Molecular Phylogenetics and Evolution 184 (2023) 107778

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

The phylogeny and divergence times of leaf-mining flies (Diptera: Agromyzidae) from anchored phylogenomics

Jing-Li Xuan [a,c,d], Sonja J. Scheffer [b,*], Matt Lewis [b], Brian K. Cassel [a], Wan-Xue Liu [d], Brian M. Wiegmann [a,*]

[a] Department of Entomology and Plant Pathology, North Carolina State University, Raleigh, NC 27695, USA
[b] Systematic Entomology Laboratory, USDA, Beltsville Agricultural Research Station, Beltsville, MD 20705, USA
[c] Anhui Bio-breeding Engineering Research Center for Water melon and Melon, College of Life Sciences, Huaibei Normal University, Huaibei 235000, PR China
[d] State Key Laboratory for Biology of Plant Diseases and Insect Pests, Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing 100193, PR China

ARTICLE INFO

Keywords:
Diptera
Leaf-miner
Phylogenomics
Divergence times

ABSTRACT

Leaf-mining flies (Diptera: Agromyzidae) are a diverse clade of phytophagous Diptera known largely for their economic impact as leaf- or stem-miners on vegetable and ornamental plants. Higher-level phylogenetic relationships of Agromyzidae have remained uncertain because of challenges in sampling of both taxa and characters for morphology and PCR-based Sanger-era molecular systematics. Here, we used hundreds of orthologous single-copy nuclear loci obtained from anchored hybrid enrichment (AHE) to reconstruct phylogenetic relationships among the major lineages of leaf-mining flies. The resulting phylogenetic trees are highly congruent and well-supported, except for a few deep nodes, when using different molecular data types and phylogenetic methods. Based on divergence time dating using a relaxed clock model-based analysis, leaf-mining flies are shown to have diversified in multiple lineages since the early Paleocene, approximately 65 million years ago. Our study not only reveals a revised classification system of leaf-mining flies, but also provides a new phylogenetic framework to understand their macroevolution.

1. Introduction

Leaf-mining flies (Diptera: Agromyzidae), sometimes known as serpentine leaf miners, are a diverse group of schizophoran flies with approximately 3,163 valid species across 26 recognized genera distributed worldwide (Winkler et al., 2009a; Lonsdale, 2014; Von Tschirnhaus, 2021a; Table S2). All leaf-mining larvae are phytophagous inside living plant tissue, damaging more than 140 plant families across liverworts, ferns, horsetails, gymnosperms, and angiosperms including monocots and eudicots (Spencer, 1990; Scheffer et al., 2007; Eiseman, 2021). Many agromyzid species are recorded as monophagous or oligophagous, where larval feeding is restricted to a single or multiple closely related plants, while relatively few species are known to be polyphagous feeding on a wide variety of host plants across distantly related plant families (Spencer, 1990; Lonsdale, 2021). Larval feeding habits of agromyzids are highly diverse in the taxonomic breadth of hosts they use, and in the range of plant tissues they consume (Spencer, 1990; Eiseman, 2021). Most species feed internally on leaves, as leaf-miners, while about 25% of known species feed on flower heads, seed pods, twigs, stems, roots, or the cambium of woody plants (Spencer, 1973, 1990; Eiseman, 2021; Lonsdale, 2021). A few species can produce mines in different parts of the host plant, for example, *Ophiomyia phaseoli* (Tryon) is able to shift from leaf to stem mining (Lonsdale, 2021). Leaf-mining larvae generate various specific types of mines including linear, linear-blotch, and blotch (Hering, 1951; Eiseman, 2021; Lonsdale, 2021), and pupate inside or outside the larval host plant tissue (Spencer, 1990; Civelek, 2003; Dempewolf, 2005; Eiseman et al., 2019; Spreadsheet in the supplementary files).

The Family Agromyzidae has traditionally been divided into two subfamilies, Agromyzinae and Phytomyzinae, which follows the system containing ‘agromyzides’ and ‘phytomyzides’ originally devised by Fallén (1823a, 1823b). This system of classification in Agromyzidae has been examined using phylogenies reconstructed with morphology (Dempewolf, 2001; 2005) and molecular gene markers (Scheffer et al., 2007). Definitions of the two subfamilies are based on larval mouthpart morphology (McAlpine et al., 1987; Dempewolf, 2001) and wing venation of adults (Spencer, 1973; Spencer and Steyskal, 1986; Lonsdale, 2021).

---

* Corresponding authors.
E-mail addresses: sonja.scheffer@usda.gov (S.J. Scheffer), bwiegman@ncsu.edu (B.M. Wiegmann).

https://doi.org/10.1016/j.ympev.2023.107778
Received 17 October 2022; Received in revised form 24 February 2023; Accepted 31 March 2023

The Subfamily Agromyzinae has been divided into three distinct clades based on morphology, with the monophyly of the entire subfamily widely accepted historically (Lonsdale, 2014; 2021). The first clade, *Ophiomyia* group, encompasses the genera *Ophiomyia*, *Melanagromyza*, *Euhexomyza* (orphan species of genus *Hexomyza*, the type species of which belongs in *Ophiomyia*, resulting in a synonymy of the genus), and the Old-World *Tropicomyia* (Lonsdale, 2014). Spencer (1966) considered *Melanagromyza* to be sister group to a monophyletic clade including genera *Euhexomyza* and *Ophiomyia* based on morphological features and life history traits. The other two Agromyzinae clades each exclusively contain a single genus, *Agromyza* and *Japanagromyza*, but these genera have sometimes been allied since many species share external features, such as developed prescutellar acrostichal setae and pale halteres (Lonsdale, 2013). Some *Japanagromyza* and *Agromyza* also have an inner-distal comb of setae on the hind tibia (von Tschirnhaus, 1991). *Japanagromyza* is distributed globally and is more diverse in male genitalic characters, suggesting that the genus may be non-monophyletic (Lonsdale, 2013).

The Subfamily Phytomyzinae has greater morphological, generic and species diversity than Agromyzinae (Lonsdale, 2021). Spencer (1990) divided this subfamily into three groups, the *Napomyza*, *Phytoliriomyza*, and *Phytobia* groups. Within the '*Napomyza* group', Winkler et al. (2009a) advanced the classification of the genus *Phytomyza* based on a molecular phylogeny generated from multiple gene markers in which *Napomyza* and *Ptochomyza* were recognized as subgenera of *Phytomyza,* and *Chromatomyia* was resynonymised with the genus *Phytomyza.* von Tschirnhaus (2021b) argued against this this action, but Lonsdale and Eiseman (2021) provided additional morphological and life history evidence in support of the synonymy of *Chromatomyia*. An additional three little-known genera (*Gymnophytomyza*, *Aulagromyza*, and *Pseudonapomyza*) are of uncertain placement in the subfamily (Zlobin, 2002, 2007; Lonsdale, 2021) which differ in their placement in morphological and molecular trees (Dempewolf, 2001; 2005; Scheffer et al., 2007).

There is no strong morphological or molecular evidence supporting the '*Phytoliriomyza* group', which includes many diverse leaf-mining genera (Scheffer et al., 2007; Lonsdale, 2021). Six genera consisting of *Calycomyza*, *Haplopeodes*, *Liriomyza*, *Metopomyza*, *Phytoliriomyza*, and *Selachops*, may be grouped by their shared tubercle-like setae on the posterodistal margin of the epandrium and surstylus (Lonsdale, 2021). This possibly monophyletic clade has been verified in a molecular phylogeny of Agromyzidae (Scheffer et al., 2007), but was not supported in a morphological phylogeny by Dempewolf (2001; 2005), who suggested that a *Amauromyza* + *Cerodontha* rendered the remainder of the genus group paraphyletic (Dempewolf, 2001; 2005). The genus *Selachops* is a small Palearctic group of three species (Zlobin, 1984; von Tschirnhaus, 2021a) that Lonsdale (2021) noted shared numerous similarities with *Metopomyza*. The '*Phytobia* group', is comprised of *Amauromyza*, *Nemorimyza*, and *Phytobia* (Spencer, 1990). This group was considered an early diverging lineage in leaf-mining flies as they share a large body size and many have stem- or cambium-boring larval feeding habits (Nowakowski, 1962; Spencer and Steyskal, 1986; Spencer, 1990), however, subsequently, other agromyzid researchers discredited this hypothesis based on morphological and molecular phylogenies (Dempewolf, 2001; 2005; Scheffer et al., 2007).

Although phylogenies of leaf-mining flies have been inferred from morphology and molecular gene markers, the relationships among most groups are still uncertain or not well-resolved (Dempewolf, 2001, 2005; Scheffer et al., 2007). A robust phylogenetic framework requires a broad sampling and more informative molecular dataset. In the present study, we employed an anchored hybrid enrichment (AHE) approach to capture orthologous single copy nuclear genes using designed probes based on dipteran genomes and transcriptomes. Our specific goals are to reconstruct the phylogenetic relationships among major lineages of leaf-mining flies using phylogenomic data; and use these phylogenomic data to estimate times of major agromyzid divergences to place leaf-mining fly lineages into a more accurate temporal framework for on-going macroevolutionary analysis (Xuan et al., 2022).

2. Materials and methods

2.1. Taxon sampling

We sampled a total of 139 fly specimens, including 136 ingroup exemplars and 3 outgroup representatives from the two families previously associated with Agromyzidae: Fergusoninidae and Odiniidae. Our ingroup sampling represents about 69% (18 out of 26) of all currently valid genera (Table S2), and includes 124 species, with additional representatives samples for subgenera (10) of *Phytomyza* and *Cerodontha* (Table S1 and S2). Most of the ingroup representatives we included in the study are monophagous species feeding on hosts in a single plant family. Five are polyphagous species feeding on greater than four plant families, and eight are oligophagous species feeding on hosts in two or three plant families (Spreadsheet in the supplementary files). The families Fergusoninidae and Odiniidae were chosen as outgroups as they have been found to be closely related to Agromyzidae based on McAlpine's (1989) classification and the recent phylogenomic study of the families of schizophoran Diptera (Bayless et al., 2021). Specimens were field collected using Malaise traps, hand nets, and/or reared from host plants. Full collection information for the specimens sampled are provided in Table S1. All specimens were preserved in 95% ethanol and stored at −80 °C prior to DNA extraction.

