Received: 15 March 2025 | Accepted: 29 April 2025

DOI: 10.1111/syen.12686

ORIGINAL ARTICLE

Systematic Entomology

Extensive hybridisation and complex evolutionary history in the leafhopper genus *Agnesiella* (Hemiptera: Cicadellidae: Typhlocybinae)

Junjie Wang[1] | **Xian Zhou**[2] | **Christopher H. Dietrich**[3] | **Yanghui Cao**[1] | **Min Huang**[1]

[1]Key Laboratory of Plant Protection Resources and Pest Management of Ministry of Education, College of Plant Protection, Northwest A&F University, Yangling, China

[2]Key Laboratory of Oasis Agricultural Pest Management and Plant Protection Resources Utilization, Xinjiang Uygur Autonomous Region, College of Agriculture, Shihezi University, Shihezi, China

[3]Illinois Natural History Survey, Prairie Research Institute, University of Illinois, Champaign, Illinois, USA

Correspondence
Yanghui Cao and Min Huang, Yanghui Cao and Min Huang, Key Laboratory of Plant Protection Resources and Pest Management of Ministry of Education, College of Plant Protection, Northwest A&F University, Yangling, Shaanxi 712100, China.
Email: caoyh@nwafu.edu.cn and huangmin@nwsuaf.edu.cn

Funding information
the Ministry of Science and Technology of the People's Republic of China; National Natural Science Foundation of China

Editor: Christopher L. Owen

Abstract

The subfamily Typhlocybinae (Cicadellidae) represents a globally distributed, species-rich lineage of leafhoppers. Despite significant advancements in morphological taxonomy and higher-level phylogenetics, species-level evolutionary dynamics within individual typhlocybine genera remain poorly understood. This study focuses on the endemic Oriental genus *Agnesiella* Dworakowska, which exhibits high species diversity in the Hengduan Mountains of Southwest China. Using whole-genome sequences of 48 individuals representing 40 *Agnesiella* species and 4 species from related genera, we reconstructed phylogenetic relationships, estimated divergence times and investigated patterns of hybridisation and introgression within this genus using single-copy orthologue sequences (SCOs), ultraconserved elements (UCEs) and single-nucleotide polymorphism sequences (SNPs). Our findings reveal a complex evolutionary history in *Agnesiella*, shaped by incomplete lineage sorting (ILS) and extensive interspecific gene flow, particularly within the subgenus *Draberiella*. The diversification of *Agnesiella* coincides with the orogenic and climatic changes in the Hengduan Mountains during the Miocene–Pliocene, which may have promoted allopatric isolation, secondary contact and hybridisation. Functional analysis of the introgressed genomic regions suggests their potential contribution to adaptive evolution, including enhanced metabolism of nitrogen compounds and plant secondary metabolites. These findings provide novel insights into the evolutionary dynamics of Typhlocybinae, emphasising the critical role of hybridisation and introgression in driving speciation and adaptation in insect lineages.

KEYWORDS
incomplete lineage sorting, introgression, leafhopper, Phylogenomics, Typhlocybinae

INTRODUCTION

The leafhopper subfamily Typhlocybinae, which includes 530 valid genera and 5113 species (Dmitriev et al., 2022), is the second largest group of Cicadellidae and is distributed worldwide. This group feeds on a wide variety of host plants, with many species exhibiting remarkable host specialisation (Dmitriev & Dietrich, 2010; Nickel & Remane, 2002). By feeding on parenchyma cell contents, they often cause serious damage to leaves (Goodchild, 1966) and many typhlocybines are serious agricultural pests. The subfamily comprises six extant tribes (Alebrini, Beameranini, Dikraneurini, Empoascini, Erythroneurini and Typhlocybini) and one extinct tribe (Protodikraneurini), all of which originated during the mid-to-late Cretaceous (Cao et al., 2023; Yan et al., 2022).

Previous research on Typhlocybinae has primarily focused on morphological taxonomy and higher-level phylogenetic relationships

(Xu et al., 2021; Lu et al., 2021; Yan et al., 2022; Cao et al., 2023; Zhou et al., 2023; Wang et al., 2022, 2024). While new taxa continue to be described (Dietrich, 2013; Wang et al., 2024), the mechanisms driving diversification within the subfamily remain largely unexplored. Previous studies suggest that intercontinental dispersal and isolation played a key role in early lineage diversification (Cao et al., 2023). Some studies have explored the phylogeography of some economically important species (Li et al., 2022; Zhao et al., 2022) and species delimitation among closely related species (Fu et al., 2014). However, the phylogenetic relationships and evolutionary dynamics within individual typhlocybine genera have received limited attention. The potential roles of hybridisation and introgression in adaptability or speciation in leafhoppers are unexplored but have been suggested by some previous authors based on evidence from morphological observations and limited molecular data (Krishnankutty et al., 2015; Ross, 1958).

