2011). Species from this group encompass a wide variety of lifestyles and adaptations and play significant functional roles, from pollination to phytophagy, scavenging, predation, parasitism and even parasitoidism (Bidzilya & Rajaei, 2024; Kristensen & Skalski, 1999; Luz et al., 2015; Rubinoff & Haines, 2005). Moreover, larvae of Gelechioidea have a striking diversity in feeding modes and lifestyles and can engage in leaf mining, gall making and case bearing (Common, 1990; Kristensen & Skalski, 1999). This group also includes numerous

economically significant agricultural pests (Gözüaçık et al., 2008; Raine, 1966; White et al., 2007; Yan et al., 2019), as well as species that provide fundamental ecological services (Buxton et al., 2018; Luo et al., 2011). Despite their great ecological and economic significance, Gelechioidea are poorly studied relative to other Lepidoptera, especially in terms of undescribed species diversity, unresolved systematics and phylogenetic relationships. A small selection of adult gelechioid diversity can be seen in Figure 1.

The work of Hodges (1998) was the first of its time to adopt a cladistic approach through the analysis of a character matrix that represented family- or subfamily-level groups as terminals to unravel the relationships among the families of Gelechioidea, which put forwards major rearrangements of family-level groups within the superfamily. The first phylogeny with exemplar taxa for Gelechioidea was published by Kaila (2004), sampling 193 larval, pupal and adult morphological characters across 143 taxa, resulting in a tree that is divided into two main lineages—one of which had genera that were grouping together with Gelechiidae and the other with Oecophoridae-associated genera, groups which he referred to as gelechiid and oecophorid lineages, respectively. Although there were no

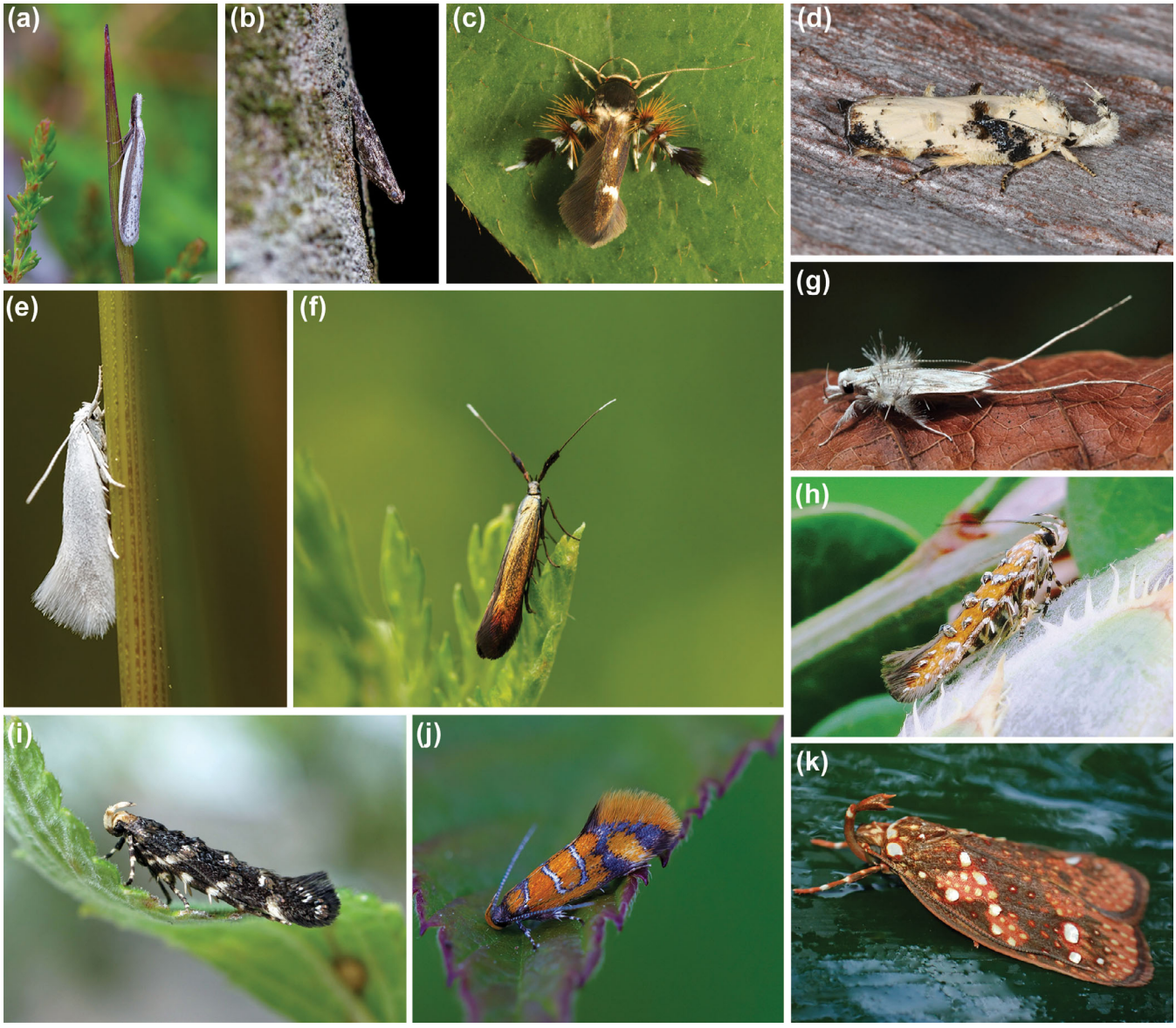

FIGURE 1 A small section of the adult morphological diversity in the superfamily Gelechioidea. (a) Oecophoridae, *Pleurota bicostella*, iNaturalist, photo by Pekka Malinen; (b) Batrachedridae, *Batrachedra praeangusta*, iNaturalist, photo by Pekka Malinen; (c) Stathmopodidae, *Atrijuglans hetaohei*, iNaturalist, pest of walnuts; (d) Depressariidae, *Tonica effractella*, iNaturalist, photo by user "domf"; (e) Elachistidae, *Elachista argentella*, iNaturalist, photo by Pekka Malinen; (f) Coleophoridae, *Coleophora deaureatella*, iNaturalist, photo by Pekka Malinen; (g) Oecophoridae, *Struthoscelis* sp., photo by Kenji Nishida; (h) Cosmopterigidae, *Eteobalea isabellella*, photo by Friedmar Graf; (i) Cosmopterigidae, *Eteobalea serratella*, photo by Friedmar Graf; (j) Oecophoridae, *Promalactis procerella*, iNaturalist, photo by Pekka Malinen; (k) Depressariidae, *Scorpiopsis pyrobola*, iNaturalist, photo by Michelle Colpus. All photographs used with permission or according to the published copyright.

suggested taxonomic revisions to come out of this study, it was the first one that evaluated interfamilial relationships in Gelechioidea using species as terminals. Soon other studies followed. Bucheli and Wenzel (2005) published the first molecular phylogeny that integrated two mitochondrial genes with morphological characters across 42 taxa. Later, Kaila et al. (2011) presented the first extensive molecular phylogenetic study, sequencing 7 genes across 109 gelechioid taxa. They found Elachistidae s.l. to be non-monophyletic as the core Elachistinae were sister to (Coleophoridae + Batrachedridae), Parametriotinae were included in a large clade which was sister to the aforementioned three groups, and the rest of Elachistidae s.l. were in a different place while not being in a monophylum themselves (Figure 2a).

Heikkilä et al. (2014) published an integrated phylogenetic work with 8 genes and 253 morphological characters across 155 taxa, and they proposed 16 monophyletic families for Gelechioidea. Notably, they redefined a monophyletic Depressariidae consisting of Depressariinae and other subfamilies that were previously placed under Elachistidae s.l., as well as additional subfamilies from Lecithoceridae, Oecophoridae and Peleopodidae (Figure 2b). Aiming to test the consensus on Gelechioidea systematics at that time, Sohn et al. (2016) analysed up to 19 genes across 89 taxa, a dataset that was largely independent from Heikkilä et al. (2014) both in terms of sampled taxa and the included genes. Although they did not recover Depressariidae sensu Heikkilä et al. as monophyletic, they proposed no substantial taxonomic changes and suggested that even though some interfamilial relationships were weakly supported in the respective studies, their recurrence in phylogenies with independent taxon sampling hints at their possible taxonomic validity. The most recent molecular phylogenetic work focused on Gelechioidea was from Wang and Li (2020), who analysed 8 genes across 89 taxa and proposed 20 families, among which were Peleopodidae and Ethmiidae (Figure 2d) that they interpreted as distinct from Depressariidae sensu Heikkilä et al. (2014). Overall, Elachistidae and Depressariidae were the families that received the most revisions throughout these studies. Although most clades above family level were weakly supported in these studies, several groupings of families recur with moderate statistical support, especially in the last three studies (Heikkilä et al., 2014; Sohn et al., 2016; Wang & Li, 2020): (1) the GC clade, or the "Gelechiid assemblage" of Sohn et al. (2016) consisting of Gelechiidae and Cosmopterigidae; (2) the BC clade consisting of Batrachedridae and Coleophoridae; (3) the SSB/M clade consisting of Stathmopodidae, Scythrididae and Blastobasidae, sometimes also with Momphidae; and (4) the AXLO clade consisting of Autostichidae, Xyloryctidae, Lecithoceridae and Oecophoridae.

Phylogenomics, the use of genomic-scale molecular data in phylogenetics, has been proven to be quite useful in resolving phylogenetic relationships at different taxonomic levels of the insect tree of life. Such are the relationships among different insect orders (Misof et al., 2014), among superfamilies of Lepidoptera (Bazinet et al., 2013; Kawahara et al., 2019; Mayer et al., 2021; Rota et al., 2022; Twort et al., 2021), among families of different superfamilies in Lepidoptera (Espeland et al., 2018; Li et al., 2024) as well as relationships between subfamilies within focal families (Laurent & Kawahara, 2019; Murillo-Ramos et al., 2023; Õunap et al., 2025). Here, we aim to contribute to the understanding of systematics and phylogenetics of Gelechioidea. To achieve this, we first gathered genomic-scale molecular data from 57 ingroup taxa that span 18 out of the 20 gelechioid families proposed by Wang and Li (2020) and analysed more than a thousand nuclear protein-coding genes in a phylogenomic framework to better resolve the backbone phylogeny. Then, we incorporated most of the available Sanger sequencing data from the previous three molecular phylogenetic studies in order to improve our taxon sampling and to examine intrafamilial relationships.

MATERIALS AND METHODS

Taxon sampling

All the available genome assemblies and raw transcriptomic reads for Gelechioidea (as of September 2022) were considered in this study.

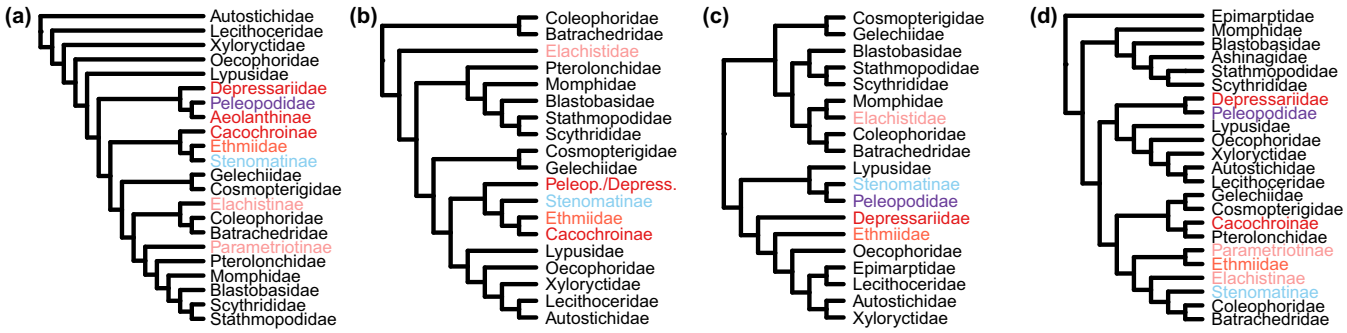

FIGURE 2 A Summary of the family-level topologies from a selection of previous molecular studies that focused on the group as a whole. All family and subfamily names are according to (Wang & Li, 2020) and/or newer studies that changed a subfamily name, such as the name Cacochroinae is from (Kaila et al., 2024). A subfamily is shown distinctively from its family only when it was not part of the rest of the family in one of the studies displayed, or when it is relevant to the discussion in this work. Colours are assigned to families and only the families that were particularly important to the main text of this work were highlighted with colours and colours match the family colours in the rest of the figures throughout this work. (a) Kaila et al. (2011); (b) Heikkilä et al. (2014); (c) Sohn et al. (2016); (d) Wang and Li (2020).