2.2. Species identification based on morphology

Most specimens were provided and identified by our collaborators and by the coauthors of this study. Male specimens were preferentially selected and used for downstream DNA extraction to allow species identification by male genitalic characters, especially for many species in which external characters of adults are homogeneous. Female samples were used only when adult males were unavailable and could be confirmed by associated rearing records. Genus-level IDs and some species identifications in this family were based on external and internal morphological characters using taxonomic keys (Spencer and Steyskal, 1986; McAlpine et al., 1987; Lonsdale, 2011; 2021). All voucher specimens and associated genomic DNA extractions (Table S1) are stored in the genomic DNA and voucher specimen archive of the Scheffer Laboratory (SJS) of the USDA ARS Systematic Entomology Laboratory, Beltsville, MD USA.

2.3. Whole-genomic DNA extraction

Whole-genomic DNA for specimen was extracted using the DNeasy Blood & Tissue Kit (Qiagen Inc., Hilden, Germany) following the manufacturer's protocol, with the exception that the final step was modified in that the genomic DNA was eluted twice with 30 ul of the AE buffer. Initial DNA concentration for each sample was quantified using dsDNA High Sensitivity Assay Kit with a Qubit® 2.0 fluorometer (Invitrogen by Life Technologies Inc., Oregon, USA). Specimens after DNA extractions were deposited back with 95% ethanol as vouchers.

2.4. Anchored hybrid enrichment lab work processing

Because most agromyzids possess small body sizes (e.g., *Liriomyza* [1–3 mm]), they often yield lower genomic DNA concentrations that cannot satisfy the minimum threshold of DNA demands for downstream library preparation. To obtain sufficient DNA concentrations, we amplified genomic DNA using the REPLI-g Mini Kit (Qiagen Inc., Hilden, Germany) following the standard manufacturer's protocol. Gel electrophoresis (0.8% agarose) was also run to assess DNA integrity. To prepare genomic template for target enrichment, we sheared whole genomic DNAs to a target fragment size of approximately 300 base pairs (bp) by sonication on a Covaris S220/E220 Focused-ultrasonicator (Covaris Inc., Massachusetts, USA) in Covaris microtubes for 30 s. Sheared DNA was used as input for library preparation.

For single-end (SE) raw reads collection, we followed the protocol published by Lemmon et al. (2012), while, for paired-end (PE) raw reads collection, library preparation was conducted following the protocol using NEBNext® Ultra^{TM} II DNA Library Prep Kit (New England Biolabs Inc., Massachusetts, USA) for Illumina NEB7103 with sample purification beads. Dual index primers (Set 1 and Set 2) supplied in the NEBNext Multiplex Oligos Kit for Illumina (New England Biolabs Inc., Massachusetts, USA), were employed. Specifically, indexes were added into each sample followed by a protocol of Meyer and Kircher (2010). Subsequently, indexed samples were pooled in normalized quantities into Lo-bind 1.7 ml tubes, and target enrichment was performed on each pool using the Diptera AHE Probes (Young et al., 2016) with SureSelect Target Enrichment System Kits (Agilent Technologies Inc., California, USA). After enrichment, DNA concentration in each pool was detected using a Qubit® 2.0 fluorometer with dsDNA High Sensitivity Assay Kit (Invitrogen by Life Technologies Inc., Oregon, USA).

Sequencing was carried out on the Illumina NovaSeq 6000 sequencing System for PE raw reads collection (2 × 150 bp), while a few samples were sequenced on the Illumina HiSeq 2500 platform for collection SE raw reads (125 bp). Information of Illumina raw reads collection for specimens is shown in Table S1. High-throughput sequencing was conducted at the Genomic Science Laboratory (GSL, https://research.ncsu.edu/gsl/), North Carolina State University (NCSU). The genomic raw reads obtained by AHE included in this study were deposited with the Sequence Read Archives (SRA) PRJNA859188 and PRJNA858472 at the National Center for Biotechnology Information (NCBI) database. And RNA-seq reads of one outgroup taxon (*Odinia conspicua*) were downloaded from NCBI (Bioproject: PRJNA587382).

2.5. Raw reads processed with bioinformatic tools

Raw reads from Illumina sequencing were demultiplexed by indexes for each sample using CASSAVA v.1.8.2 (GSL, NCSU). The quality of raw reads was checked using FASTQC v.0.11.8 in Galaxy (https://usegalaxy.org/). Subsequently, adaptors and low-quality raw reads were removed using the trimming program TRIMMOMATIC v.0.36 (Bolger et al., 2014). And the parameter settings in TRIMMOMATIC are as follows: clipping adapters (ILLUMINACLIP:TruSeq3-PE.fa:2:30:10); removal of leading and trailing sequences below quality = 15 (LEADING:15 TRAILING:15); scanning by 4-base sliding window and cutting when the average quality per base drops below 15 (SLIDINGWINDOW:4:15); dropping reads below 36 bp in length (MINLEN:36). *De novo* assembly was conducted using the trimmed reads without reference genomes implemented in TRINITY v.2.2 (Grabherr et al., 2011). Default parameters were used with the exception that we assigned 9 CPUs and 80G maximum memory on the NCSU STATGEN Bioinformatics Research Center and High Performance Computing Henry2 clusters. We also used SPADES software (Bankevich et al., 2012) to conduct *de novo* assembly for multiple taxa with very large amounts of reads using the default parameters with the exception that we assigned 7 threads and 50G memory limit on the NCSU clusters.

To identify orthologs within the assembled contigs captured using our probe designs, we carried out orthology prediction using a graph-based, reciprocal blast approach with hidden Markov model predictions (pHMMs) implemented in OrthoGraph v.0.6.1 (Petersen et al., 2017). Specifically, orthologous genes were identified based on an orthologous reference set ‘Mecopterida’ database, which contains orthologous gene clusters identified from five reference species: *Bombyx mori* Linnaeus (Lepidoptera: Bombycidae), *Danaus plexippus* (Linnaeus) (Lepidoptera: Nymphalidae), *Aedes aegypti* (Linnaeus) (Diptera: Culicidae), *Drosophila melanogaster* (Diptera: Drosophilidae) and *Glossina morsitans* Westwood (Diptera: Glossinidae).

After ortholog prediction, we conducted a cleaning step (removal of non-fly genes) in an exhaustive locus-by-locus check by searching against genome and transcriptome databases of Diptera species available from NCBI GenBank (Bethesda, Maryland, USA). All the orthologous gene sequences obtained were verified using the program BLAST (Altschul et al., 1990) as top hits for dipteran genes. The non-fly genes were removed using package ‘taxonomizr’ (https://github.com/sherrillmix/taxonomizr) in RStudio v.1.2 (RStudio Team, 2020). Specifically, if orthologs were not matched in the database as top hits for Diptera, they were removed from processing in downstream pipeline steps for phylogenetic analysis. After contaminants removal, ortholog amino acid sequences were aligned using MAFFT v.7.481 (Katoh and Standley, 2013) with L-INS-i algorithm (–localpair, –addfragments, and –maxiterate 1000 flags). To assess the quality of multiple sequence alignments (MSAs), we followed the procedures of Misof et al. (2014) to define outlier sequences and remove them from both amino acid and nucleotide alignments (Evangelista et al., 2019). Reference-taxon sequences and all gap-only data columns were removed from sequence alignments.

To reduce the effects of random, ambiguous, or erroneous alignment regions at the amino acid level, we checked each orthologous gene set using ALISCORE v.2.2 (Misof and Misof, 2009; Kück et al., 2010; Misof et al., 2014; Buenaventura et al., 2020; Bayless et al., 2021). Ambiguously or erroneously aligned regions or sections were removed from alignment sequences using ALICUT and a customized Perl script (Misof and Misof, 2009; Kück et al., 2010). Refined nucleotide sequences were realigned using the amino acid alignments as blueprint in PAL2NAL (Suyama et al., 2006; Misof et al., 2014; Pauli et al., 2018; Bayless et al., 2021). Trailing end of indels were filled with ‘X’ in amino acids and with ‘N’ in nucleotides, respectively. Finally, we separately concatenated the MSAs for amino acids and nucleotides into the respective phylogenetic supermatrics using FASCONCAT-G (Kück and Longo, 2014).

2.6. Phylogenetic analysis

Maximum Likelihood (ML) trees were inferred from amino acid (TAA) and nucleotide datasets with the third protein-coding positions retained (NT123) and removed (NT12) using IQ-TREE v.2.1.2 (Nguyen et al., 2015; Minh et al., 2020) either implemented on XSEDE of the CIPRES (Cyberinfrastructure for Phylogenetic Research) Science Gateway v.3.3 (CSG) (Miller et al., 2010) or on the Henry2 computing cluster at NCSU. We conducted ML analysis within the ‘-m TESTNEW’ option that allows each locus to have its own evolutionary rate, and a specified partition scheme using ‘-spp’ command. A model-selection method was implemented in ModelFinder (Kalyaanamoorthy et al., 2017) to select the best-fitting substitution model for each locus based on Bayesian Information Criterion (BIC) score using the rcluster algorithm (Lanfear et al., 2014). We used Ultrafast bootstrap approximation approach (UFBoot; Minh et al., 2013; Hoang et al., 2018) to estimate the bootstrap support value with 1000 replicates (-bb 1000) and set the ‘bnni’ parameter to reduce overestimating support with ‘-bnni’ option. For individual branch support, we conducted the Shimodaira-Hasegawa-like approximate Likelihood Ratio Test (SH-aLRT; Shimodaira and Hasegawa, 1999) with 1000 replicates using the ‘-alrt 1000′ command. UFBoot and SH-aLRT support was assigned in a single run. Totally, we conducted ten replicate maximum likelihood tree search runs for each dataset.

Phylogenetic trees were also constructed using Bayesian inference (BI) from concatenated nucleotide alignments under the General Time-Reversible (GTR) model of nucleotide substitution (Tavaré, 1986) with four independent runs and four coupled chains starting from parsimony trees. Markov Chain Monte Carlo (MCMC) sampling generations were set as 4×10^8 for each chain implemented in EXABAYES v.1.5 (Aberer et al., 2014) run on Henry2 computing cluster at NCSU. It was set to draw a sample from each cold chain every 1000 generations via the option ‘samplingFreq’. Average standard deviation of split frequencies (asdsf) was used to implement diagnostics for topological congruence. After each completed MCMC run, the initial 20% of all sampled trees were discarded as burn-in. A consensus tree was constructed from four independent runs using the post-processing tool 'consense' included in EXABAYES package, and a parameter file was generated using the postprocessing tool 'postProcParam' to check how well the parameters were sampled in the settings. We also assessed the effective sample size (ESS) value for each parameter and the potential scale reduction factor (PSRF) for statistical congruence among multiple runs.