Recent genome-wide studies have revealed that gene flow among species is more widespread than previously thought (Aguillon et al., 2022; Edelman & Mallet, 2021; Mallet et al., 2016), often leading to phylogenetic incongruence and reticulate evolutionary patterns in closely related groups (Jiao et al., 2020; Steenwyk et al., 2023). Such phenomena have been documented in diverse taxa, including mosquitoes (Fontaine et al., 2015), Darwin's finches (Lamichhaney et al., 2015), Malawi cichlids (Malinsky et al., 2018), *Heliconius* butterflies (Mavárez et al., 2006; Thawornwattana et al., 2023), fruit flies (Suvorov et al., 2022) and ancient humans (Green et al., 2010). Previous studies suggest that approximately 25% of flowering plants and 10% of animal species experience hybridisation or introgression (Mallet, 2005).

While gene flow was traditionally thought to have a homogenising effect that suppresses speciation, recent research indicates that introgression can generate substantial genetic variation, potentially driving adaptive radiation (Grant & Grant, 2019; Meier et al., 2017). The genus *Heliconius* provides a striking example of homoploid hybrid speciation under rigorous criteria (Rosser et al., 2024; Schumer et al., 2014), where introgression has enabled the acquisition of adaptive traits that establish reproductive isolation from parental species, including wing colouration patterns (Sánchez et al., 2015), visual preference (Rossi et al., 2024) and host plant preference (Merrill et al., 2013).

The microleafhopper genus *Agnesiella* Dworakowska (Hemiptera: Cicadellidae), belonging to tribe Typhlocybini Kirschbaum, comprises two subgenera: *Agnesiella* (14 species) and *Draberiella* (42 species). Recent phylogenetic studies indicate that *Agnesiella* can be specifically placed within the *Linnavuoriana* genus complex, representing a recently diverged lineage within Typhlocybinae (Lu et al., 2021; Yan et al., 2022; Cao et al., 2023; Zhou et al., 2023). Species of *Agnesiella* predominantly feed on alder trees (*Alnus*, Betulaceae) and are primarily distributed in the Oriental region, with particularly high species diversity concentrated in the southwestern mountains of China, a global biodiversity hotspot (Mi et al., 2021; Wang et al., 2022, 2024; Yan & Yang, 2016, 2019). East and Southeast Asian mountain ranges, particularly the Tibetan Plateau and its environs, are considered prime locations for investigating evolutionary processes (Igea & Tanentzap, 2021; Spicer et al., 2020). The Hengduan Mountains of the southeastern Tibetan Plateau, a region characterised by complex orogenesis and climate change, serve as a compelling example of how geological and climatic forces drive species diversification (Ding et al., 2020; Li et al., 2021; Xing & Ree, 2017). This observation aligns with broader research suggesting a positive relationship between biodiversity and the magnitude and rate of tectonic uplift (Marder et al., 2025). *Agnesiella*, a relatively recently diverged lineage with high species richness in the Hengduan Mountains, appears to represent a rapidly diversifying group. This rapid diversification, however, poses challenges for resolving interspecific relationships, as processes such as incomplete lineage sorting (ILS), hybridisation and introgression can generate conflicting evolutionary signals across the genome, leading to phylogenetic incongruence among genetic loci. These complexities make the genus *Agnesiella* an exceptional model for investigating the interplay of speciation, diversification, hybridisation and introgression in leafhoppers within montane ecosystems.

In order to resolve the evolutionary relationships within the genus *Agnesiella*, we sequenced whole genomes of multiple species. Utilising datasets of single-copy orthologous sequences (SCOs) and ultraconserved elements (UCEs) obtained from these genomic resources, we reconstructed a robust phylogeny of *Agnesiella* and estimated its divergence times. This study represents the first genome-wide investigation of hybridisation and introgression in leafhoppers. By analysing single-nucleotide polymorphisms (SNPs) and gene trees, we uncovered evidence for extensive hybridisation events spanning the early and later diverging phases of *Agnesiella*.