Apart from two samples that had raw genome/exome sequencing data, all of them were used in downstream parts of the analyses. This resulted in 57 ingroup taxa, out of which 42 were raw transcriptomic reads and 15 genome assemblies (Table S1). At its final extent, our dataset included 18 out of 20 families that were proposed for Gelechioidea by Wang and Li (2020), although families Momphidae, Scythrididae, Batrachedridae, Ethmiidae and Epimarptidae were represented by only a single taxon. The 10 outgroup taxa used in this study span six obtectomeran superfamilies: Alucitoidea, Calliduloidea, Papilionoidea, Pterophoroidea, Pyraloidea and Thyridoidea.

Whole-genome sequencing and de novo genome assembly

DNA extracts for the two elachistids *Haplochrois buvati* and *Urodeta hibernella* were provided by Marko Mutanen (University of Oulu). Library preparation and sequencing followed the museomics protocol described in Twort et al. (2021). Generated paired-end libraries were sequenced for 300 cycles (150 cycles each for forward and reverse reads) on the Illumina NovaSeq platform at the Swedish National Genomics Infrastructure (NGI) distributed over two lanes per sample for each sample. The sequencing experiment yielded 33,329,738 reads for *Haplochrois buvati* and 30,847,549 reads for *Urodeta hibernella*.

Raw data for both species was preprocessed as follows. FASTQ files from the two different lanes were merged for each read orientation before further preprocessing. PRINSEQ was used with default parameters to filter out low complexity reads and homopolymer stretches. Then, Trimmomatic (Bolger et al., 2014) with ILLUMINACLIP mode was used to trim sequencing adapters. Finally, trimmed and cleaned reads were used as input to SPAdes (Prjibelski et al., 2020) for de novo genome assembly. Resulting assemblies had N50 of 831 and 2110, and had 508,205 and 276,406 total contigs for *Haplochrois buvati* and *Urodeta hibernella* respectively.

Preprocessing of the public data

Genome assemblies (Boyes & Blaxter, 2024; Boyes & Boyes, 2023; Boyes & Boyes, 2024a; Boyes & Boyes, 2024b; Boyes & Hammond, 2023; Boyes & Holland, 2023a; Boyes & Holland, 2023b; Boyes & Boyes, 2024; Boyes & Hutchinson, 2023; Boyes & Lees, 2023; Boyes & Parkerson, 2022; Boyes & Hutchinson, 2023; Boyes & Kerry, 2023; Holland, 2023; Li et al., 2020; Mead et al., 2021; Poelchau et al., 2015; Stahlke et al., 2023; Tabuloc et al., 2019; Zhang et al., 2022) were downloaded from NCBI (Sayers et al., 2024), using the provided command-line tool datasets (O'Leary et al., 2024). Raw transcriptomic reads for each transcriptome sample (Calla et al., 2017; Kawahara and Breinholt, 2014; Li et al., 2020; Wang et al., 2020; Xu et al., 2021) were downloaded from ENA (O'Cathail et al., 2025). Direct download links were obtained from the accession numbers using ffq v0.0.4 (Gálvez-Merchán et al., 2023). Raw reads were processed as follows. Adapters and potential obvious contamination were removed with cutadapt (Martin, 2011) according to the FastQC (Andrews, 2010) reports; reads that were shorter than 50 nucleotides or those that had average base quality below 30 (below 20) were discarded for Illumina data (for older platforms). Reads that had homopolymer stretches or those that had overall low complexity were discarded using PRINSEQ^{++} (Cantu et al., 2019). TRINITY v2.13.2 was used on the cleaned transcriptomic reads for de novo transcriptome assembly. Resulting assemblies were then clustered into unique sequences that had less than 95% sequence identity with CD-HIT-EST (Li & Godzik, 2006).

The resulting transcriptome assemblies and the downloaded genome assemblies were processed with BUSCO v5.2.2 to search for single-copy orthologous (SCO) genes for Lepidoptera as defined in OrthoDB v10. BUSCO output directories per species were then used as input to an in-house python script to extract the nucleotide and amino-acid sequences for the found SCO genes. Only the genes that were present in at least 70% of the transcriptomes and 90% of the genomes were included for further analysis.

Species identity

The identity of species was checked by extracting the DNA barcode (mitochondrial COI gene) from each assembly, and the sequence was searched in BOLD (Ratnasingham & Hebert, 2007). The species assigned to each assembly was changed only if the DNA barcode matched with a 99% or higher identity to a species that was different from the original identification.

Creating the core phylogenomic dataset

Amino-acid sequences for each gene were used to create multiple-sequence alignments (MSAs) with MAFFT v7.490 (Katoh & Standley, 2013) with the L-INS-i algorithm. Resulting MSAs were back-translated to nucleotide alignments using the Python script "pal2nal.py" (Archive S1). Having false-positive hits for a few sequences across extracted genes in a phylogenomic dataset is common and removing potential outliers and paralogs has been reported to improve the compiled datasets in terms of reduced violation of substitution models used for phylogenetic inference (Yang & Smith, 2014). To account for this, a gene-by-gene sequence exclusion method was implemented as follows. First, IQ-TREE v2.16 (Minh et al., 2020) was used to create individual gene trees for each nucleotide MSA. Then, for each gene tree, the relative length of the longest branch (hereafter referred to as P_{max}) was calculated and genes with $P_{max} \geq 0.1$ were flagged as potential outliers. All gene trees that were flagged as potential outliers were re-evaluated with a recursive tree bisection method at the longest branch of the current tree to arrive at

a "non-outlier" gene tree that still had the 95% of the tips of the starting tree while still satisfying the global 70% taxon coverage threshold. Finally, sequences from the problematic tips were removed from the genes that could be 'saved' with this method and the resulting alignments were retained for downstream analyses. The code to evaluate gene trees with the above-mentioned approach is implemented in the R script "inspect_gene_trees.R" (Archive S1).

Amino-acid alignments for the kept genes after the above-mentioned gene filtering approach were trimmed using TrimAl v1.2rev59 (Capella-Gutierrez et al., 2009) with the provided heuristic method -automated1. The resulting filtered alignments were back-translated as before and were subjected to the gene filtering method as explained above. Resulting nucleotide alignments were concatenated with the Python script "concat-aln" (Archive S1) with the use of the Biopython package (Cock et al., 2009).

Integrating Sanger sequencing samples

The aligned dataset of 24 genes across 324 gelechioid taxa (see Table S2 for voucher and gene information) was obtained from an in-house instance of the voucher sequence database VoSeq (Peña & Malm, 2012). Then the concatenated matrix was split into individual files for each gene. From each gene, the most complete sequence was kept to be used as a reference sequence when searching these genes across the genomic samples. These reference sequences were mapped against one of the gelechiid genome assemblies with available gene annotation (*Pectinophora gossypiella* Saunders) so that they could be put into the correct reading frame. The in-frame reference sequences for all the genes except COI were translated into amino acids and were used in an in-house gene finding pipeline (see Archive S2 Õunap et al., 2025) akin to BUSCO, which uses MetaEuk (Levy Karin et al., 2020) for targeted gene discovery and hmmsearch from the HMMER suite v3.1b2 (Eddy, 2008, 2011) to identify the homologues. COI sequences were searched with tblastn v2.13.0 (Altschul et al., 1990; Camacho et al., 2009) using the invertebrate mitochondrial genetic code (−db_gencode 5). Found genes and the VoSeq genes were re-aligned together as amino acid translations and were back-translated as explained above.

PHYLOGENETIC ANALYSES

Maximum-likelihood inference

Both RAxML-NG (Kozlov et al., 2019) and IQ-TREE v2.16 were used to infer species trees from the concatenated matrix. The resulting concatenated phylogenomic dataset was analysed as both nucleotides and amino acids across five analyses that differed with respect to the molecule type (NT vs. AA), partitioning scheme (X gene vs. X gene X codon), inference software (IQ-TREE vs. RAxML-NG) and treatment of third codon positions for the nucleotide data (NT12 vs. NT123). For the amino acid data, the dataset was partitioned according to genes and the best-fit models were determined using ModelFinder (Kalyaanamoorthy et al., 2017) within IQ-TREE while restricting the model search to only WAG, LG and Q.insect substitution models (−mset WAG, LG, Q.insect). For the remaining analyses (NTXXX), the best-fit models were searched with ModelFinder, while restricting the model search to different rate heterogeneity and invariable sites extensions of the GTR model (−mset GTR and -mrate I + R). To assess branch support, Ultrafast bootstrap (Hoang et al., 2018) and SH-like aLRT (Guindon et al., 2010) methods were used with all the analyses listed here, with 5000 and 1000 replicates, respectively. Further details regarding all independent analyses that were conducted with the phylogenomic data can be found in Table 1. The resulting maximum likelihood (ML) tree from the analysis #3 had the fewest poorly supported nodes and was therefore selected to be further evaluated in analysis #6 by searching through all the included nucleotide substitution models in IQ-TREE and with 10 concurrent independent tree searches.

For the analysis of the combined dataset, the best performing method for the phylogenomic dataset (NT123 X gene X codon) was attempted but showed signs of overparameterisation of the already highly gap-rich data, as several partitions consist of the third codon position for only a single gene. For this reason, the combined data were partitioned manually according to data source (genomic vs. Sanger data) and genetic code (nuclear Sanger genes vs. mitochondrial Sanger genes), which resulted in three partitions. The best substitution model for each partition was searched using ModelTest-NG (Darriba et al., 2020), with the option -T raxml. As for the phylogenetic inference software, RAxML-NG was preferred because of two reasons: (1) the initial attempt with IQ-TREE showed poor ultrafast bootstrapping performance, with speed that was not significantly faster than standard Felsenstein bootstraps, and (2) RAxML-NG uses subtree pruning and regrafting (SPR) as the tree search heuristic, which is proven to perform better at exploring the tree space, especially with taxon-rich datasets (see the benchmarks in supplementary material of Kozlov et al., 2019). To save time in computations, the suboptimal IQ-TREE ML topology was used as the starting tree for the RAxML-NG tree search. Later reanalyses of the data without the custom starting tree did not yield a better topology in terms of likelihood. The maximum likelihood analysis was initially constrained with the backbone tree from the phylogenomic analyses, but further analyses without constraints showed that this was unnecessary, as the obtained backbone topology did not change. The resulting support values from 100 standard bootstraps were mapped onto the ML tree that was inferred without a topological constraint after observing that constraining the tree search did not affect the resulting backbone topology.

ASTRAL inference

Best-fitting models for each gene alignment were searched by running IQ-TREE with the ModelFinder option (-m MF, suppressing

TABLE 1 Summary of the six independent phylogenetic analyses conducted with the core phylogenomic dataset.

	ML software	Data (#columns)	Partitioning (# partitions)	Model/partitioning scheme search	Model set	Rate heterogeneity
1	IQ-TREE	NT123 (2,046,984)	By gene (1767)	ModelFinder	-mset GTR	-mrate I + R
2	RaxML-NG	NT123 (2,046,984)	By gene (1767)	ModelTest-NG	-T raxml	Default (+I and + G)
3	IQ-TREE	NT123 (2,046,984)	By gene and codon pos. (408)	ModelFinder (w/partition merging)	-mset GTR	-mrate I + R
4	IQ-TREE	NT12 (1,364,656)	By gene (1767)	ModelFinder	-mset GTR	-mrate I + R
5	IQ-TREE	AA (682,328)	By gene (1767)	ModelFinder	-mset WAG, LG, Q. insect	Default (all possible)
(Main) 6	IQ-TREE	NT123 (2,046,984)	By gene and codon pos. (408)	ModelFinder (w/partition merging)	Default (all possible)	Default (all possible)

the subsequent tree inference step); then the found models were used in 40 independent IQ-TREE runs for phylogenetic inference with 1000 UltraFast bootstraps (Hoang et al., 2018) and the -bnni options enabled.

Best-scoring trees out of 40 independent runs were selected by re-evaluating the likelihoods using RAxML-NG (--evaluate mode, which re-optimises both the model parameters and the branch lengths) and used as input for wASTRAL-h (Zhang & Mirarab, 2022), which is a new implementation of ASTRAL (Zhang et al., 2018) that can weight quartets by using the branch lengths and branch support values among the input gene trees.

Concordance factors

Gene and site concordance factors (gCF and sCF) were calculated on both the main ML topology and the ASTRAL topology using IQ-TREE v2.4.0 in order to benefit from the newly implemented --scfl option, which uses likelihood distributions when sampling quartet states instead of relying on parsimony. gCF for a branch could be summarised as the proportion of gene trees (among the trees that have the relevant branches present) that are in agreement with the focal branch, while sCF calculate an estimate of the proportion of sites that support the particular split caused by the branch in question among all sites that are present in the concatenated alignment by effectively sampling random quartet states around the focal branch, regardless of the distribution of said sites among the genes. Quartet concordance factors were calculated using ASTRAL-III (Zhang et al., 2018) by fixing the input species tree topology (-q option) and just calculating the full set of support metrics available (-t 2 setting). Resulting concordance (and discordance) factors were collected and visualised using the concordance vector approach described in Lanfear and Hahn (2024) to calculate confidence intervals around the point estimates and to better investigate the discordant cases.