Multispecies coalescent (MSC) analysis was performed using the gene tree summary-based method implemented in ASTRAL-III v.5.6.3 (Zhang et al., 2018). Individual gene trees were built from the partitioned orthologous loci with the third position sites included and excluded, respectively, using ML approach under GTR model implemented in IQ-TREE.

2.7. Likelihood mapping analysis

To assess phylogenetic information content, we separately performed likelihood mapping analysis (Strimmer and von Haeseler, 1997) using NT123 and TAA datasets implemented in IQ-TREE. This approach evaluated the support for all possible nodal quartets and displayed them in a triangular graph. A sequence model was automatically selected using the '-m test' command option. The best-fitting model chosen for nucleotide alignments was GTR + F + I + G4, and for amino acid alignments with JTTDCMut + F + I + G4, according to the Bayesian Information Criterion (BIC) score, respectively. A subsequent tree search was skipped during likelihood mapping analysis. Two nexus files were defined containing the taxon clusters to test specific sets of phylogenetic relationships. The first defines four clusters as follows: cluster 1 = *Japanagromyza* (containing 1 taxon); cluster 2 = *Agromyza* (consisting of 17 taxa); cluster 3 = 'TM' clade (including genera *Tropicomyia*, *Euhexomyza*, *Ophiomyia*, and *Melanagromyza*; containing 23 taxa), cluster 4 = Phytomyzinae (including 95 taxa). The second nexus file was defined the following clusters for testing: cluster 1 = Agromyzinae (41 taxa); cluster 2 = 'PNP' clade (comprising *Pseudonapomyza*, *Nemorimyza*, *Phytobia*; 8 taxa); cluster 3 = 'AP' clade (consisting of *Aulagromyza* and *Phytomyza*; 32 taxa); cluster 4 = 'CL' clade (containing *Cerodontha*, *Amauromyza*, *Calycomyza*, *Phytoliriomyza*, *Selachops*, *Metopomyza*, and *Liriomyza*; 55 taxa). Outgroups are not included in the likelihood mapping analysis.

2.8. Divergence time estimation

The ML tree inferred from dataset NT123 was used to estimate divergence times among leaf-mining fly clades using the approximate likelihood calculation method implemented in MCMCTREE of the PAML package v.4.10 (dos Reis and Yang, 2011). Divergence times were estimated in the following steps. First, Gradient (G) and Hessian (H) matrices of branch lengths for each data partition were estimated on the ML tree using MCMCTREE and BASEML programs, with the DNA substitution model set to HKY85, and alpha set to 0.5 for gamma distributed rates among sites. Second, we conducted MCMC sampling from the posterior. We adopted an independent-rate model where the rate of evolution in each lineage on the tree was assumed fully independent (Drummond et al., 2006; Rannala and Yang, 2007; Paradis, 2013). Three fossil calibration points were used in this study: *Melanagromyza tephrias* Melander (age: 37.8–33.9 million years ago [Ma]; Melander et al., 1949), *Agromyza praecursor* Melander (age: 37.8–33.9 Ma; Melander et al., 1949), *Phytomyza vetusta* Theobald (age: 33.9–28.5 Ma; Theobald, 1937) (EDNA Fossil Insect Database, Mitchell, 2013; https://fossilinsectdatabase.co.uk/search.php) (Evenhuis, 1994; https://hbs.bishopmuseum.org/fossilcat). The minimum-age calibration for *Melanagromyza* and *Agromyza* were respectively constrained at 38 Ma, and 34 Ma. The root age was constrained at less than 65 Ma to reflect the previously estimated age of origin for the Agromyzidae at 65 Ma (Wiegmann et al., 2011). This root-node constraint is also bounded by the more general finding that, despite significant dipteran diversity recovered from Cretaceous-aged amber deposits, no members of the cyclorrhaphan fly clade Schizophora (including Agromyzidae and other acalyptrate families) have ever been recovered from the Cretaceous (145–65.5 Ma) (Grimaldi 2018). The birth–death process parameters were set as $\lambda = \mu_{BD} = 1$ and $\rho = 0.1$. We calculated the likelihood under the HKY85 substitution model and set alpha = 0.5 for gamma rates at sites. The gamma-Dirichlet prior was set on the mean substitution rate for partition and the rate variance parameter σ^2 (dos Reis et al., 2014). The concentration parameter was set to 1 for a symmetric Dirichlet distribution that is used to spread the rate prior across partitions (Anisimova, 2019). The total MCMC chain ran for 4,020,000 iterations, initial 10% of which was discarded as burn-in.

Finally, we assessed convergence on multiple independent MCMC runs. Here, we conducted six independent MCMC runs simultaneously and plotted the posterior distributions of mean age and their corresponding highest posterior density (HPD) 95% confidence interval (CI) on each node using packages dplyr, reshape, ggplot2, and HDInternal in Rstudio. The six independent MCMC runs converged on a single posterior distribution (Fig. S1). Therefore, we combined the six mcmc files using MCMCtree summarizer (Flouri, 2018) into a single combined file, then mapped the summarized results of this combined file on the tree using the MCMCTREE program.

3. Results

3.1. AHE locus capture

The complete nucleotide supermatrix includes 496 loci (329,685 sites; 170,639 parsimony-informative sites) for 139 terminals with an average of 51% missing data in the matrix. The mean nucleotide locus length is 622 bp, and 73% of loci are greater than 400 bp in length.

3.2. Phylogeny

Tree topologies generated from different types of data (NT123; NT12; and TAA) and distinct phylogenetic methods (ML; BI; and MSC) are highly congruent in almost all species-level and genus-level nodes in both bootstrap support and phylogenetic relationships, however, BI and MSC analyses show only weak or moderate support for some deeper nodes among agromyzid genera (Fig. 1 and S2-S6). In nearly all analyses, the traditional subfamily Agromyzinae is found to be paraphyletic due to alternative placements of the genus *Agromyza* found in various analyses, while the subfamily Phytomyzinae is generally well-supported as monophyletic across analyses and data sets (Fig. 1 and S4-S8). One exception was found however, in which coalescent-based species trees calculated from loci scored as amino acids (TAA) place *Agromyza* inside Phytomyzinae, although with low support (Fig. S7).

Based on the strong support of our results from standard concatenated ML analyses, we here define a new subfamily 'Ophiomyiinae' for the consistently monophyletic early branching grouping containing the genera *Melanagromyza*, *Euhexomyza*, *Ophiomyia* and *Tropicomyia*. This clade is marked 'TM' on the phylogenetic tree of Fig. 1. Given this result, we subsequently redefine the Agromyzinae to include only the genera *Agromyza* and *Japanagromyza*.

The monophyly of new subfamily Ophiomyiinae is robustly supported by all datasets and phylogenetic approaches (Fig. 1 and S4-S8). All genera with at least two species in this group have been found to be monophyletic, with the exception of *Euhexomyza,* a finding that is fully supported by all datasets and phylogenetic methods (SH-aLRT / UFBoot = 100/100; PP = 1; Fig. 1 and S2-S6). In *Euhexomyza*, the two sampled species form a sister group with full support using NT123 data regardless of phylogenetic approaches but are paraphyletic in analyses of NT12 and TAA datasets (Figs. 1, S4, and S5). In the subfamily Ophiomyiinae, two included polyphagous species, *Tropicomyia theae* (Cotes) and *T. polyphyta* (Kleinschmidt) are supported as sister species, while the oligophagous species *Melanagromyza cleomae* Spencer and *M. minimoides*