Topology tests

To test phylogenetic hypotheses regarding some of the interesting groupings for both our phylogenomic and denser trees, we conducted tree topology tests (Kishino et al., 1990; Kishino & Hasegawa, 1989; Shimodaira, 2002; Shimodaira & Hasegawa, 1999; Strimmer & Rambaut, 2002) that are implemented in IQ-TREE (Minh et al., 2020). For the phylogenomic tree, the resulting topologies from the six independent analyses were tested with the best nucleotide models that resulted in the presented topology (Figure 4). For the denser tree result, seven alternative placements (Figure 3b–h) for the taxa that belong to Pterolonchidae sensu Heikkilä et al. (2014) were tested against the maximum-likelihood topology (Figure 3a, Figure S1) to assess confidence in the (non)monophyly of the family and in the placement of *Coelopoeta glutinosi* as sister to the rest of Gelechioidea, with the molecular data analysed here.

Visualisation

All phylogenetic trees were visualised using R (R Core Team, 2022), with the relevant functions that were provided by the R packages ape (Paradis & Schliep, 2019) and ggtree (Yu et al., 2017). All manual modifications of the figures were done with Inkscape v1.3.2.

RESULTS AND DISCUSSION

General topology, support values and concordance factors

We present a maximum-likelihood phylogenomic tree from 1767 genes across a total of 67 taxa (Figure 4 and Table S1), which, for the first time, enables the examination of interfamilial relationships in Gelechioidea with high amounts of molecular data. The tree has maximal branch support throughout except for one node (see the

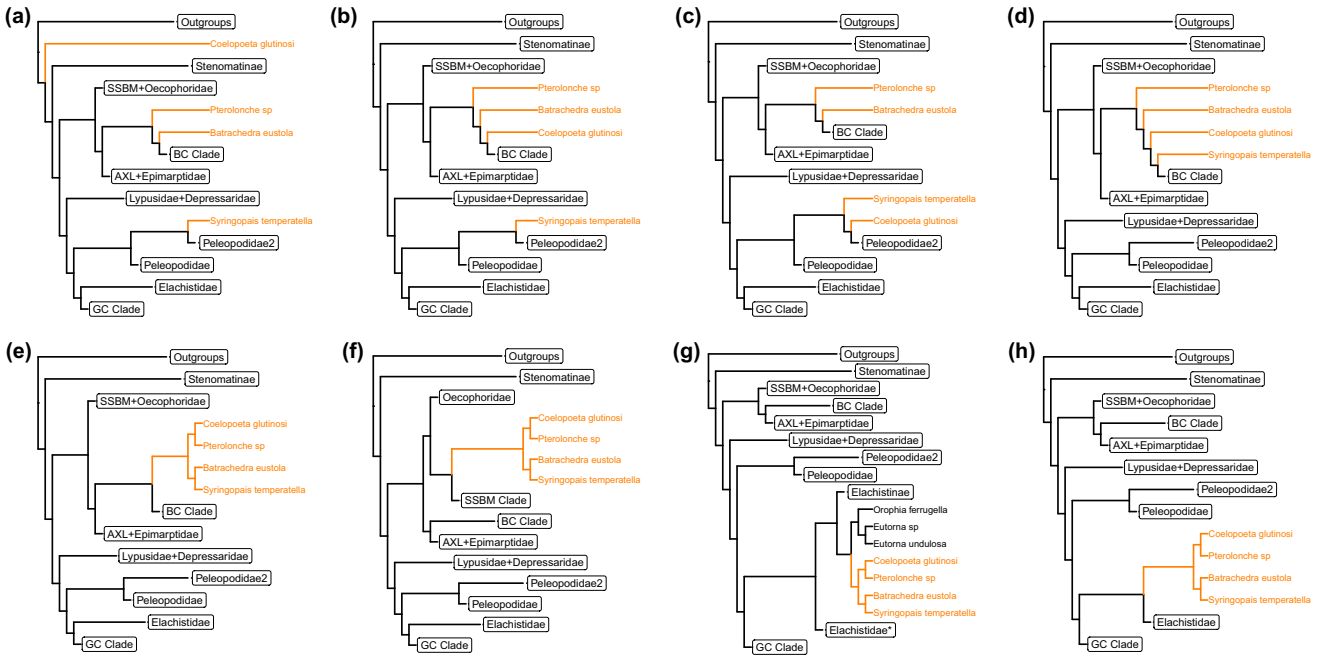

FIGURE 3 Summary cladograms of the eight alternative topologies used for the topology tests conducted using IQ-TREE regarding the monophyly/placement of the four included pterolonchid taxa. (a) Original topology; (b) *Coelopoeta glutinosi* is in the partial pterolonchid grade and is sister to the BC clade; (c) *Coelopoeta glutinosi* branches after *Syringopais temperatella*, the one pterolonchid that was not in the pterolonchid grade, and is sister to Peleopodidae; (d) Pterolonchids as a grade, forming a clade together with the BC clade; (e) All pterolonchids monophyletic and sister to the BC clade; (f) All pterolonchids monophyletic and sister to the SSBM clade, placed according to Heikkilä et al. (2014); (g) All pterolonchids monophyletic and sister to (*Orophia* and *Eutorna*), placed according to Wang and Li (2020); (h) All pterolonchids monophyletic and sister to the clade that includes Elachistidae.

labelled node in Figure 4 that has SH-aLRT value of 77.9/100 and UFBS value of 100/100). When the tree is rooted with respect to the included outgroups, overall four main lineages appear in the backbone of the tree: (1) Stenomatinae (Depressariidae); (2) A clade with Gelechiidae, Cosmopterigidae, Elachistidae, Ethmiidae, Lypusidae, Depressariidae and Peleopodidae; (3) A clade with Blastobasidae, Stathmopodidae, Scythrididae, Momphidae and Oecophoridae; 4) A clade with Autostichidae, Xyloryctidae, Batrachedridae, Coleophoridae, Epimarptidae and Lecithoceridae. The topology could be summarised as (Outgroups,(1,(2, (3,4)))). While we were mostly able to recover these four lineages regardless of the ML software, molecule type and different partitioning strategies across our six different analyses (Table 1), there were differences in the more recent splits across our different analyses (Figures S2–S7). Nevertheless, the results of the analysis #6, which yielded the maximum-likelihood tree with the fewest poorly supported nodes (only one node), and which we present here as the preferred topology (Figure 4), provide insights into those recent splits. It is also worth noting that despite the significant support across almost all nodes, most deep splits have relatively short branches.

Concordance factors were highly variable throughout the tree, ranging 0.57–99.6, 30.29–99.87, 27.94–91.27 for the gCF, qCF and sCF, respectively (Figure S8). Excluding five families with a sole representative, the average CF values for branches that lead to the 13 family-level clades were 52.9 for gCF, 45.1 for sCF and 72.7 for qCF while the values themselves ranged 2.05–99.6 (gCF), 33.23–91.27 (sCF) and 33.47–99.85 (qCF). Gene discordance factors due to non-monophyly of clades around a focal branch across the discordant gene trees (referred to as gDFP values by IQ-TREE) were generally high and displayed a left-skewed distribution with the mean of 58.96 and the median of 70.98. Moreover, in most cases when the gCF values were low, gDFP values were the highest value among the other values in the concordance vectors. This may indicate a particularly high prevalence of gene tree estimation error within our dataset.

Another interesting metric on our gene alignment that can shed light on this phenomenon is the difficulty score (hereafter referred to as *DS*) reported by Pythia (Haag et al., 2022), which is a machine-learning estimator that uses inherent properties of the alignment such as sites/taxa ratio, percentage of missing data and topological distance among parsimony trees inferred from the alignment to report an estimated difficulty score between 0 and 1 (1 being the most difficult). It has been shown before through the analysis of more than 19,000 gene alignments across 15 empirical datasets that the number of independent tree inferences can severely affect the accuracy of gene-tree inference, and it has been found that the *DS* of an alignment predicted by Pythia in particular, can be largely predictive regarding the number of tree searches required to find the maximum-likelihood tree. When the difficulties of alignments are divided into

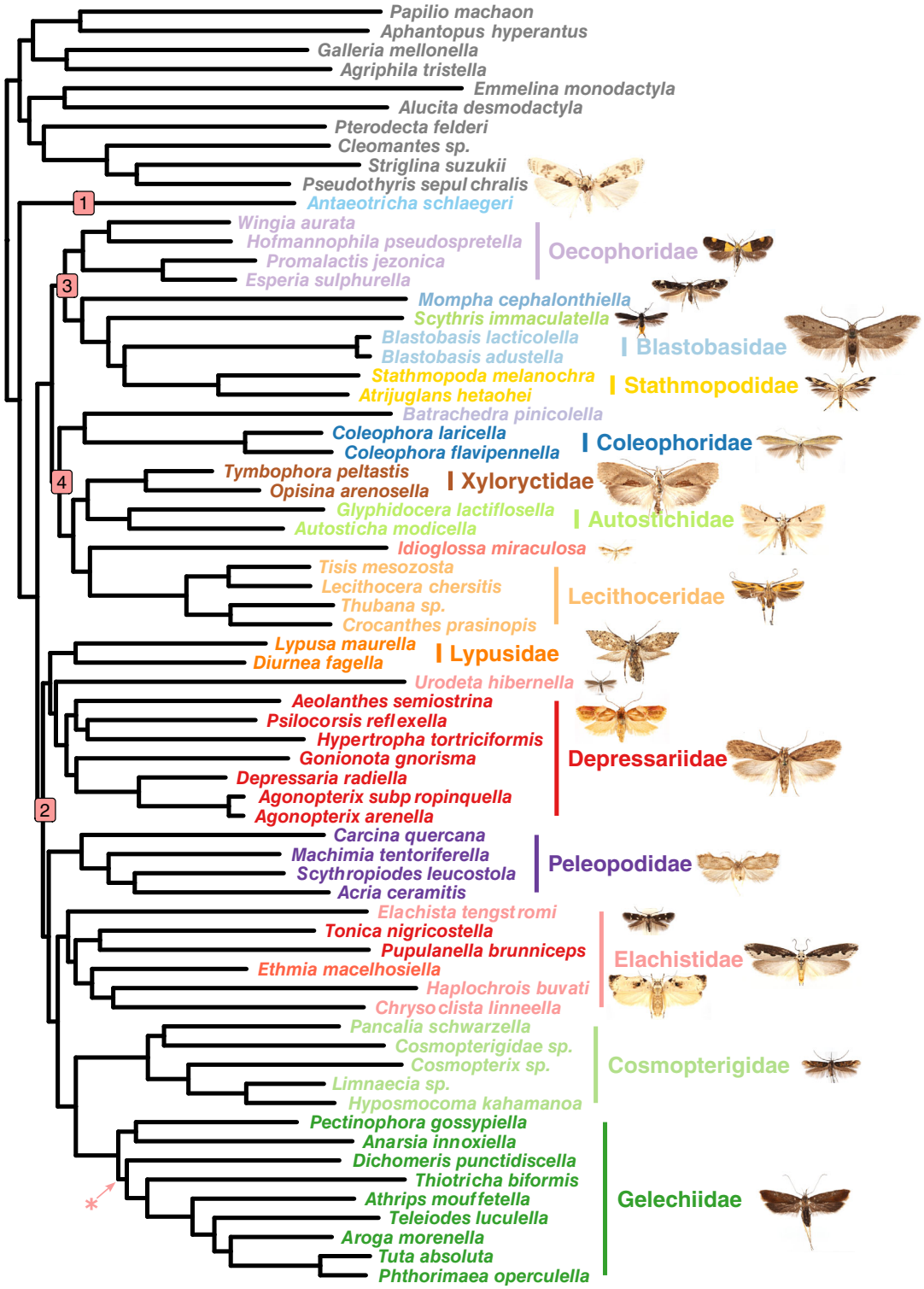

FIGURE 4 Maximum-likelihood tree from the NT123 analysis. All the nodes except the one that is labelled with an asterisk have *SH − aLRT* ≥ 80 and *UFBS* ≥ 95. Tip labels are coloured according to the assigned family names prior to this study. From top to bottom, images represent *Antaeotricha schlaegeri*, *Oecophora bractella*, *Mompha conturbatella*, *Scythris sinensis*, *Hypatopa binotella*, *Stathmopoda melanochra*, *Coleophora ornatipennella*, *Tymbophora peltastis*, *Stibaromacha ratella*, *Idioglossa polliacola*, *Tisis sp*, *Diurnea fagella*, *Urodeta hibernella*, *Aeolanthes* sp., *Depressaria radiella*, *Acria emarginella*, *Elachista quadripunctella*, *Ethmia bipunctella*, *Tonica effractella*, *Pancalia schwarzella*, *Dichomeris ustalella*.

three categories as easy (*DS* < 0.3), intermediate (0.3 ≤ *DS* ≤ 0.7) and hard (*DS* > 0.7) alignments, only 34.09% of the alignments with intermediate difficulty would yield the best known ML topology reliably (at least 90% of the time, among 100 tries), whereas this percentage was 65.49 for the easy and 0.42 for the hard alignments (Liu et al., 2024). Among the 1767 genes that comprise our dataset,

19.75% were easy, 68.25% were intermediate and 12% were hard alignments, according to the same Pythia score qualitative categorisation criteria. This demonstrates a prevalent problem in phylogenomic datasets: the gene-tree estimation error is inevitable to a large degree with the current standard methods of maximum-likelihood tree inference for the gene trees, and it needs to be accounted for when the results of analyses that rely on inferred gene trees are being discussed. The same study (Liu et al., 2024), however underlines that the effect of using possibly suboptimal ML gene trees as input for tree-reconciliation methods might not be as detrimental. They show that as the number of replicates for gene-tree inference increases, false-positive branches would get increasingly lower local posterior probability support.

The resulting topology from the ASTRAL analysis (Figure S9) differed from the one presented in Figure 4. However, for the majority of disagreements between the two topologies, movement of a single branch several nodes away in a rootwards direction could explain the configuration observed in the ASTRAL tree. Some examples of such cases are: (1) *Elachista tengstromi* (Elachistidae, Elachistinae) jumped to a sister position to Gelechiidae+Cosmopterigidae instead of being sister to a clade that also includes Parametriotinae (Elachistidae), (2) *Urodeta hibernella* resolved as sister to Lypusidae + core Depressariidae, instead of being sister to the latter, (3) *Idioglossa miraculosa* (Epimarptidae) is sister to Autostichidae + Xyloryctidae + Lecithoceridae, instead of being sister to Lecithoceridae, (4) *Mompha cephalonthiella* is sister to SSBO instead of being sister to SSB, (5) *Ethmia macelhosiella* (Ethmiidae) is sister to *Chrysoclista linneella* instead of being sister to *Chrysoclista linneella* + *Haplochrois buvati*, which is a clade present in the ML tree of two genera under the same subfamily Parametriotinae (Elachistidae), which was recovered as monophyletic in all three previous molecular works that sampled this subfamily (Figure 2). A detailed exploration of the example (1) mentioned above is given in Figure S10. As exemplified by the Ψ_4 values for the gene concordance on the heatmaps, this dataset suffers from a severe case of gene-tree estimation error. In this specific case, the grouping inferred by ASTRAL has decreased gCF and sCF values compared to the ML-inferred grouping whereas the qCF seems to be increased in the ASTRAL arrangement when compared to the ML arrangement. The opposite of this pattern—ASTRAL arrangement increasing gCF—is also observed among the other examples mentioned above. In theory, qCF would be able mitigate low levels gene-tree estimation error where the gDFP values would be slightly elevated (Lanfear & Hahn, 2024), but when the gene-tree estimation errors are particularly high, values in the sCF and quartet concordance vectors (CF, DF1, DF2) will be biased to be arbitrarily higher than the true biological parameter they are estimating since the fourth value is by definition zero for these factors. Another concern here is the uneven sampling, especially for branches for which one of the four groups around it is a sole representative of a family in our tree. In a case of more even sampling, qCF and gCF would be able to sample quartets where different representatives of the undersampled family to test the concordance.

Overall, the discordance between the ASTRAL and ML topologies is most probably resulting from noise in our gene-tree topologies rather than a meaningful biological signal, which we would not be able to detect or rule out due to the high level of gene-tree estimation error in our dataset. The fact that this holds even with 40 independent tree searches for each gene alignment along with the observation of short branches in the backbone of the ML tree remains consistent with the previously suggested hypotheses of rapid radiations during the early divergence of Gelechioidea (Kaila et al., 2011).

Position of the root

When we complemented this phylogenomic dataset with Sanger sequencing data from the previous three molecular phylogenetic studies and inferred a maximum-likelihood phylogenetic tree with this new dataset with denser taxon sampling (hereafter referred to as denser tree), we see that most families and their positions in the backbone are stable (Figure S1, see also the discussion below regarding Pterolonchidae). We were able to recover the same family-level backbone topology with the denser taxon sampling dataset both with and without using the phylogenomic tree's family-level backbone topology as the topological constraint for the denser tree phylogenetic inference.

Throughout the earlier phylogenetic works focused on Gelechioidea, the recovered position of the root varied (Figure 2). Here we find with very good branch support in our phylogenomic tree that the depressariid subfamily Stenomatinae is sister to the rest of Gelechioidea (Figures 4 and 5, see also the discussion below regarding Pterolonchidae). A peculiar pupal character, so-called lateral condyles in abdominal segments preventing lateral movement, was suggested by Minet (1990) to be a synapomorphy for an expanded concept of Elachistidae. This trait is also present in Stenomatinae. The finding that Stenomatinae is the sister to the rest of Gelechioidea would suggest that this character is actually a synapomorphy, yet repeatedly reversed, for the Gelechioidea.

Depressariidae, Lypusidae, Peleopodidae, Elachistidae, Gelechiidae and Cosmopterigidae

After Stenomatinae, the tree splits into two main lineages: Lineage #2 and the clade including Lineages #3 and #4 as denoted above. The first split in Lineage #2 separates most of Depressariidae and Lypusidae—which was previously placed as sister to either the "AXLO" clade (Heikkilä et al., 2014; Wang & Li, 2020), or (Peleopodidae + Stenomatinae) (Sohn et al., 2016)—from the rest. Wang and Li (2020) regarded Depressariidae sensu Heikkilä et al. (2014) to be better split into two families and they redefined Depressariidae to only include subfamilies Depressariinae, Hypertrophinae, Hypercalliinae and several additional genera while they redefined a separate Peleopodidae to include Peleopodinae and

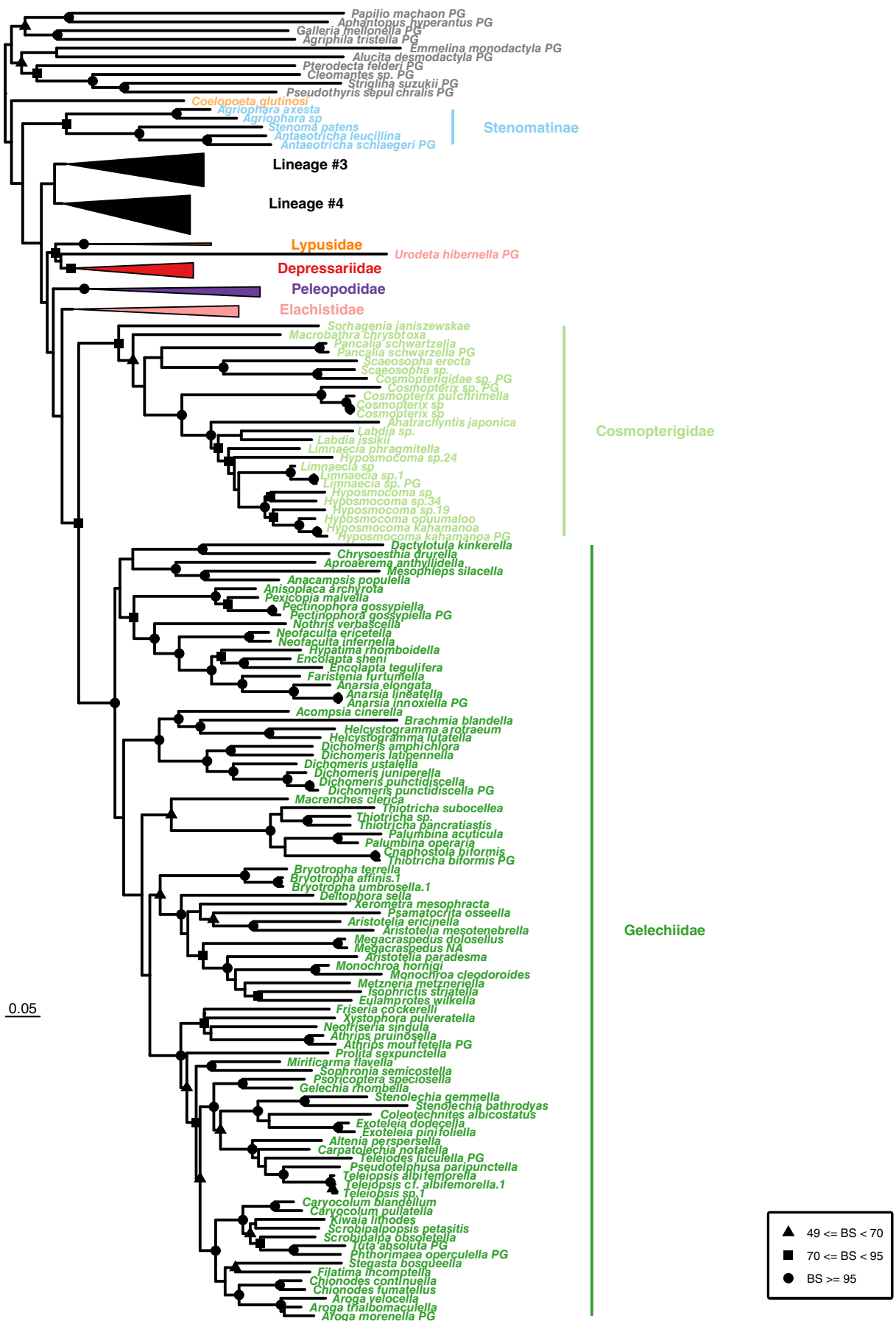

FIGURE 5 Legend on next page.

Acriinae. As opposed to what was observed by Wang and Li, we do not recover Peleopodidae and core Depressariidae as sister groups. Instead, Peleopodidae are found here as sister to the rest of the second clade within Lineage #2, a clade that consists of Gelechiidae, Cosmopterigidae, Elachistidae, Ethmiidae and two sister depressariid genera *Tonica* and *Pupulanella* (Adamski & Nishida, 2021). It is worth noting here that although the position of Peleopodidae in our phylogenomic tree is maximally supported, in the amino-acid tree, this family is recovered as sister to Lypusidae (Figure S7) with high support ($UFBS \geq 95$, $SH-aLRT \geq 80$) although the sister relationship between (Lypusidae + Peleopodidae) and core Depressariidae is not as well supported ($UFBS = 67$). This suggests that the position of Peleopodidae within the lineage #2 in our tree is not yet resolved. The observation that nucleotide and amino-acid trees can yield conflicting topologies or an apparent discrepancy in support values is a known phenomenon in phylogenomics and has been explored methodically for Lepidoptera (Rota et al., 2022), a dipteran family (Gillung et al., 2018), and for arthropods in general (Zwick et al., 2012). It has been suggested that this phenomenon most likely stems from so-called systematic errors, a term that describes the situation in which the assumptions of the models of molecular evolution used are violated in a biased manner.

Heikkilä et al. (2014) demonstrated that Elachistidae sensu lato as defined by Kaila et al. (2011) is polyphyletic, and the name should describe the clade comprising subfamilies Elachistinae, Parametriotinae and Agonoxeninae. However, Wang and Li (2020) reported Elachistinae + Parametriotinae —previously suggested as core groups of Elachistidae—to be non-monophyletic; therefore, leaving Elachistidae paraphyletic. We observe a third pattern here, in which one should include *Tonica*, *Pupulanella* and Ethmiidae to have a monophyletic Elachistidae that includes both Parametriotinae and most of Elachistinae. A single elachistid taxon, *Urodeta hibernella* Staudinger, however, remains separated from the rest of Elachistidae sensu Wang and Li (2020) and instead is sister to the core Depressariidae. Despite the apparent discrepancy between the position of *Urodeta hibernella* here and in the main results of Heikkilä et al. (2014), in the DNA-only ML topology presented there, this taxon is recovered as sister to *Aeolanthes* sp. and within a clade of other depressariid taxa. The fact that *Urodeta hibernella* is recovered far away from the rest of Elachistidae despite the available genomic data across more than a million characters hints at the validity of this placement near the core Depressariidae. The phylogenomic topology for this clade [(Elachistinae, ((Tonica, Pupulanella), (Parametriotinae, Ethmiidae)))] is corroborated by our denser tree with the addition of Cacochroinae (Kaila et al., 2024), (sampled only in the denser tree), as sister to Elachistinae (Figure 6).