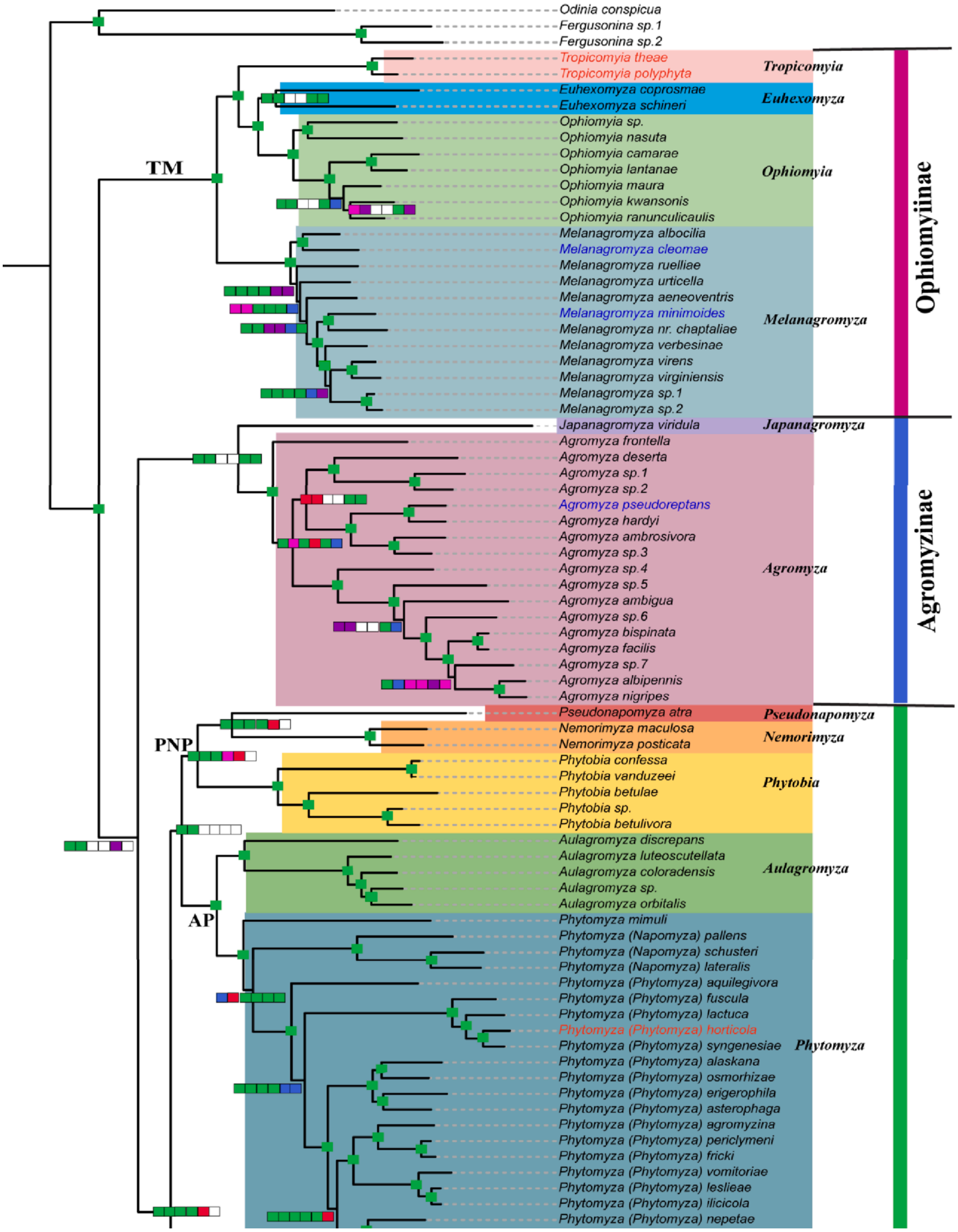

Fig. 1. Maximum likelihood tree of Agromyzidae based on 496 concatenated orthologs scored as nucleotides (NT123) implemented in IQ-TREE. Squares on nodes of the phylogenetic tree indicate bootstrap support values obtained from distinct phylogenetic methods and datasets. The squares from left to indicate results of the following tests: the first two represent SH-aLRT (Shimodaira-Hasegawa-like approximate Likelihood Ratio Test) and Ultrafast bootstrap approximation (UFBoot) values generated by ML analysis using nucleotides within three protein-coding positions retained (NT123); the middle two indicate SH-aLRT and UFBoot from ML analysis with amino acid data (TAA); the last two represent the posterior probabilities separately obtained by Bayesian inference with NT123 dataset and by ASTRAL coalescent-based tree summary of individual gene trees generated in IQ-TREE. The color coding of each square refers to specific support level categories shown at the bottom left corner. A single filled green square on the node indicates strong support (≥ 0.95) across all analysis and data types. A white box indicates that the node is not recovered by the corresponding type of dataset or phylogenetic approach. Monophyletic clades are colored to indicate generic/subgeneric representations in Agromyzidae classification. Subgenera in the genera *Phytomyza* and *Cerodontha* are shown in parentheses corresponding to specific terminal taxa on the phylogenetic tree. All *Phytomyza* species marked by purple boxes on this tree were previously in *Chromatomyia*, and then transferred to *Phytomyza* based on Winkler et al. (2009a). Polyphagous and oligophagous species are separately marked by red and blue fonts on the tree, while all remaining species are monophagous. Abbreviations used to denote deep level monophyletic clades represent as follows: TM (Ophiomyiinae) = *Tropicomyia* + *Euhexomyza* + *Ophiomyia* + *Melanagromyza*; PNP = *Pseudonapomyza* + *Nemorimyza* + *Phytobia*; AP = *Aulagromyza* + *Phytomyza*; CL = *Cerodontha* + *Amauromyza* + *Calycomyza* + *Phytoliriomyza* + *Selachops* + *Metopomyza* + *Liriomyza*; CAL = *Calycomyza* + *Phytoliriomyza* + *Selachops* + *Metopomyza* + *Liriomyza*.

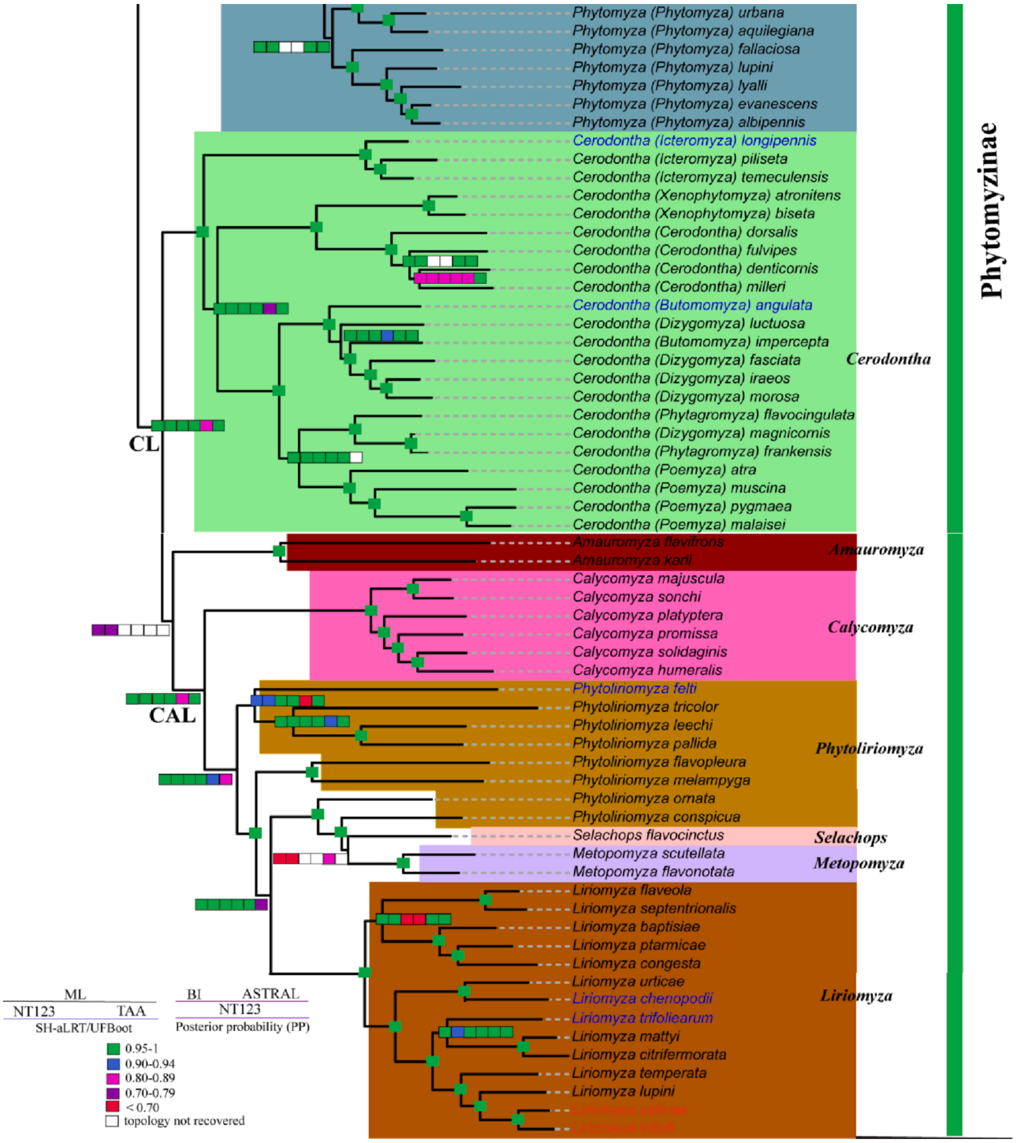

Fig. 1. (*continued*).

Spencer are not closest relatives within *Melanagromyza* (Fig. 1).

Among analyses using various datasets and methods, the only major difference found is among support for placement of the single sampled species of the genus *Japanagromyza* (Fig. 1 and S4-S5). Specifically, *Japanagromyza* is placed as sister to *Agromyza* with strong support in the ML tree generated using NT123 dataset (Fig. 1), while it is placed as a sister group to the new subfamily Ophiomyiinae using the TAA and NT1 + 2 datasets in ML analysis with strong support (Figs. S4, S5).

In Phytomyzinae, 10 of the 12 genera sampled for at least two species are confirmed as monophyletic except for *Phytoliriomyza*, in which sampled taxa were recovered as paraphyletic or polyphyletic on the phylogeny (Fig. 1 and S4-S8). To summarize support for deeper nodes in this subfamily, we divided the group into four large clades: 'PNP', 'AP', 'CL', and 'CAL' (Fig. 1). The 'PNP' clade consisting of *Pseudonapomyza*, *Nemorimyza*, and *Phytobia*, is fully supported as monophyletic in ML analysis of NT123 (Fig. 1). In this clade, *Pseudonapomyza atra* (Meigen) is placed as sister to the genus *Nemorimyza* with strong support in ML trees generated using NT123 and TAA datasets (Fig. 1). Nevertheless, this grouping is only weakly or moderately supported using Bayesian inference (PP = 0.55; Fig. S6) and in the coalescent-based species tree produced from gene trees with the third positions removed (PP = 0.75; Fig. S8), but not supported in coalescent based trees with all nucleotide sites retained (Fig. S7).

The 'AP' clade containing two genera, *Aulagromyza* and *Phytomyza*, is strongly supported to be monophyletic across all data and analysis types (Fig. 1 and S4-S8). Within the diverse genus *Phytomyza*, previous *Chromatomyia* marked in the purple boxes in Fig. 1 is again recovered as non-monophyletic with species recovered in three independent lineages on the phylogenetic tree. *Phytomyza mimuli* Spencer is strongly supported as sister to all remaining *Phytomyza* in all phylogenetic analyses (Fig. 1 and S4-S8). Species from the subgenus *Napomyza* form a monophyletic subclade with strong support, which is placed among the earliest diverging branches of the genus *Phytomyza* with strong branch bootstrap support values (Fig. 1). The 'AP' clade is placed as sister to the 'PNP' clade with strong branch support (SH-aLRT / UFBoot = 99.9/99) in ML analysis of the NT123 dataset, but not recovered by ML analysis with amino acids and other phylogenetic methods (Fig. 1 and S4-S8). On the contrary, the 'AP' clade is placed as sister to the 'CL' clade with robust support in ML analysis using the TAA dataset (SH-aLRT / UFBoot = 99.4/98; Fig. S5).

The 'CL' clade in Phytomyzinae, comprising the genera *Cerodontha*, *Amauromyza*, *Calycomyza*, *Phytoliriomyza*, *Selachops*, *Metopomyza*, and *Liriomyza*, forms a monophyletic group with full support in both ML and coalescent analyses (SH-aLRT / UFBoot = 100/100; PP = 1; Fig. 1).

Within *Cerodontha*, we included samples from seven subgenera including *Icteromyza*, *Cerodontha*, *Xenophytomyza*, *Butomomyza*, *Dizygomyza*, *Phytagromyza*, and *Poemyza*. The subgenus *Dizygomyza* is found to be paraphyletic, and these placements break up the monophyly of *Butomomyza* and *Phytagromyza* (Fig. 1). Each of the remaining named subgenera we sampled, including *Icteromyza*, *Xenophytomyza*, *Cerodontha* and *Poemyza*, separately form a monophyletic group, and subgenus *Poemyza* is nested into the polyphyletic clade involving *Butomomyza*, *Dizygomyza*, and *Phytagromyza* (Fig. 1).