Remaining families

As already explained above, sister to the Lineage #2 is a clade in which Lineages #3 and #4 are sisters. Within Lineage #3, we recover Oecophoridae as sister to the "SSBM" clade that consists of Stathmopodidae, Blastobasidae, Scythrididae and Momphidae. Oecophoridae were previously found to be sister to the "AXL" clade consisting of Autostichidae, Xyloryctidae and Lecithoceridae by Heikkilä et al. (2014) and Wang and Li (2020), although with low support. The rest of the Lineage #3, the "SSBM" clade, is a group of families that were recovered together before by Kaila et al. (2011), Heikkilä et al. (2014), Sohn et al. (2016) and Wang and Li (2020), sometimes with a different order of splits within the group than the written order of the abbreviation. We recover the SSBM clade in our denser tree too (Figure 7). However, there is an additional family that was represented with only Sanger-sequenced samples, Ashinagidae sensu Wang and Li (2020). Despite the lack of genomic data for Ashinagidae, it is recovered as monophyletic and its position within the SSBM clade is consistent with the topology in Wang and Li (2020). In Lineage #4, a clade consisting of Batrachedridae and Coleophoridae is sister to the "AXL" clade. Batrachedridae and Coleophoridae together as a clade were found as either sister to the rest of Gelechioidea (Heikkilä et al., 2014) (Figure 2b), sister to Elachistinae (Wang & Li, 2020) (Figure 2d) or sister to (Elachistinae + Momphidae) (Sohn et al., 2016) (Figure 2c). Interestingly, we see that *Idioglossa*, the only sampled epimarptid in the phylogenomic tree, is positioned within the "AXL" clade, as sister to Lecithoceridae. Sohn et al. (2016) recovered *Idioglossa* sister to Lecithoceridae also, while the two epimarptids, including *Idioglossa*, in Wang and Li (2020) were recovered as sister to the rest of Gelechioidea. The finding that Epimarptidae is sister to Lecithoceridae is corroborated, albeit with low bootstrap support, in our denser tree that includes another epimarptid, *Epimarptis hiranoi* Sugisima (Figure 8). The nucleotide versus amino-acid discrepancy highlighted above is pronounced also here regarding the relative positions of Batrachedridae, Coleophoridae and *Idioglossa* (Figure S6). It is clear that more methodical exploration of this phenomenon is required in future work.

There is a family apart from the ones mentioned above that is scattered across our denser tree, Pterolonchidae sensu Heikkilä et al. (2014), which was sampled only with four Sanger-sequenced taxa. One pterolonchid is recovered as sister to (Stenomatinae + Gelechioidea) (Figure 5), two as the first two successive splits in the clade that also includes (Batrachedridae + Coleophoridae) (Figure 8), and the remaining one within Peleopodidae (Figure 6). Although it brings concern regarding the validity of the group, this is a pattern that is hard to comment on with high confidence since we do not have genomic samples from this family. Indeed, we see that the presented placement here is not the only supported topology when we

FIGURE 5 Gelechiidae and Cosmopterigidae within Lineage #2 (Figure 4) in the denser tree with Sanger sequencing samples integrated. Shapes at the nodes show the standard bootstrap support. Tips representing the phylogenomic samples are marked with a "PG" suffix to distinguish them from the Sanger-sequenced samples.

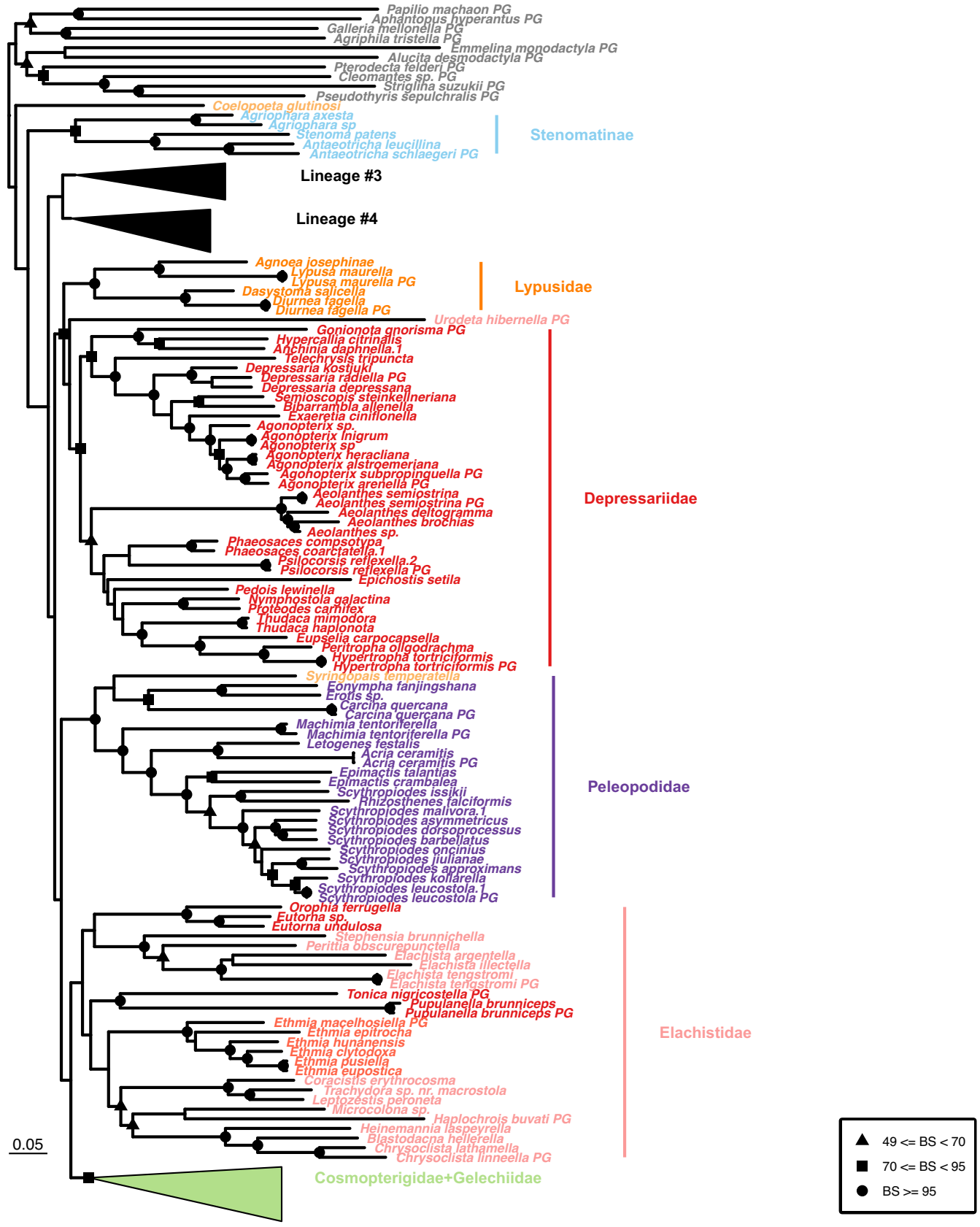

FIGURE 6 Remaining families within Lineage #2 (Figure 4) in the denser tree with Sanger sequencing samples integrated. Shapes at the nodes show the standard bootstrap support. Tips representing the phylogenomic samples are marked with a "PG" suffix to distinguish them from the Sanger-sequenced samples.

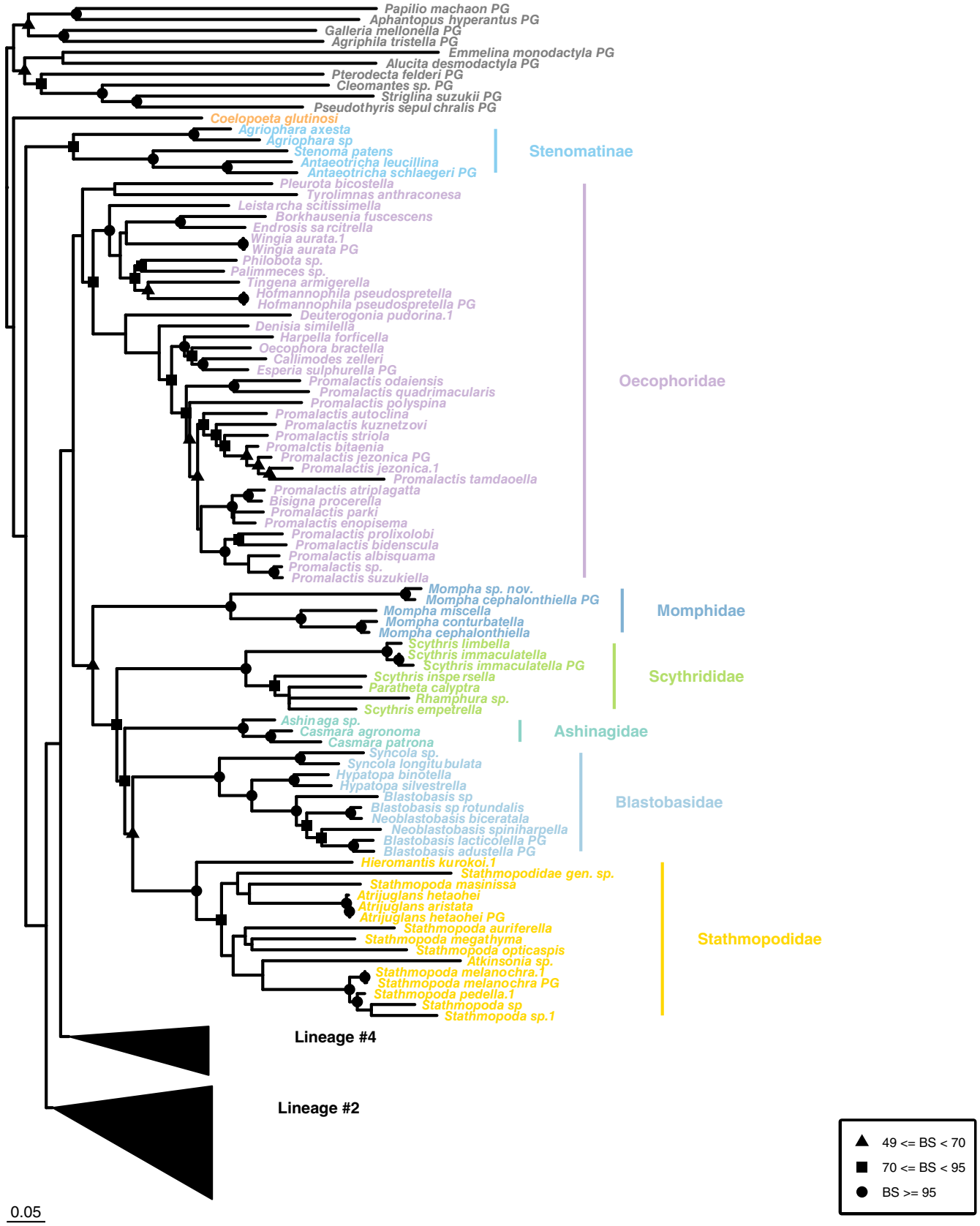

FIGURE 7 Lineage #3 (Figure 4) in the denser tree with Sanger sequencing samples integrated. Shapes at the nodes show the standard bootstrap support. Tips representing the phylogenomic samples are marked with a "PG" suffix to distinguish them from the Sanger-sequenced samples.

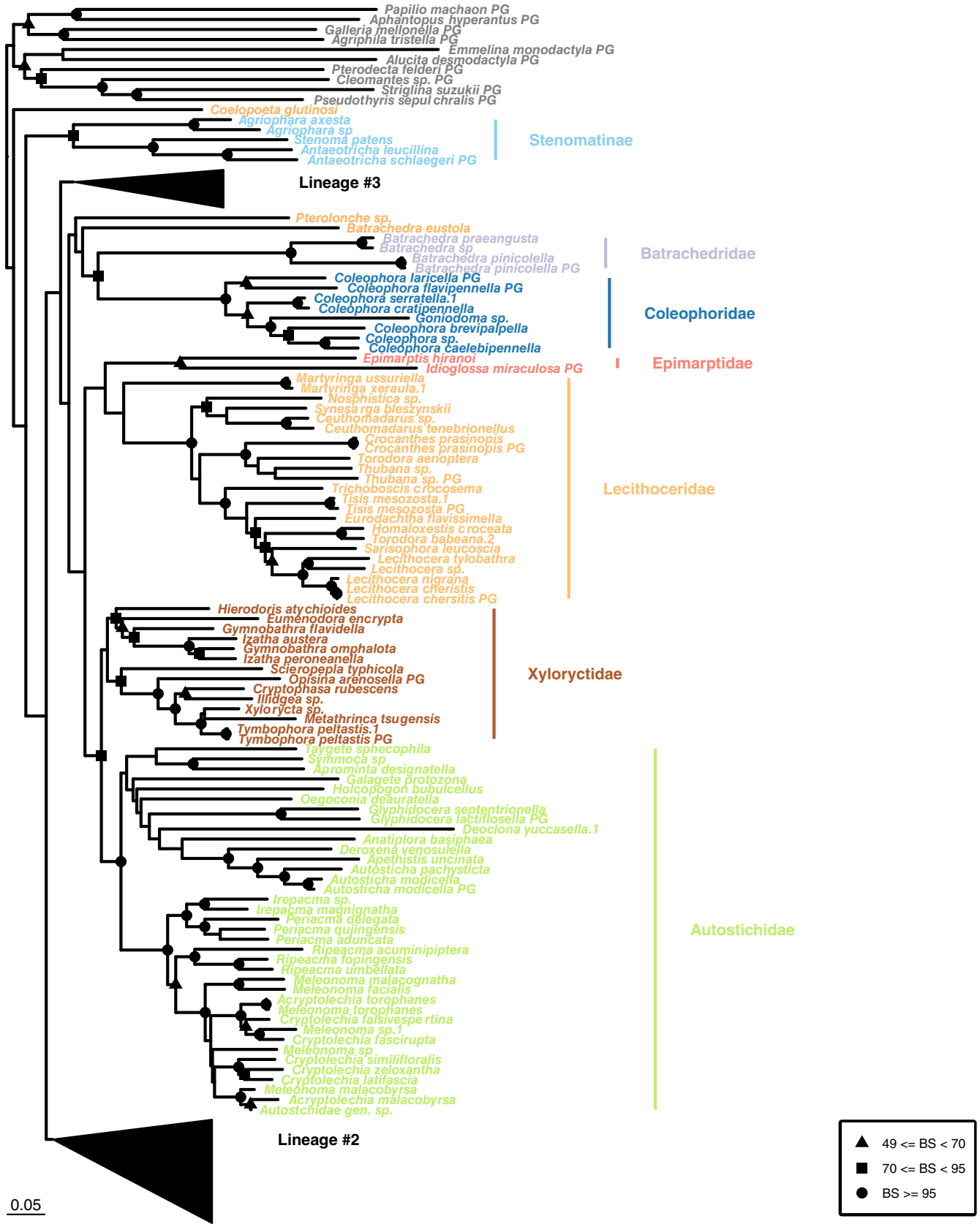

FIGURE 8 Lineage #4 (Figure 4) placement in the denser tree with Sanger sequencing samples integrated. Shapes at the nodes show the standard bootstrap support. Tips representing the phylogenomic samples are marked with a "PG" suffix to distinguish them from the Sanger-sequenced samples.

TABLE 2 Results from topology tests. Summary cladograms of the alternative topologies can be found in Figure 3.

Tree	deltaL	Bp-RELL	p-KH	p-SH	p-WKH	p-WSH	c-ELW	p-AU
1	0	0.369	0.644	1	0.605	0.927	0.367	0.66
2	15.193	**0.012**	0.066	0.485	0.066	0.241	**0.0127**	0.0712
3	5.0739	0.232	0.356	0.713	0.356	0.681	0.232	0.451
4	37.136	**0**	**0.0099**	0.071	**0.0041**	**0.0215**	**4.34e-07**	**0.000802**
5	7.7636	0.102	0.339	0.767	0.339	0.835	0.102	0.391
6	5.7742	0.242	0.395	0.736	0.395	0.769	0.242	0.517
7	32.227	**0.004**	0.0818	0.16	0.0578	0.184	**0.00457**	**0.0278**
8	18.093	0.04	0.192	0.399	0.192	0.561	0.0399	0.183

Note: All tests performed 10,000 resamplings using the RELL method. deltaL: LogL difference from the maximal logL in the set. bp-RELL: Bootstrap proportion using the RELL method (Kishino et al., 1990). p-KH: *p*-value of one-sided Kishino and Hasegawa (1989). p-SH: *p*-value of Shimodaira and Hasegawa (1999). p-WKH: *p*-value of weighted KH test. p-WSH: *p*-value of weighted SH test. c-ELW: Expected Likelihood Weight (Strimmer & Rambaut, 2002). p-AU: *p*-value of approximately unbiased (AU) test (Shimodaira, 2002). Significant values (rejected topologies by a given test) are shown in boldface.

conducted topology tests with alternative topologies regarding the placement of the included pterolonchid taxa. For instance, placing all four pterolonchid taxa with the same ingroup topology and the placement as they are recovered in Heikkilä et al. (2014) (Figure 3h) does not appear to be a significantly worse topology than the presented topology (Table 2 Tree #6). Regardless of the weak molecular evidence for the validity of the group, the presence of nearly or entirely immobile male valvae gives support for the monophyly of a part of the Pterolonchidae as described in Heikkilä et al. (2014).

REVISED CLASSIFICATION

Elachistidae sensu nov.

As illustrated above, our results indicate that Elachistidae is not monophyletic. We transfer Cacochroinae (Depressariidae), Ethmiinae **stat. nov**. and genera *Pupulanella* and *Tonica* to Elachistidae sensu *nov*. A unique synapomorphy of the revised concept of Elachistidae is the arrangement of larval dorsal setae D1 and D2 in abdominal segment A9: within Elachistidae sensu *nov*. the dorsal seta D1 is mesially placed as compared to D2.

Urodeta hibernella, however, is excluded from this revised Elachistidae as it is not recovered together with the rest of Elachistidae sensu *nov*. in our molecular analyses and treated as unplaced to family.

Stenomatidae stat. nov.

As shown in both the phylogenomic and the 'denser' trees, Stenomatinae is recovered as sister to the rest of Gelechioidea, away from other Depressariidae. We raise it to Stenomatidae **stat. nov**. to restore the monophyly of the core Depressariidae. Lack of secondary SV setae in larval anal legs distinguishes Stenomatidae **stat. nov**. from the rest of Depressariidae sensu *nov*. and other associated lineages such as Ethmiinae **stat. nov**.

Depressariidae sensu nov.

As a result of the revisions listed above, the extent of Depressariidae has now changed and includes the following subfamilies: (i) Depressariinae (without *Pupulanella* and *Tonica*), (ii) Aeolanthinae, (iii) Hypercalliinae and (iv) Hypertrophinae.

CONCLUSIONS

The hyperdiverse Gelechioidea has long been intractable for evolutionary studies due to the poor resolution of family-level relationships. Utilising more than 2 million base pairs of molecular data across 57 ingroup taxa, we now have a much-improved backbone phylogenetic hypothesis for Gelechioidea as compared to previously suggested alternatives, with very high branch support. Although our phylogenomic results based on nucleotide data provide further insights into Gelechioidea phylogenetic systematics, there are several issues that became apparent when the results from different analyses on the phylogenomic core dataset and from the denser nucleotide tree based on a much greater taxon sampling (but including up to 24 genes from the Sanger-sequenced samples) are critically compared. These issues need further inspection in future works. Such are the not-so-stable positions of Peleopodidae, Batrachedridae, Coleophoridae and *Idioglossa*; and the validity of Pterolonchidae in its current delimitation in light of the (lack of) available molecular evidence.

AUTHOR CONTRIBUTIONS

Etka Yapar: contributed to formal analysis, writing – original draft, writing – review and editing, and visualisation. **Andrea Chiocchio:** contributed to writing – review and editing and data curation.

Maria Heikkilä: contributed to writing – review and editing and data curation. **Jadranka Rota:** contributed to funding acquisition, writing – review and editing. **Lauri Kaila:** contributed to writing – review and editing, supervision, conceptualisation. **Niklas Wahlberg:** contributed to funding acquisition, supervision, conceptualisation, writing – review and editing.

ACKNOWLEDGEMENTS

A considerable part of the public transcriptome samples included in this study was generated thanks to a United States National Science Foundation grant to the late Charles Mitter and to Michael Cummings (University of Maryland). We thank Adam Bazinet (Battelle National Biodefense Institute, LLC) and Kim Mitter (University of Maryland) for their laboratory and informatic work in generating and submitting the data to Genbank. We thank Marko Mutanen (University of Oulu) for providing the DNA extracts of *Haplochrois buvati* and *Urodeta hibernella* samples. We also thank the Darwin Tree of Life project for making the high-quality Lepidoptera genomes available. We were able to complete the computations regarding phylogenetic inference significantly faster thanks to resources provided by LUNARC, The Centre for Scientific and Technical Computing at Lund University. We thank Fabien Condamine and three anonymous reviewers for constructive comments on the initially submitted manuscript.

FUNDING INFORMATION

E.Y. was funded by ELLIIT with the project: "Call-C02-Developing core–technologies for tree-based models" to N.W.; A.C. was funded by a Crafoord Foundation (Crafoordska Stiftelsen, Sweden) grant awarded to J.R. (grant no. 20180794).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study (Yapar, 2025) are openly available in Zenodo at https://zenodo.org/records/16173549. Raw sequencing data generated for this study (Lund University & LU SBG, 2025) are deposited on ENA under the project accession PRJEB87822 https://www.ebi.ac.uk/ena/browser/view/PRJEB87822.

ORCID

Etka Yapar https://orcid.org/0000-0001-7938-1325
Andrea Chiocchio https://orcid.org/0000-0002-0067-7025
Maria Heikkilä https://orcid.org/0000-0002-9048-4381
Jadranka Rota https://orcid.org/0000-0003-0220-3920
Lauri Kaila https://orcid.org/0000-0003-0277-1872
Niklas Wahlberg https://orcid.org/0000-0002-1259-3363

REFERENCES

Adamski, D. & Nishida, K. (2021) A new genus *Pupulanella* and a new species, Adamski and Nishida (lepidoptera: Gelechioidea: Depressariidae) from Costa Rica. *The Journal of the Lepidopterists' Society*, 75(1), 1–13. Available from: https://doi.org/10.18473/lepi.75i1.a1

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403–410. Available from: https://doi.org/10.1016/S0022-2836(05)80360-2

Andrews, S. (2010) FastQC a quality control tool for high throughput sequence data. Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Bazinet, A.L., Cummings, M.P., Mitter, K.T. & Mitter, C.W. (2013) Can RNA-seq resolve the rapid radiation of advanced moths and butterflies (Hexapoda: lepidoptera: Apoditrysia)? An exploratory study. *PLoS One*, 8(12), e82615. Available from: https://doi.org/10.1371/journal.pone.0082615

Bidzilya, O.V. & Rajaei, H. (2024) The Gelechiidae (lepidoptera) of The Democratic Republic of the Congo. *Revue Suisse de Zoologie*, 131(2), 487–510. Available from: https://doi.org/10.35929/RSZ.0133

Bolger, A.M., Lohse, M. & Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. Available from: https://doi.org/10.1093/bioinformatics/btu170

Boyes, D., Blaxter, M.L. & University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding Collective, Wellcome Sanger Institute Tree of Life Programme, Wellcome Sanger Institute Scientific Operations: DNA Pipelines Collective, Tree of Life Core Informatics Collective, Darwin Tree of Life Consortium. (2024) The genome sequence of the tipped oak case-bearer, *Coleophora flavipennella* (Duponchel 1843). *Wellcome Open Research*, 9, 3. Available from: https://doi.org/10.12688/wellcomeopenres.19917.1

Boyes, D. & Boyes, C. (2023) The genome sequence of the acer sober, *Anarsia innoxiella* (Gregersen & Karsholt, 2017). *Wellcome Open Research*, 8, 357. Available from: https://doi.org/10.12688/wellcomeopenres.19514.1

Boyes, D. & Boyes, C. (2024a) The genome sequence of the dotted Grey groundling, *Athrips mouffetella* (Linnaeus, 1758). *Wellcome Open Research*, 9, 42. Available from: https://doi.org/10.12688/wellcomeopenres.20840.1

Boyes, D., Boyes, C. & University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding Collective, Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory Team, Wellcome Sanger Institute Scientific Operations: Sequencing Operations, Wellcome Sanger Institute Tree of Life Core Informatics Team, Darwin Tree of Life Consortium. (2024b) The genome sequence of the crescent groundling, *Teleiodes luculella* (Hübner, 1813). *Wellcome Open Research*, 9, 143. Available from: https://doi.org/10.12688/wellcomeopenres.21090.1

Boyes, D., Hammond, J. & University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding Collective, Wellcome Sanger Institute Tree of Life Programme, Wellcome Sanger Institute Scientific Operations: DNA Pipelines Collective, Tree of Life Core Informatics Collective, Darwin Tree of Life Consortium. (2023) The genome sequence of the Ruddy flat-body, *Agonopterix subpropinquella* (Stainton, 1849). *Wellcome Open Research*, 8, 487. Available from: https://doi.org/10.12688/wellcomeopenres.20170.1

Boyes, D., Holland, P.W. & University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding Collective, Wellcome Sanger Institute Tree of Life Programme, Wellcome Sanger Institute Scientific Operations: DNA Pipelines Collective, Tree of Life Core Informatics Collective, Darwin Tree of Life Consortium. (2023a) The genome sequence of the brindled flat-body, Agonopterix arenella (Denis & Schiffermüller, 1775). *Wellcome Open Research*, 8, 214. Available from: https://doi.org/10.12688/wellcomeopenres.19252.1

Boyes, D., Holland, P.W. & University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding Collective, Wellcome Sanger Institute Tree of Life Programme, Wellcome

Sanger Institute Scientific Operations: DNA Pipelines Collective, Tree of Life Core Informatics Collective, Darwin Tree of Life Consortium. (2023b) The genome sequence of the Brown house-moth, *Hofmannophila pseudospretella* (Stainton, 1849). *Wellcome Open Research*, 8, 230. Available from: https://doi.org/10.12688/wellcomeopenres.19394.1