The genus *Amauromyza* is a sister group to the 'CAL' clade with moderate support in ML analyses using NT123 and NT12 datasets (Figs. 1 and S4), but this grouping is not recovered in BI and coalescent species trees (Fig. 1 and S6-S8). A monophyletic *Calycomyza* is sister to a clade comprising genera *Phytoliriomyza*, *Selachops*, *Metopomyza* and

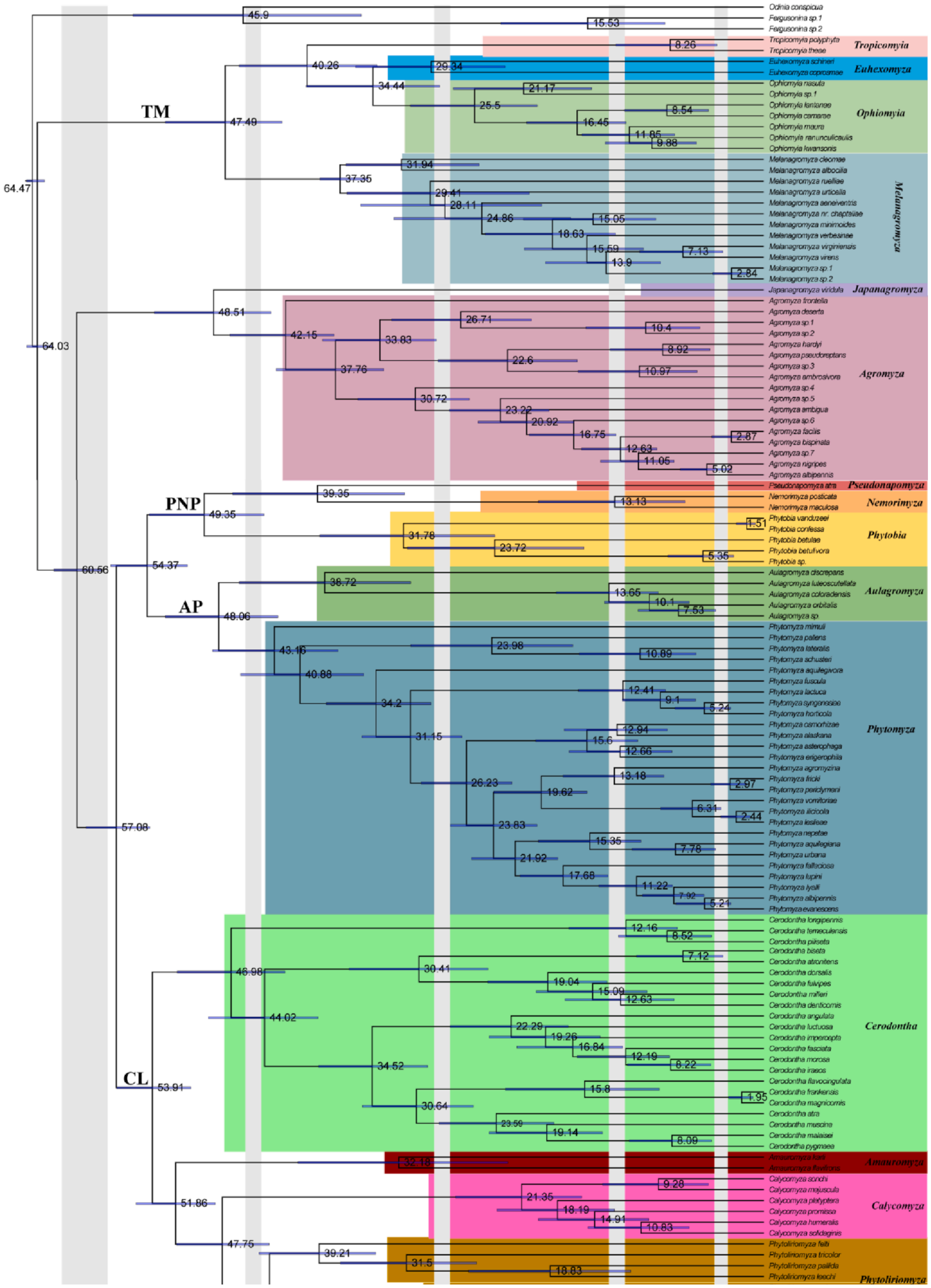

Fig. 2. Divergence time estimates using ML tree generated from NT123 dataset conducted in MCMCTREE. Values on internal nodes represent the mean divergence time in million years ago (Ma), and the node bar indicates the highest posterior density (HPD) 95% confidence interval (CI) of the average age. Geologic time scale (Ma) is shown at the bottom of time tree. Abbreviations in geologic time scale indicates as follows: Plei. = Pleistocene; Ho. = Holocene; Quat. = Quaternary. Vertical gray bars from left to right on the time tree sequentially indicate the middle of each epoch: Paleocene, Eocene, Oligocene, Miocene, and Pliocene.

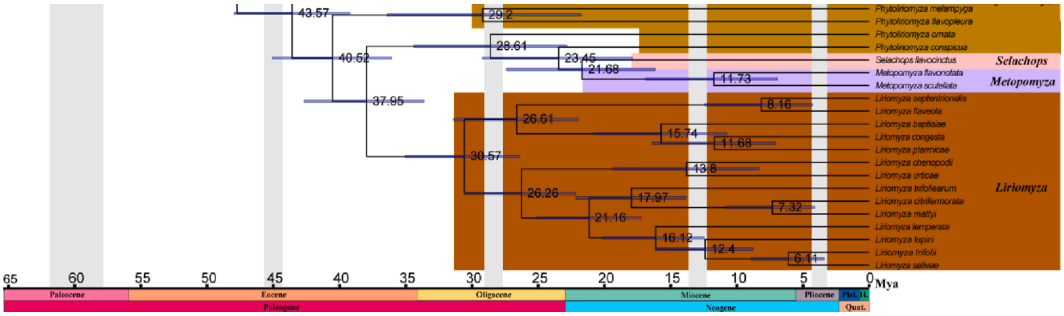

Fig. 2. (*continued*).

Liriomyza, with strong support in ML and coalescent analyses, but only moderately supported by BI (Fig. 1 and S6-S8). Several species of *Phytoliriomyza* are are also placed with the genera *Selachops*, *Metopomyza*, and *Liriomyza* in all ML analyses, making *Phytoliriomyza* as currently defined paraphyletic (Fig. 1 and S2-S3). The genus *Selachops* is grouped with *Metopomyza* in ML and BI trees with weak or moderate support (Figs. 1, S4, and S6). In the large genus *Liriomyza*, nearly all the species-level relationships are strongly supported by all phylogenetic analyses of the various datasets. Two polyphagous species, *L. sativae* and *L. trifolii*, form a sister group with strong support in all analyses, nevertheless, two oligophagous species *L. chenopodii* (Watt) and *L. trifoliearum* Spencer are not closely related on the phylogenetic trees (Fig. 1 and S4-S8).

3.3. Likelihood mapping

Two major differences are found in our phylogenetic reconstructions of higher-level relationships: one is the placement of the genus *Japanagromyza*, and the other is the position of the ‘AP’ clade (Fig. S1). To evaluate the phylogenetic information content of our nucleotide and amino acid alignments for these conflicting relationships, we conducted likelihood mapping analysis. Based on this analysis, we have found that the quartets produced from both concatenated nucleotide (NT123) and amino acid (TAA) alignments are relatively decisive as very few quartets provide ambiguous information. In other words, there are few or no points falling into the center or middle edges of the triangle diagram, suggesting that both molecular data sets contain decisive phylogenetic signals. Of the quartets sampled, 98.9% support placement of the genus *Japanagromyza* together with genus *Agromyza*, while 1.0% support the grouping of *Japangaromyza* as sister to subfamily Phytomyzinae using nucleotide alignments (Fig. S2a). In contrast, for amino acids, we find that 97.8% of quartets support genus *Japanagromyza* as a sister to the ‘TM’ clade (Ophiomyiinae), whereas only 1.4% support a sister relationship between *Japanagromyza* and *Agromyza* (Fig. S2b).

For the placement of ‘AP’ clade, 42.4% of quartets support a sister group between ‘AP’ and ‘PNP’ clades with NT123 (Fig. S2c), and this grouping is also strongly supported in the ML tree with this nucleotide data (SH-aLRT / UFBoot = 99.9/99; Fig. 1). Conversely, 36.1% of quartets analyzed with TAA data support the ‘AP’ + ‘PNP’ grouping (Fig. S2d), but this is not reflected in the ML tree analysis as the ‘AP’ clade is sister to the ‘CL’ clade with strong support using the TAA dataset (SH-aLRT / UFBoot = 99.4/98; Fig. S2d). Nevertheless, 31.4% of quartets support the sister group between ‘AP’ and ‘CL’ clades (Fig. S2d). Thus, nucleotides are more decisive in this area of the tree than amino acids, with greater uncertainty found for quartets involving the resolution of the ‘AP’, ‘CL’, and ‘PNP’ clades (Fig. S2c and S2d).

3.4. Divergence time estimation

Six independent MCMC runs yielded highly congruent posterior distributions of clade ages (Fig. S3). Mean node age and HPD 95% CI are provided on each node of the dated phylogeny (Fig. 2). Leaf-mining flies are likely to have originated in the early Paleocene (64.47 Ma). This finding is consistent with the fossil record for both agromyzid leaf mines (Winkler et al. 2010) and flies (Grimaldi 2018). Sampled crown group members of the subfamily Ophiomyiinae diversified about 47.49 Ma (HPD 95% CI: 37.14–49.25; Table S3), while subfamily Phytomyzinae diversified at about 57.08 Ma (HPD 95% CI: 54.13–59.77). Three large clades ‘CL’, ‘PNP’, and ‘AP’ in this subfamily diverged at about 53.91, 49.35, and 48.06 Ma, respectively (Fig. 2; Table S3). The large genera *Phytomyza* and *Liriomyza* are, separately, dated to approximately 48.06 Ma and 37.95 Ma (Fig. 2).

4. Discussion

4.1. Phylogenetic relationships of leaf-mining flies

Our phylogenomic study of leaf-mining flies provides new evidence on the earliest divergences among lineages of this entirely phytophagous family. Most significantly, the results of our study do not support the division of the family into just two reciprocally monophyletic subfamilies under all analytical conditions (Fig. 1). This contradicts the currently accepted classification system (Fallén, 1823a; 1823b) as well as previous results from both morphological (Dempewolf, 2001; 2005) and molecular studies (Scheffer et al., 2007). The large size and scope of phylogenomic analyses are currently revising phylogenetic classifications for many groups (Buenaventura et al., 2020; 2021; Yan et al., 2021a; 2021b), especially in those where rapid radiations have generated considerable phylogenetic uncertainty (Bayless et al. 2021). Our strongly supported trees based on hundreds of orthologous nuclear genomic loci identify support for three major agromyzid clades (Fig. 1). Based on these data, we support reclassification of the family to include three subfamilies. Thus, we propose a new subfamily Ophiomyiinae for the earliest lineage in the tree corresponding to our ‘TM’ clade including *Melanagromyza*, *Euhexomyza*, *Ophiomyia* and *Tropicomyia*. We also hereby designate *Ophiomyia* Brazhnikov as the type genus of Ophiomyiinae and have registered this taxon as a new family-group name in Zoobank. This removes the lineage previously named the ‘*Ophiomyia* genus group’ from the Agromyzinae, leaving the Agromyzinae restricted to only *Agromyza* and *Japanagromyza*. All other genera in the family Agromyzidae remain in the subfamily Phytomyzinae (Fig. 1).

Bayesian and coalescent-based analyses reveal some discordant signals for deep relationships in our data, and these methods fail to strongly support the earliest divergences in the family unlike those shown in our strongest findings from maximum likelihood (Figs. S6-S8). However, these methods may be particularly sensitive to contradictory signal contributed by individual loci or to uneven sampling of genes or taxa (Simion et al., 2020; Young and Gillung, 2020), and so gene tree discordance may play a role in obscuring signal in these instances. Further sampling and exploration of the differential contribution of phylogenetic information content in these data will be critical to identify and resolve remaining uncertainty among the leaf-mining flies.

4.1.1. The origin and diversification of Agromyzidae

Our Bayesian relaxed clock divergence times analysis was given a firm root node maximum age of 65 Ma based on the complete absence of any schizophoran family in the Cretaceous fossil record (Grimaldi 2018). This constraint was also used in similar studies that estimated divergences for higher-level fly clades in the Cyclorrhapha (Wiegmann et al. 2011; Junqueira et al. 2016). Analyses carried out lacking a Cretaceous-Paleogene boundary for Schizophora routinely yield much deeper ages for Schizophoran families (approx. 120 Ma), including an older age for the origin of Agromyzidae (81.9 Ma, Song et al, 2022). These older estimates reflect a current major disparity in our understanding of fly evolution and these estimates are heavily influenced by high rates of species diversification and genetic divergence in Schizophora, as well as the lack of necessary contextualizing evidence from the fossil record. A broader sample of schizophoran clades would be required to fully evaluate the impact of these issues on root node estimates for the Agromyzidae.

4.1.2. Ophiomyiinae and Agromyzinae

Our phylogenomic data thus do not support the traditional divisions of the Agromyzidae into two reciprocally monophyletic subfamilies, a finding which also confirms von Tschirnhaus's (1971) speculation that the lineages of leaf-mining flies cannot be forced into two distinct halves. Our phylogenomic data instead supports the recognition of three major lineages within the group historically interpreted as the subfamily Agromyzinae, which are also supported by morphological characters (Lonsdale, 2021). The first of these, our new subfamily Ophiomyiinae ('TM clade'; Fig. 1), was strongly supported as monophyletic.

This clade has been called the '*Ophiomyia* genus group' in morphological assessments, and includes the genera *Ophiomyia*, *Melanagromyza*, *Euhexomyza*, plus *Tropicomyia* (Eiseman, 2021; Lonsdale, 2021). As a general diagnosis of Ophiomyiinae we list the following features shared by members of this group: (1) large and bulging lunule with the insertion of the antenna below the midpoint of the head, and (2) subepandrial sclerite divided into two well-sclerotized plates (Lonsdale, 2014; 2021). Most species in Ophiomyiinae are internal borers in stems or seeds, except for *Tropicomyia,* whose species produce irregular mines in the upper surface of leaves (Spencer, 1973; Lonsdale, 2021).

Agromyzinae now consists solely of *Agromyza* and *Japanagromyza*, which typically share the features of prescutellar acrostichal setae and the mesophallus located basal to the distiphallus (Lonsdale, 2021). This redefined subfamily is placed as the sister group to Phytomyzinae with strong support in our ML analyses using nucleotides (Fig. 1). Morphological data is conflicting on the position of *Japanagromyza.* A stridulatory file is present in *Agromyza* adults of both sexes, but is lacking in *Japanagromyza* (Von Tschirnhaus, 1971; Lonsdale, 2021), but this feature is convergently derived in *Liriomyza*. *Agromyza* species possess distinctive saw-like stridulatory file on the lateral margin of tergite 2 in males and females, while in *Liriomyza*, the anterolateral stridulatory organ is located in the abdominal membrane only for some male adults (Lonsdale, 2021). Previously, *Japanagromyza* has been treated as an intermediate between *Ophiomyia* and *Agromyza* due to a combination of morphological characteristics shared among them (von Tschirnhaus, 1991; Sasakawa, 2010). Here, we found alternative placements of this genus near the larger groups containing *Ophiomyia* or *Agromyza* (Figs. 1, S4, and S5). Scheffer et al. (2007) also found *Japanagromyza* as sister to *Agromyza*, while Lonsdale (2013) postulated instead that *Japanagromyza* may itself be non-monophyletic because of their genitalic diversity (Dempewolf 2001, 2005). Thus, the exact placement of *Japanagromyza* remains uncertain, but as the group is consistently placed sister to *Agromyza* in all nucleotide-based analyses (Figs. 1, S6, and S7), we have decided to retain its position within the Agromyzinae. Notably, we were only able to sample a single specimen of the most common American species, *J. viridula* (Coquillett) and so further confirmation must await future investigations that include increased sampling.

4.1.3. Phytomyzinae

All remaining Agromyzidae comprise the monophyletic subfamily Phytomyzinae, a diverse and heterogeneous set of lineages (Lonsdale, 2021). We found support for three main lineages within Phytomyzinae that we refer to here as the 'PNP', 'AP', and 'CL' clades (Fig. 1). Most genera have been found to be monophyletic including *Nemorimyza*, *Phytobia*, Aulagromyza, *Phytomyza*, *Cerodontha*, *Amauromyza*, *Calycomyza*, *Metopomyza*, and *Liriomyza*, with full bootstrap and branch support (Fig. 1 and S2-S6). These findings confirm previous studies (Dempewolf, 2005; Scheffer et al., 2007; Xuan et al. 2022), but also provide new positions of some genera with strong support, such as for sampled members of *Phytoliriomyza* and *Aulagromyza* (Fig. 1).

Within Phytomyzinae, our 'PNP' clade corresponds approximately to the '*Phytobia* group' recognized by Spencer (1990) for *Amauromyza*, *Nemorimyza* and *Phytobia*. While our trees do not support placement of *Amauromyza* in this clade, the supported position of *Amauromyza* is weak in nearly all analyses (Fig. 1 and S4-S8) and so an association with this group is still possible. Previous authors considered the *Phytobia* group to be the earliest diverging lineage of all Agromyzidae since they share larger body sizes and seemingly plesiomorphic feeding modes, e.g., stem and cambium feeding (Nowakowski 1964; Spencer, 1990), but results based on molecular data, both here and in Scheffer et al. (2007), have not supported this hypothesis.

Spencer (1990) also recognized a '*Napomyza* group', corresponding approximately to our monophyletic 'AP' clade. Spencer's grouping consisted of *Aulagromyza*, *Phytomyza*, *Napomyza*, *Chromatomyia*, *Gymnophytomyza*, and *Pseudonapomyza*. *Napomyza* and *Ptochomyza* are now recognized as subgenera of the diverse genus *Phytomyza* based on findings of a large multigene phylogenetic study, and *Chromatomyia* was supported as an artificial assemblage of species that was synonymized with *Phytomyza* (Winkler et al., 2009a,b). While some still support retention of *Chromatomyia* (see von Tschirnhaus (2021b), for example), extensive morphological and life history data also supports *Chromatomyia* as a synonym of *Phytomyza* summarized in Lonsdale and Eiseman (2021). Results of the present study (Fig. 1 and S2-S6) further strongly support the synonymy of *Chromatomyia*. The genera *Aulagromyza* and *Pseudonapomyza* are superficially similar, but can be differentiated by characters of wing venation (Lonsdale, 2021). In our study, we found these two genera to consistently represent distinct lineages. Sampled specimens of *Aulagromyza* (notably excluding members of the possibly unrelated *A. populicola* group) comprise a lineage that is sister group to *Phytomyza*, supported by high bootstrap values and posterior probabilities (Fig. 1). *Pseudonapomyza*, represented by only one species of the *A. atra* group (possibly unrelated to the *P. acanthacearum* group), is supported as sister group to *Nemorimyza* in the 'PNP' group (Fig. 1 and S4-S8). However, genera *Aulagromyza* and *Pseudonapomyza* are strongly suspected to be non-monophyletic (Spencer, 1990; Dempewolf, 2001; Scheffer et al., 2007; Zlobin, 2002; 2007; Winkler et al., 2009a), thus, further study is needed to investigate these placements with additional sampling.

All nine remaining genera in Phytomyzinae form a large monophyletic lineage corresponding to Spencer's '*Phytoliriomyza* group' (Spencer, 1990), here labeled as the 'CL' clade in Fig. 1. This '*Phytoliriomyza* group' includes *Cerodontha*, *Calycomyza*, *Phytoliriomyza*, *Metopomyza*, *Selachops*, *Liriomyza*, *Haplopeodes*, *Xeniomyza*, and *Pseudoliriomyza* (Spencer, 1990), but the last three genera were not available for this study. The phylogenetic position and relationships of *Cerodontha* have not been convincingly resolved in previous studies (Dempewolf, 2001; 2005; Scheffer et al., 2007). Here, we sampled 22 *Cerodontha* species across all seven subgenera and found strong support for the monophyly of the genus, for non-monophyly of some subgenera, and for a basal position of *Cerodontha* in the 'CL' clade (Fig. 1 and S2-S6).