Boyes, D., Hutchinson, F. & University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding Collective, Wellcome Sanger Institute Tree of Life Programme, Tree of Life Core Informatics Collective, Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective, Darwin Tree of Life Consortium. (2023) The genome sequence of the dingy Dowd, *Blastobasis adustella* (Walsingham, 1894). *Wellcome Open Research*, 8, 407. Available from: https://doi.org/10.12688/wellcomeopenres.19900.1

Boyes, D. & Kerry, B. (2023) The genome sequence of the Common plume moth, *Emmelina monodactyla* (Linnaeus, 1758) [version 1; peer review: 1 approved with reservations]. *Wellcome Open Research*, 8(97). Available from: https://doi.org/10.12688/wellcomeopenres.19035.1

Boyes, D., Lees, D. & University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding Collective, Wellcome Sanger Institute Tree of Life Programme, Wellcome Sanger Institute Scientific Operations: DNA Pipelines Collective, Tree of Life Core Informatics Collective, Darwin Tree of Life Consortium. (2023) The genome sequence of the long-horned flat-body, *Carcina quercana* (Fabricius, 1775). *Wellcome Open Research*, 8, 16. Available from: https://doi.org/10.12688/wellcomeopenres.18596.1

Boyes, D. & Parkerson, L. (2022) The genome sequence of the common grass-veneer, *Agriphila tristella* (Denis & Schiffermu Ller, 1775) [version 1; peer review: 3 approved]. *Wellcome Open Research*, 7(304). Available from: https://doi.org/10.12688/wellcomeopenres.18568.1

Bucheli, S.R. & Wenzel, J. (2005) Gelechioidea (Insecta: lepidoptera) systematics: a reexamination using combined morphology and mitochondrial DNA data. *Molecular Phylogenetics and Evolution*, 35(2), 380–394. Available from: https://doi.org/10.1016/j.ympev.2005.02.003

Buxton, M., Anderson, B. & Lord, J. (2018) The secret service – analysis of the available knowledge on moths as pollinators in New Zealand. *New Zealand Journal of Ecology*, 42(1), 1–9. Available from: https://doi.org/10.20417/nzjecol.42.11

Calla, B., Noble, K., Johnson, R.M., Walden, K.K.O., Schuler, M.A., Robertson, H.M. et al. (2017) Cytochrome P450 diversification and hostplant utilization patterns in specialist and generalist moths: birth, death and adaptation. *Molecular Ecology*, 26(21), 6021–6035. Available from: https://doi.org/10.1111/mec.14348

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. et al. (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, 10(1), 421. Available from: https://doi.org/10.1186/1471-2105-10-421

Cantu, V.A., Sadural, J. & Edwards, R. (2019) PRINSEQ++, a multi-threaded tool for fast and efficient quality control and preprocessing of sequencing datasets [preprint]. PeerJ Preprints. Available from: https://doi.org/10.7287/peerj.preprints.27553v1

Capella-Gutierrez, S., Silla-Martinez, J.M. & Gabaldon, T. (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, 25(15), 1972–1973. Available from: https://doi.org/10.1093/bioinformatics/btp348

Cock, P.J.A., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A. et al. (2009) Biopython: freely available python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25(11), 1422–1423. Available from: https://doi.org/10.1093/bioinformatics/btp163

Common, I. (1990) *Moths of Australia*. Melbourne, Victoria: Melbourne University Press.

Darriba, D., Posada, D., Kozlov, A.M., Stamatakis, A., Morel, B. & Flouri, T. (2020) ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. *Molecular Biology and Evolution*, 37(1), 291–294. Available from: https://doi.org/10.1093/molbev/msz189

Eddy, S.R. (2008) A probabilistic model of local sequence alignment that simplifies statistical significance estimation. *PLoS Computational Biology*, 4(5), e1000069. Available from: https://doi.org/10.1371/journal.pcbi.1000069

Eddy, S.R. (2011) Accelerated profile HMM searches. *PLoS Computational Biology*, 7(10), e1002195. Available from: https://doi.org/10.1371/journal.pcbi.1002195

Espeland, M., Breinholt, J., Willmott, K.R., Warren, A.D., Vila, R., Toussaint, E.F.A. et al. (2018) A comprehensive and dated phylogenomic analysis of butterflies. *Current Biology*, 28(5), 770–778.e5. Available from: https://doi.org/10.1016/j.cub.2018.01.061

Gálvez-Merchán, Á., Min, K.H., Pachter, L. & Booeshaghi, A.S. (2023) Metadata retrieval from sequence databases with ffq. *Bioinformatics*, 39(1), btac667. Available from: https://doi.org/10.1093/bioinformatics/btac667

Gillung, J.P., Winterton, S.L., Bayless, K.M., Khouri, Z., Borowiec, M.L., Yeates, D. et al. (2018) Anchored phylogenomics unravels the evolution of spider flies (Diptera, Acroceridae) and reveals discordance between nucleotides and amino acids. *Molecular Phylogenetics and Evolution*, 128, 233–245. Available from: https://doi.org/10.1016/j.ympev.2018.08.007

Gözüaçık, C., Çetin Erdoğan, Ö. & Beyarslan, A. (2008) Note: *Syringopais temperatella* Lederer, 1855 and its parasitoids in wheat and barley fields in the southeast Anatolian region of Turkey. *Phytoparasitica*, 36(5), 489–490. Available from: https://doi.org/10.1007/BF03020295

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, 59(3), 307–321. Available from: https://doi.org/10.1093/sysbio/syq010

Haag, J., Höhler, D., Bettisworth, B. & Stamatakis, A. (2022) From easy to hopeless—predicting the difficulty of phylogenetic analyses. *Molecular Biology and Evolution*, 39(12), msac254. Available from: https://doi.org/10.1093/molbev/msac254

Heikkilä, M., Mutanen, M., Kekkonen, M. & Kaila, L. (2014) Morphology reinforces proposed molecular phylogenetic affinities: a revised classification for Gelechioidea (lepidoptera). *Cladistics*, 30(6), 563–589. Available from: https://doi.org/10.1111/cla.12064

Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q. & Vinh, L.S. (2018) UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, 35(2), 518–522. Available from: https://doi.org/10.1093/molbev/msx281

Hodges, R.W. (1998) The Gelechioidea. In: Kristensen, N.P. (Ed.) *Lepidoptera: Moths and butterflies. Handbook of Zoology/Handbuch der Zoologie 35*, Vol. 1. Berlin: Walter de Gruyter, pp. 131–158. Available from: https://doi.org/10.1515/9783110804744.131

Holland, P.W. & University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding Collective, Wellcome Sanger Institute Tree of Life Programme, Wellcome Sanger Institute Scientific Operations: DNA Pipelines Collective, Tree of Life Core Informatics Collective, Darwin Tree of Life Consortium. (2023) The genome sequence of the Sulphur Tubic, *Esperia sulphurella* (Fabricius, 1775). *Wellcome Open Research*, 8, 156. Available from: https://doi.org/10.12688/wellcomeopenres.19212.1

Kaila, L. (2004) Phylogeny of the superfamily Gelechioidea (lepidoptera: Ditrysia): an exemplar approach. *Cladistics*, 20(4), 303–340. Available from: https://doi.org/10.1111/j.1096-0031.2004.00027.x

Kaila, L., Karsholt, O. & Revilla, T. (2024) The genus *Zizyphia* Chrétien, 1908, with notes on its systematic position and the first record of *Z. cleodorella* Chrétien, 1908 from Europe (lepidoptera, Depressariidae, Cacochroinae). *Nota Lepidopterologica*, 47, 19–28. Available from: https://doi.org/10.3897/nl.47.115542

Kaila, L., Mutanen, M. & Nyman, T. (2011) Phylogeny of the mega-diverse Gelechioidea (lepidoptera): adaptations and determinants of success. *Molecular Phylogenetics and Evolution*, 61(3), 801–809. Available from: https://doi.org/10.1016/j.ympev.2011.08.016

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A. & Jermiin, L.S. (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14(6), 587–589. Available from: https://doi.org/10.1038/nmeth.4285

Katoh, K. & Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), 772–780. Available from: https://doi.org/10.1093/molbev/mst010

Kawahara, A.Y. & Breinholt, J.W. (2014) Phylogenomics provides strong evidence for relationships of butterflies and moths. *Proceedings of the Royal Society B: Biological Sciences*, 281(1788), 20140970. Available from: https://doi.org/10.1098/rspb.2014.0970

Kawahara, A.Y., Plotkin, D., Espeland, M., Meusemann, K., Toussaint, E.F., Donath, A. et al. (2019) Phylogenomics reveals the evolutionary timing and pattern of butterflies and moths. *Proceedings of the National Academy of Sciences of the United States of America*, 116(45), 22657–22663. Available from: https://doi.org/10.1073/pnas.1907847116

Kishino, H. & Hasegawa, M. (1989) Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *Journal of Molecular Evolution*, 29(2), 170–179. Available from: https://doi.org/10.1007/BF02100115

Kishino, H., Miyata, T. & Hasegawa, M. (1990) Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. *Journal of Molecular Evolution*, 31(2), 151–160. Available from: https://doi.org/10.1007/BF02109483

Kozlov, A.M., Darriba, D., Flouri, T., Morel, B. & Stamatakis, A. (2019) RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, 35(21), 4453–4455. Available from: https://doi.org/10.1093/bioinformatics/btz305

Kristensen, N.P. & Skalski, A.W. (1999) Phylogeny and Palentology. In: Kristensen, N.P. (Ed.) *Lepidoptera: moths and butterflies. Handbook of zoology/Handbuch der Zoologie 35*, Vol. 1. Berlin: Walter de Gruyter, pp. 7–25.

Lanfear, R. & Hahn, M.W. (2024) The meaning and measure of concordance factors in Phylogenomics. *Molecular Biology and Evolution*, 41(11), msae214. Available from: https://doi.org/10.1093/molbev/msae214

Laurent, R.A.S. & Kawahara, A.Y. (2019) Reclassification of the sack-bearer moths (lepidoptera, Mimallonoidea, Mimallonidae). *ZooKeys*, 815, 1–114. Available from: https://doi.org/10.3897/zookeys.815.27335

Levy Karin, E., Mirdita, M. & Söding, J. (2020) MetaEuk—sensitive, high-throughput gene discovery, and annotation for large-scale eukaryotic metagenomics. *Microbiome*, 8(1), 48. Available from: https://doi.org/10.1186/s40168-020-00808-x

Li, F., Li, Y., Wang, T., Li, T., Zhu, S., Pei, P. et al. (2020) Screening potential reference genes for reverse transcription-quantitative polymerase chain reaction normalization in *Atrijuglans hetaohei* (lepidoptera: Gelechioidea). *Entomological Research*, 50(6), 317–326. Available from: https://doi.org/10.1111/1748-5967.12444

Li, W. & Godzik, A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22(13), 1658–1659. Available from: https://doi.org/10.1093/bioinformatics/btl158

Li, X., Breinholt, J.W., Martinez, J.I., Keegan, K., Ellis, E.A., Homziak, N.T. et al. (2024) Large-scale genomic data reveal the phylogeny and evolution of owlet moths (Noctuoidea). *Cladistics*, 40(1), 21–33. Available from: https://doi.org/10.1111/cla.12559

Li, X., Fan, D., Zhang, W., Liu, G., Zhang, L., Zhao, L. et al. (2015) Outbred genome sequencing and CRISPR/Cas9 gene editing in butterflies. *Nature Communications*, 6, 8212. Available from: https://doi.org/10.1038/ncomms9212

Liu, C., Zhou, X., Li, Y., Hittinger, C.T., Pan, R., Huang, J. et al. (2024) The influence of the number of tree searches on maximum likelihood inference in Phylogenomics. *Systematic Biology*, 73(5), 807–822. Available from: https://doi.org/10.1093/sysbio/syae031

Lund University, & LU SBG. (2025) Resolving the phylogenetic bush of Gelechioidea (Lepidoptera) using phylogenomics [Data set]. European Nucleotide Archive. Available from: https://www.ebi.ac.uk/ena/browser/view/PRJEB87822

Luo, S., Li, Y., Chen, S., Zhang, D. & Renner, S.S. (2011) Gelechiidae moths are capable of chemically dissolving the pollen of their host plants: first documented sporopollenin breakdown by an animal. *PLoS One*, 6(4), e19219. Available from: https://doi.org/10.1371/journal.pone.0019219

Luz, F.A., Gonçalves, G.L., Moreira, G.R. & Becker, V.O. (2015) Description, molecular phylogeny, and natural history of a new kleptoparasitic species of gelechiid moth (lepidoptera) associated with Melastomataceae galls in Brazil. *Journal of Natural History*, 49(31–32), 1849–1875. Available from: https://doi.org/10.1080/00222933.2015.1006284

Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, 17(1), 10–12. Available from: https://doi.org/10.14806/ej.17.1.200

Mayer, C., Dietz, L., Call, E., Kukowka, S., Martin, S. & Espeland, M. (2021) Adding leaves to the lepidoptera tree: capturing hundreds of nuclear genes from old museum specimens. *Systematic Entomology*, 46(3), 649–671. Available from: https://doi.org/10.1111/syen.12481

Mead, D., Saccheri, I., Yung, C., Lohse, K., Lohse, C., Ashmole, P. et al. (2021) The genome sequence of the ringlet, *Aphantopus hyperantus* Linnaeus 1758 [version 1; peer review: 2 approved with reservations]. *Wellcome Open Research*, 6(165). Available from: https://doi.org/10.12688/wellcomeopenres.16983.1

Minet, J. (1990) Remaniement partiel de la classification des Gelechioidea, essentiellement en fonction de caractères pré-imaginaux (Lepidoptera Ditrysia). *Alexanor (Paris)*, 16(3–4), 239–255.