Specifically within *Cerodontha*, only four of the seven subgenera *Icteromyza*, *Xenophytomyza*, *Cerodontha* and *Poemyza*, are found to be monophyletic with full branch support in any phylogenetic analysis, and the subgenus *Icteromyza* is supported as basal in the genus (Fig. 1 and S4-

S8). The subgenera *Cerodontha* and *Xenophytomyza* are strongly supported as sister groups, and share an angulate first flagellomere and loss of the lateral scutellar seta (Lonsdale, 2021). In our study, the two sampled species of the subgenus *Butomomyza* were not recovered together within a large clade consisting of most sampled *Dizygomyza* species. *Poemyza* was recovered as monophyletic, and sister-group to a lineage consisting of a *Phytagromyza* rendered paraphyletic by one *Dizygomyza* species (Fig. 1). The non-monophyly of the subgenera *Butomomyza*, *Dizygomyza* and *Phytagromyza* is both well-supported and quite unexpected as species apparently defined by unusual morphological characters are not recovered together. Overall, *Cerodontha* is a large, complex genus in need of much additional analysis, and our results suggest that many morphological characters potentially useful for defining subclades are in need of re-evaluation and broader phylogenetic sampling.

Another monophyletic lineage in our 'CL' group is labeled here as the 'CAL' clade, which includes the genera *Calycomyza*, *Phytoliriomyza*, *Metopomyza*, *Selachops*, *Liriomyza* as well as *Haplopeodes*, which was not available for sequencing in the present study (Fig. 1). A synapomorphy for this group is the shared presence of tubercle-like setae on the posterodistal margin of the epandrium and surstylus (Lonsdale, 2021). Among the genera sampled here, only *Phytoliriomyza* was found to be paraphyletic or polyphyletic on our trees. Sampled species were split into four independent lineages on the phylogeny (Fig. 1 and S4-S8), reinforcing the diverse interrelationships of this group reported in previous studies (Dempewolf, 2001; 2005; Scheffer et al., 2007). Lonsdale (2021) also suggested that *Phytoliriomyza* should be divided into at least two groups to maintain the monophyly of *Metopomyza* and *Liriomyza*, both of which show affinities for different *Phytoliriomyza* taxa. Our sampling in this area of the phylogeny also included the enigmatic Palearctic species *Selachops flavocinctus* Wahlberg, which was recovered as sister-taxon to *Metopomyza*, a relationship previously suggested by morphological evidence (Dempewolf, 2001; 2005).

5. Conclusions

Anchored hybrid enrichment provides access to hundreds of loci across the genome that are shown to be useful for reconstructing the phylogenetic relationships from subfamily to species level. Despite their small body size, AHE gene capture is a valuable method for Agromyzidae phylogenetics using probe designs that recover conserved dipteran genes. Our maximum likelihood phylogenetic tree estimates based on nucleotides provide new insights into the deepest splits, prompting a revised classification system of Agromyzidae. Model-based analysis of agromyzid divergence times within the phylogeny provide a temporal context for the diversification of the family and a revised timeframe for placing species radiations within major genera such as *Phytomyza*, *Liriomyza*, *Cerodontha* and *Agromyza*.

CRediT authorship contribution statement

Jing-Li Xuan: Investigation, Formal analysis, Data curation, Conceptualization, Visualization, Writing – original draft, Writing – review & editing. **Sonja J. Scheffer:** Conceptualization, Resources, Writing – review & editing, Supervision. **Matt Lewis:** Investigation, Data curation. **Brian K. Cassel:** Investigation, Data curation. **Wan-Xue Liu:** Resources, Supervision. **Brian M. Wiegmann:** Conceptualization, Resources, Writing – review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

I have shared the link to my data at the attach file step

Acknowledgments

This research was funded by the US National Science Foundation (DEB-1542395 to Sonja J. Scheffer and Brian M. Wiegmann, and DEB-2030345 to Brian M. Wiegmann) and also supported by China Scholarship Council (No. 201807990002). We thank Drs. Michael von Tschirnhaus, Owen Lonsdale, Patrick Häfliger, Ghislaine Cortat, and Charley Eiseman for generously donating their valuable specimens, especially Dr. Owen Lonsdale for his careful review of an early draft of the manuscript and for providing valuable and insightful comments and suggestions. We also sincerely thank Dr. John Soghigian from the University of Calgary and Dr. Cristian Beza-Beza from North Carolina State University for their assistance in data analyses and comparative methods. In addition, mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the USDA. USDA is an equal opportunity provider and employer.

Data Availability Statement

The sequence raw reads (SRA) that support the findings of this study are publicly available in the National Center for Biotechnology Information (NCBI) SRA database, and the links of SRA are: http://www.ncbi.nlm.nih.gov/bioproject/859188 and http://www.ncbi.nlm.nih.gov/bioproject/858472. The according BioProject accession numbers PRJNA859188 and PRJNA858472. Nucleotide sequence alignments, maximum likelihood tree, and concatenated sequence locus-level partitions, are deposited on Zenodo: https://doi.org/10.5281/zenodo.7193143

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2023.107778.

References

Aberer, A.J., Kobert, K., Stamatakis, A., 2014. ExaBayes: massively parallel Bayesian tree inference for the whole-genome era. Biol. Evol. 31 (10), 2553–2556.
Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215 (3), 403–410.
Anisimova, M. 2019. Evolutionary genomics: statistical and computational methods. Humana.
Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19 (5), 455–477.
Bayless, K.M., Trautwein, M.D., Meusemann, K., Shin, S., Petersen, M., Donath, A., Podsiadlowski, L., Mayer, C., Niehuis, O., Peters, R.S., Meier, R., 2021. Beyond Drosophila: resolving the rapid radiation of schizophoran flies with phylotranscriptomics. BMC Biol. 19 (1), 1–17.
Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinf. 30 (15), 2114–2120.
Buenaventura, E., Szpila, K., Cassel, B.K., Wiegmann, B.M., Pape, T., 2020. Anchored hybrid enrichment challenges the traditional classification of flesh flies (Diptera: Sarcophagidae). Syst. Entomol. 45 (2), 281–301.
Buenaventura, E., Lloyd, M.W., Perilla Lopez, J.M., González, V.L., Thomas-Cabianca, A., Dikow, T., 2021. Protein-encoding ultraconserved elements provide a new phylogenomic perspective of Oestroidea flies (Diptera: Calyptratae). Syst. Entomol. 46 (1), 5–27.
Civelek, H.S., 2003. Checklist of Agromyzidae (Diptera) of Turkey, with a new record. Phytoparasitica 31 (2), 132–138.
Dempewolf, M., 2005. Dipteran leaf miners. Biology, ecology and evolution of gall-inducing arthropods Volume 1 and 2, 407–429.
Dempewolf, M., 2001. Larval morphologie und phylogenie der Agromyzidae (Diptera). PhD thesis, University of Bielefeld, Bielefeld, Germany.
dos Reis, M.D., Yang, Z., 2011. Approximate likelihood calculation on a phylogeny for Bayesian estimation of divergence times. Mol. Biol. Evol. 28 (7), 2161–2172.
dos Reis, M.D., Zhu, T., Yang, Z., 2014. The impact of the rate prior on Bayesian estimation of divergence times with multiple loci. Syst. Biol. 63 (4), 555–565.
Drummond, D.A., Raval, A., Wilke, C.O., 2006. A single determinant dominates the rate of yeast protein evolution. Mol. Biol. Evol. 23 (2), 327–337.

Eiseman, C.S., Lonsdale, O., Feldman, T.S., 2019. Nine new species of Agromyzidae from North Carolina, USA, with new host and distribution records for additional species. Zootaxa 4571 (3), 3–27.

Eiseman C.S., 2021. Leafminers of North America (2nd Edn.). Self-published e-book. http://charleyeiseman.com/leafminers/.

Evangelista, D.A., Wipfler, B., Béthoux, O., Donath, A., Fujita, M., Kohli, M.K., Legendre, F., Liu, S., Machida, R., Misof, B., Peters, R.S., 2019. An integrative phylogenomic approach illuminates the evolutionary history of cockroaches and termites (Blattodea). Proc. R. Soc. B. 286, 1–9.

Evenhuis, N.L., 1994. Agromyzidae. In: Catalog of the fossil flies of the world (Insecta: Diptera) website. Available at: http://hbs.bishopmuseum.org/fossilcat/Agromyzidae.html. 20 Jan. 2014.

Fallén, C.F., 1823a. Agromyzides Sveciae. Berlingianis.

Fallén, C.F., 1823b. Phytomyzides et Ochtidiae Sveciae. Berlingianis.

Flouri, T., 2018. MCMCtree summarizer. Retrieved from https://github.com/xflou ris/summa rizer.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nat. Biotechnol. 29 (7), 644–652.

Grimaldi, D., 2018. Basal Cyclorrhapha in amber from the Cretaceous and Tertiary (Insect: Diptera), and their relationships: Brachycera in Cretaceous amber Part IX. Bull. Am. Mus. Nat. Hist. 423, 1–97.

Hering, E.M., 1951. Biology of the leaf miners. Springer Science, Business Media.

Hoang, D.T., Chernomor, O., Von Haeseler, A., Minh, B.Q., Vinh, L.S., 2018. UFBoot2: improving the ultrafast bootstrap approximation. Mol. Biol. Evol. 35 (2), 518–522.

Junqueira, A.C.M., Alzaredo-Espin, A.M., Paulo, D.F., Marinho, M.A.T., Tomsho, L.P., Drautz-Moses, D.I., Purbojati, R.W., Ratan, A., Schuster, S.C., 2016. Large scale mitogenomics enables insights into Schizophora (Diptera) radiation and population diversity. Sci. Rep. 6, 21762.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K., von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14 (6), 587–589.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30 (4), 772–780.

Kück, P., Longo, G.C., 2014. FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. Front. Zool. 11 (1), 1–8.

Kück, P., Meusemann, K., Dambach, J., Thormann, B., von Reumont, B.M., Wägele, J.W., Misof, B., 2010. Parametric and non-parametric masking of randomness in sequence alignments can be improved and leads to better resolved trees. Front. Zool. 7 (1), 1–12.

Lanfear, R., Calcott, B., Kainer, D., Mayer, C., Stamatakis, A., 2014. Selecting optimal partitioning schemes for phylogenomic datasets. BMC Evol. Biol. 14 (1), 1–14.

Lemmon, A.R., Emme, S.A., Lemmon, E.M., 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. Syst. Biol. 61 (5), 727–744.

Lonsdale, O., 2011. The *Liriomyza* (Agromyzidae: Schizophora: Diptera) of California. Zootaxa 2850 (1), 1–123.