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A. et al. (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, 37(5), 1530–1534. Available from: https://doi.org/10.1093/molbev/msaa015

Misof, B., Liu, S., Meusemann, K., Peters, R.S., Donath, A., Mayer, C. et al. (2014) Phylogenomics resolves the timing and pattern of insect evolution. *Science*, 346(6210), 763–767. Available from: https://doi.org/10.1126/science.1257570

Murillo-Ramos, L., Twort, V., Wahlberg, N. & Sihvonen, P. (2023) A phylogenomic perspective on the relationships of subfamilies in the family Geometridae (lepidoptera). *Systematic Entomology*, 48(4), 618–632. Available from: https://doi.org/10.1111/syen.12594

O'Cathail, C., Ahamed, A., Burgin, J., Cummins, C., Devaraj, R., Gueye, K. et al. (2025) The European nucleotide archive in 2024. *Nucleic Acids Research*, 53(D1), D49–D55. Available from: https://doi.org/10.1093/nar/gkae975

O'Leary, N.A., Cox, E., Holmes, J.B., Anderson, W.R., Falk, R., Hem, V. et al. (2024) Exploring and retrieving sequence and metadata for species across the tree of life with NCBI datasets. *Scientific Data*, 11(1), 732. Available from: https://doi.org/10.1038/s41597-024-03571-y

Õunap, E., Nedumpally, V., Yapar, E., Lemmon, A.R. & Tammaru, T. (2025) Molecular phylogeny of north European Geometridae (lepidoptera: Geometroidea). *Systematic Entomology*, 50(1), 32–67. Available from: https://doi.org/10.1111/syen.12638

Paradis, E. & Schliep, K. (2019) Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35(3), 526–528. Available from: https://doi.org/10.1093/bioinformatics/bty633

Peña, C. & Malm, T. (2012) VoSeq: a voucher and DNA sequence web application. *PLoS One*, 7(6), e39071. Available from: https://doi.org/10.1371/journal.pone.0039071

Poelchau, M., Childers, C., Moore, G., Tsavatapalli, V., Evans, J., Lee, C.-Y. et al. (2015) The i5k workspace@NAL—enabling genomic data access, visualization and curation of arthropod genomes. *Nucleic Acids Research*, 43(D1), D714–D719. Available from: https://doi.org/10.1093/nar/gku983

Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A. & Korobeynikov, A. (2020) Using SPAdes De Novo assembler. *Current Protocols in Bioinformatics*, 70(1), e102. Available from: https://doi.org/10.1002/cpbi.102

R Core Team. (2022) *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/

Raine, J. (1966) Life history of *Dasystoma salicellum* Hbn. (lepidoptera: Oecophoridae), a new Pest of blueberries in British Columbia. *The Canadian Entomologist*, 98(3), 331–334. Available from: https://doi.org/10.4039/Ent98331-3

Ratnasingham, S. & Hebert, P.D.N. (2007) Bold: the barcode of life data system (http://www.Barcodinglife.Org). *Molecular Ecology Notes*, 7(3), 355–364. Available from: https://doi.org/10.1111/j.1471-8286.2007.01678.x

Rota, J., Twort, V., Chiocchio, A., Peña, C., Wheat, C.W., Kaila, L. et al. (2022) The unresolved phylogenomic tree of butterflies and moths (lepidoptera): assessing the potential causes and consequences. *Systematic Entomology*, 47, 1–20. Available from: https://doi.org/10.1111/syen.12545

Rubinoff, D. & Haines, W.P. (2005) Web-spinning Caterpillar stalks snails. *Science*, 309(5734), 575. Available from: https://doi.org/10.1126/science.1110397

Sayers, E.W., Beck, J., Bolton, E.E., Brister, J.R., Chan, J., Comeau, D.C. et al. (2024) Database resources of the National Center for biotechnology information. *Nucleic Acids Research*, 52(D1), D33–D43. Available from: https://doi.org/10.1093/nar/gkad1044

Shimodaira, H. (2002) An approximately unbiased test of phylogenetic tree selection. *Systematic Biology*, 51(3), 492–508. Available from: https://doi.org/10.1080/10635150290069913

Shimodaira, H. & Hasegawa, M. (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution*, 16(8), 1114–1116. Available from: https://doi.org/10.1093/oxfordjournals.molbev.a026201

Sohn, J.-C., Regier, J.C., Mitter, C., Adamski, D., Landry, J.-F., Heikkilä, M. et al. (2016) Phylogeny and feeding trait evolution of the mega-diverse Gelechioidea (lepidoptera: Obtectomera): new insight from 19 nuclear genes. *Systematic Entomology*, 41(1), 112–132. Available from: https://doi.org/10.1111/syen.12143

Stahlke, A.R., Chang, J., Chudalayandi, S., Heu, C.C., Geib, S.M., Scheffler, B.E. et al. (2023) Chromosome-scale genome assembly of the pink bollworm, *Pectinophora gossypiella*, a global pest of cotton. *G3: Genes, Genomes, Genetics*, 13(4), jkad040. Available from: https://doi.org/10.1093/g3journal/jkad040

Strimmer, K. & Rambaut, A. (2002) Inferring confidence sets of possibly misspecified gene trees. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 269(1487), 137–142. Available from: https://doi.org/10.1098/rspb.2001.1862

Tabuloc, C.A., Lewald, K.M., Conner, W.R., Lee, Y., Lee, E.K., Cain, A.B. et al. (2019) Sequencing of *Tuta absoluta* genome to develop SNP genotyping assays for species identification. *Journal of Pest Science*, 92(4), 1397–1407. Available from: https://doi.org/10.1007/s10340-019-01116-6

Twort, V.G., Minet, J., Wheat, C.W. & Wahlberg, N. (2021) Museomics of a rare taxon: placing Whalleyanidae in the lepidoptera tree of life. *Systematic Entomology*, 46(4), 926–937. Available from: https://doi.org/10.1111/syen.12503

van Nieukerken, E.J., Kaila, L., Kitching, I.J., Kristensen, N.P., Lees, D.C., Minet, J. et al. (2011) Order lepidoptera Linnaeus, 1758. In: Zhang, Z.-Q. (ed.) animal biodiversity: an outline of higher-level classification and survey of taxonomic richness. *Zootaxa*, 3148(1), 212–221. Available from: https://doi.org/10.11646/zootaxa.3148.1.41

Wang, D., Tao, J., Lu, P., Luo, Y. & Hu, P. (2020) The whole body transcriptome of *Coleophora obducta* reveals important olfactory proteins. *PeerJ*, 8, e8902. Available from: https://doi.org/10.7717/peerj.8902

Wang, Q.-Y. & Li, H.-H. (2020) Phylogeny of the superfamily Gelechioidea (lepidoptera: Obtectomera), with an exploratory application on geometric morphometrics. *Zoologica Scripta*, 49(3), 307–328. Available from: https://doi.org/10.1111/zsc.12407

White, W.H., Reagan, T.E., Carlton, C., Akbar, W. & Beuzelin, J.M. (2007) *Elachista saccharella* (lepidoptera: Elachistidae), a Leafminer infesting sugarcane in Louisiana. *Florida Entomologist*, 90(4), 792–794. Available from: https://doi.org/10.1653/0015-4040(2007)90[792:ESLEAL]2.0.CO;2

Xu, D., Yang, H., Zhuo, Z., Lu, B., Hu, J. & Yang, F. (2021) Characterization and analysis of the transcriptome in *Opisina arenosella* from different developmental stages using single-molecule real-time transcript sequencing and RNA-seq. *International Journal of Biological Macromolecules*, 169, 216–227. Available from: https://doi.org/10.1016/j.ijbiomac.2020.12.098

Yan, J.-J., Zhang, M.-D. & Gao, Y.-L. (2019) Biology, ecology and integrated management of the potato tuber moth, *Phthorimaea operculella* (lepidoptera: Gelechiidae). *Acta Entomologica Sinica*, 62(12), 1469–1482. Available from: https://doi.org/10.16380/j.kcxb.2019.12.012

Yang, Y. & Smith, S.A. (2014) Orthology inference in nonmodel organisms using transcriptomes and low-coverage genomes: improving accuracy and matrix occupancy for Phylogenomics. *Molecular Biology and Evolution*, 31(11), 3081–3092. Available from: https://doi.org/10.1093/molbev/msu245

Yapar, E. (2025) Integrating Sanger and next generation sequencing data sheds light on phylogenetic relationships among gelechioid moths (Lepidoptera: Gelechioidea) [Data set]. Zenodo. Available from: https://zenodo.org/records/16173549

Yu, G., Smith, D.K., Zhu, H., Guan, Y. & Lam, T.T.-Y. (2017) Ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, 8(1), 28–36. Available from: https://doi.org/10.1111/2041-210X.12628

Zhang, C. & Mirarab, S. (2022) Weighting by gene tree uncertainty improves accuracy of quartet-based species trees. *Molecular Biology and Evolution*, 39(12), msac215. Available from: https://doi.org/10.1093/molbev/msac215

Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. (2018) ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*, 19(6), 153. Available from: https://doi.org/10.1186/s12859-018-2129-y

Zhang, M., Cheng, X., Lin, R., Xie, B., Nauen, R., Rondon, S.I. et al. (2022) Chromosomal-level genome assembly of potato tuberworm, *Phthorimaea operculella*: a pest of solanaceous crops. *Scientific Data*, 9(1), 748. Available from: https://doi.org/10.1038/s41597-022-01859-5

Zwick, A., Regier, J.C. & Zwickl, D.J. (2012) Resolving discrepancy between nucleotides and amino acids in deep-level arthropod Phylogenomics: differentiating serine codons in 21-amino-acid models. *PLoS One*, 7(11), e47450. Available from: https://doi.org/10.1371/journal.pone.0047450

SUPPORTING INFORMATION

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

Figure S1. Maximum-likelihood tree from the denser dataset. Branch support (standard Felsenstein bootstraps, FBP) is shown categorically with node symbols.

Figure S2. ML tree from the phylogenomic analysis #1 (Table 1). Node symbols indicate if a given node is absent in any other five phylogenomic trees.

Figure S3. ML tree from the phylogenomic analysis #2 (Table 1). Node symbols indicate if a given node is absent in any other five phylogenomic trees.

Figure S4. ML tree from the phylogenomic analysis #4 (Table 1). Node symbols indicate if a given node is absent in any other five phylogenomic trees.

Figure S5. ML tree from the phylogenomic analysis #3 (Table 1). Node symbols indicate if a given node is absent in any other five phylogenomic trees.

Figure S6. ML tree from the phylogenomic analysis #6 (Table 1). Node symbols indicate if a given node is absent in any other five phylogenomic trees.

Figure S7. ML tree from the phylogenomic analysis #5 (Table 1). Node symbols indicate if a given node is absent in any other five phylogenomic trees.

Figure S8. Cladogram of the ML tree topology from the phylogenomic dataset (Figure 4) with gene, site and quartet concordance factors shown as node labels as gCF/sCF/qCF.

Figure S9. Cladogram of the ASTRAL topology from the phylogenomic dataset (Figure 4) with local posterior probabilities shown as node labels.

Figure S10. An example case of the discordance between the ML and the ASTRAL topologies displaying the relationship between *Elachista tengstromi*, the remaining Elachistidae, the GC clade and the rest of the tree. GC clade: Gelechiidae+Cosmopterigidae.

Table S1. Taxon sampling for the core phylogenomic dataset. The consortium/initiative, the project or the government institution is given as the source when a publication is not available.

Table S2. Voucher code and gene information table for the Sanger sequencing samples. "-" and "X" denote absence and presence, respectively. GenBank accession code is given for the present genes if applicable.

How to cite this article: Yapar, E., Chiocchio, A., Heikkilä, M., Rota, J., Kaila, L. & Wahlberg, N. (2026) Integrating Sanger and next-generation sequencing data sheds light on phylogenetic relationships among gelechioid moths (Lepidoptera: Gelechioidea). *Systematic Entomology*, 51(1), e70009. Available from: https://doi.org/10.1111/syen.70009