Lonsdale, O., 2013. *Japanagromyza* Sasakawa (Diptera: Agromyzidae) of Africa. Zootaxa 3709, 445–460.

Lonsdale, O., 2014. Redefinition and synonymy of genera in the *Ophiomyia* genus group, with the description of *Euhexomyza* new genus (Diptera: Agromyzidae). Can. Entomol. 146 (5), 481–513.

Lonsdale, O., 2021. Manual of North American Agromyzidae (Diptera, Schizophora), with revision of the fauna of the "Delmarva" states. ZooKeys 1051, 1–481.

Lonsdale, O., Eiseman, C.S., 2021. What makes a 'good' genus? Reconsideration of *Chromatomyia* Hardy (Diptera: Agromyzidae). Dipterists Digest 28, 221–249.

McAlpine, J.F., 1989. Phylogeny and classification of the Muscomorpha. Manual of Nearctic Diptera 3, 1397–1518.

McAlpine, J.F., Peterson, B.V., Shewell, G.E., Teskey, H.J., Vockeroth, J.R., Wood, D.M., 1987. Manual of Nearctic Diptera 2.

Melander, A.L., Cockerell, T.D.A., Cockerell, W.P., Wheeler, W.M., Rohwer, S.A., Lacoe, R.D., Expeditions, F., 1949. A report on some Miocene Diptera from Florissant, Colorado. American Museum novitates; no. 1407.

Meyer, M., Kircher, M., 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harbor Protocols 2010 (6), 1–10.

Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In 2010 gateway computing environments workshop 2010, 1-8.

Minh, B.Q., Nguyen, M.A.T., von Haeseler, A., 2013. Ultrafast approximation for phylogenetic bootstrap. Mol. Biol. Evol. 30 (5), 1188–1195.

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., Von Haeseler, A., Lanfear, R., 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 37 (5), 1530–1534.

Misof, B., Liu, S., Meusemann, K., Peters, R.S., Donath, A., Mayer, C., Frandsen, P.B., Ware, J., Flouri, T., Beutel, R.G., Niehuis, O., 2014. Phylogenomics resolves the timing and pattern of insect evolution. Sci. 346 (6210), 763–767.

Misof, B., Misof, K., 2009. A Monte Carlo approach successfully identifies randomness in multiple sequence alignments: a more objective means of data exclusion. Syst. Biol. 58 (1), 21–34.

Mitchell A.A., 2013. The Fossil Insect Database https://fossilinsectdatabase.co.uk/. EDNA last update: 10th June 2022.

Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32 (1), 268–274.

Nowakowski, J.T., 1962. Introduction to a systematic revision of the family Agromyzidae (Diptera) with some remarks on host plant selection by these flies. Polska Akad. Nauk. Inst. Zool. Ann. Zool. 20, 67–183.

Nowakowski, J.T., 1964. Studien über Minierfliegen (Diptera: Agromyzidae) 9. Revision der Artengruppe *Agromyza* reptans Fall. A. rufipesMeig. Dtsch. Entomol. Z. 11, 175–213.

Paradis, E., 2013. Molecular dating of phylogenies by likelihood methods: a comparison of models and a new information criterion. Mol. Phylogenet. Evol. 67 (2), 436–444.

Pauli, T., Burt, T.O., Meusemann, K., Bayless, K., Donath, A., Podsiadlowski, L.A., Mayer, C., Kozlov, A., Vasilikopoulos, A., Liu, S., Zhou, X., 2018. New data, same story: Phylogenomics does not support Syrphoidea (Diptera: Syrphidae, Pipunculidae). Syst. Entomol. 43, 447–459.

Petersen, M., Meusemann, K., Donath, A., Dowling, D., Liu, S., Peters, R.S., Podsiadlowski, L., Vasilikopoulos, A., Zhou, X., Misof, B., Niehuis, O., 2017. Orthograph: a versatile tool for mapping coding nucleotide sequences to clusters of orthologous genes. BMC Bioinf. 18 (1), 1–10.

Rannala, B., Yang, Z., 2007. Inferring speciation times under an episodic molecular clock. Syst. Biol. 56 (3), 453–466.

RStudio Team, 2020. RStudio: Integrated Development for R. RStudio, PBC, Boston, MA, http://www.rstudio.com/.

Sasakawa, M., 2010. Agromyzidae (Diptera) from Kermadec Islands. New Zealand. N. Z. Entomol. 33 (1), 14–16.

Scheffer, S.J., Winkler, I.S., Wiegmann, B.M., 2007. Phylogenetic relationships within the leaf-mining flies (Diptera: Agromyzidae) inferred from sequence data from multiple genes. Mol. Phylogenet. Evol. 42 (3), 756–775.

Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16 (8), 1114–1116.

Simion, P., Delsuc, F., Philippe, H., 2020. To what extent current limits of phylogenomics can be overcome? Phylogenetics in the Genomic Era 2 (1), 1–34.

Spencer, K.A., 1966. A revision of European species of the genera *Melanagromyza* Hendel and *Hexomyza* Enderlein, with a supplement on the genus *Ophiomyia* Braschnikov (Diptera: Agromyzidae). Beiträge zur Entomologie = Contributions to Entomology 16, 3–60.

Spencer, K.A., 1973. Stem-Borers, Gall-Causers and Leaf-Miners on Trees. In: Agromyzidae (Diptera) of Economic Importance. Ser. entomol. 9, 298–316.

Spencer, K.A., Steyskal, G.C., 1986. Manual of the Agromyzidae (Diptera) of the United States. U.S. Department of Agriculture, Agriculture Handbook No. 638.

Spencer K.A., 1990. Host Specialization in the World Agromyzidae (Diptera). Kluwer Academic Publishers: London.

Strimmer, K., Von Haeseler, A., 1997. Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. Proc. Natl. Acad. Sci. 94 (13), 6815–6819.

Suyama, M., Torrents, D., Bork, P., 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 34, 609–612.

Tavaré, S., 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. Lect. math. life sci. 17 (2), 57–86.

Theobald, N., 1937. Les insectes fossiles des terrains oligocènes de France Nancy. France: George Thomas.

von Tschirnhaus, M.V., 1971. Unbekannte Stridulationsorgane bei Dipteren und ihre Bedeutung fur Taxonomie und Phylogenetik der Agromyziden. (Diptera: Agromyzidae et Chamaemyiidae). Beitr. Entomol. 21, 551–579.

von Tschirnhaus, M.V., 1991. New results on the ecology, morphology, and systematics of Agromyzidae (Diptera). Proc. Second Int. Congr. of Dipterology 285–313.

von Tschirnhaus, M., 2021. Morphological arguments for the retainment of three long time established genera in Agromyzidae (Diptera): *Chromatomyia* Hardy, *Napomyza* Westwood, and *Ptochomyza* Hering. Dipterist Digest 28, 105–117.

von Tschirnhaus, M., 2021a. (in prep.): "World Agromyzidae online", hosted by Senckenberg DEI. http://sdei.senckenberg.de/tschirnhaus.

Wiegmann, B.M., Trautwein, M.D., Winkler, I.S., Barr, N.B., Kim, J.W., Lambkin, C., Bertone, M.A., Cassel, B.K., Bayless, K.M., Heimberg, A.M., Wheeler, B.M., 2011. Episodic radiations in the fly tree of life. Proc. Natl. Acad. Sci. 108 (14), 5690–5695.

Winkler, I.S., Mitter, C., Scheffer, S.J., 2009a. Repeated climate-linked host shifts have promoted diversification in a temperate clade of leaf-mining flies. Proc. Natl. Acad. Sci. 106, 18103–18108.

Winkler, I.S., Scheffer, S.J., Mitter, C., 2009b. Molecular phylogeny and systematics of leaf-mining flies (Diptera: Agromyzidae): delimitation of *Phytomyza* Fallén sensu lato and included species groups, with new insights on morphological and host-use evolution. Syst. Entomol. 34 (2), 260–292.

Winkler, I.S., Labandeira, C.C., Wappler, T., Wilf, P., 2010. Distinguishing Agromyzidae (Diptera) leaf mines in the fossil record: new taxa from the Paleogene of North America and Germany and their evolutionary implications. J. Paleontol. 84, 935–954.

Xuan, J.L., Scheffer, S.J., Lonsdale, O., Cassel, B.K., Lewis, M.L., Eiseman, C.S., Liu, W.X., Wiegmann, B.M., 2022. A genome-wide phylogeny and the diversification of genus *Liriomyza* (Diptera: Agromyzidae) inferred from anchored phylogenomics. Syst. Entomol. 1–20.

Yan, L., Buenaventura, E., Pape, T., Narayanan Kutty, S., Bayless, K.M., Zhang, D., 2021a. A phylotranscriptomic framework for flesh fly evolution (Diptera, Calyptratae, Sarcophagidae). Cladistics 37, 540–558.

Yan, L., Pape, T., Meusemann, K., Kutty, S.N., Meier, R., Bayless, K.M., Zhang, D., 2021b. Monophyletic blowflies revealed by phylogenomics. BMC Biol. 19 (1), 1–14.

Young, A.D., Gillung, J.P., 2020. Phylogenomics - principles, opportunities, and pitfalls of big-data phylogenetics. Syst. Entomol. 45 (2), 225–247.

Young, A.D., Lemmon, A.R., Skevington, J.H., Mengual, X., Ståhls, G., Reemer, M., Jordaens, K., Kelso, S., Lemmon, E.M., Hauser, M., De Meyer, M., 2016. Anchored enrichment dataset for true flies (order Diptera) reveals insights into the phylogeny of flower flies (family Syrphidae). BMC Evol. Biol. 16 (1), 143–156.

Zhang, C., Rabiee, M., Sayyari, E., Mirarab, S., 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinf. 19 (6), 53–75.

Zlobin, V.V., 2002. Contribution to the knowledge of the genus *Pseudonapomyza* Hendel (Diptera: Agromyzidae), with descriptions of twenty-four old world species. Dipterol. Res. 13 (4), 205–245.

Zlobin, V.V., 2007. Review of mining flies of the genus *Aulagromyza* Enderlein (Diptera: Agromyzidae). II. Int. J. Dipter. Res. 18 (4), 223–226.

Zlobin V.V., 1984. New species of *Selachops* Wahlberg and *Metopomyza* Enderlein (Diptera: Agromyzidae) from the Far East. In Diptera (Insecta) their Systematics, Geographic Distribution and Ecology 45-52.