MATERIALS AND METHODS

Taxa sampling, DNA extraction and sequencing

This study included 48 individuals: 43 ingroup specimens representing 40 *Agnesiella* species and five outgroup specimens from four related genera within the *Linnavuoriana* complex (*Amurta*, *Linnavuoriana*, *Sannella* and *Vatana*). Specimens were collected primarily from the southwestern mountains of China (Figure 1). *Sannella crucifera* was selected to root the tree as it is the most divergent outgroup based on phylogenetic analysis using mitochondrial genome sequences (13 protein-coding genes and 2 rRNAs). Specimens were identified morphologically, and voucher specimens of each species were preserved in 100% ethanol and stored at −80°C before DNA extraction at the Entomological Museum of Northwest A&F University (NWAFU), Yangling, China. A full list of sampled individuals is provided in Supplementary Table S1.

Genomic DNA was extracted from individual specimens using the TIANamp Micro DNA Kit (TIANGEN, Beijing, China). DNA concentration was then quantified using a NanoDrop One spectrophotometer. Library preparation with a 350 base-pair (bp) insert size and paired-end 150 bp sequencing were performed on Illumina NovaSeq or MGI-seq platforms at Novogene and Berry Genomics (Beijing, China). For

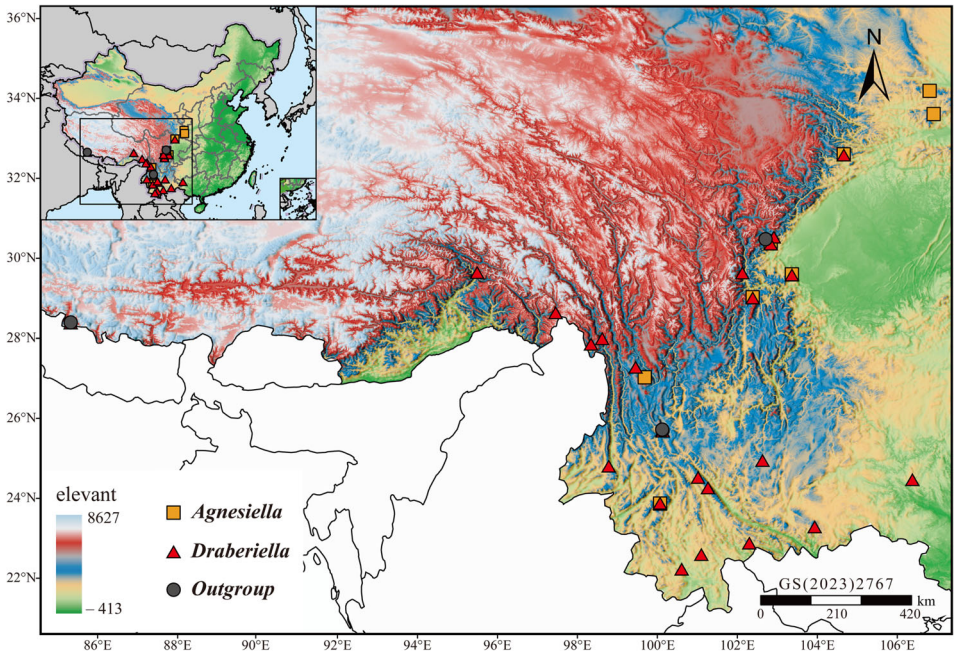

FIGURE 1 Geographic distribution and sampling localities. Geographic distribution and sampling localities of the 48 individuals, with different shapes and colours representing distinct species groups.

the majority of samples, approximately 20 GB of sequencing data were generated, resulting in an estimated 10–20× coverage (Zhang et al., 2019). An exception was *A. juglandis* sample1, for which 100 GB of data were generated to facilitate genome size estimation.

Genome assembly, annotation, variant calling and filtering

Genome assemblies were generated using a modified PLWS v1.0.6 pipeline (Zhang et al., 2019). Genome completeness for all 48 individuals was assessed using BUSCO v5.7.0 with the hemiptera_odb10 dataset ($n = 2510$) (Manni et al., 2021). To adopt a more relaxed completeness criterion for BUSCO classification, we increased the standard deviation threshold (σ) of the mean BUSCO length to 2σ. For the three high-quality genome assemblies (*Agnesiella roxana*, *Agnesiella longisagittata* and *Vatana ogromna*) exhibiting BUSCO completeness scores above 94%, we employed BRAKER2 v2.1.6 (Brůna et al., 2021) to integrate ab initio gene predictions with protein homology-based evidence. Prior to gene prediction, repetitive and low-complexity DNA regions were masked using WindowMasker (Morgulis et al., 2006). In the absence of transcriptomic data, we selected the Arthropoda protein database from OrthoDB v11 as the source of protein homology evidence for BRAKER2. Subsequently, protein-coding genes were annotated across all assembled genomes using GeMoMa v1.9 (Keilwagen et al., 2019), gene annotations from the three BRAKER2-processed assemblies and the genome annotation of *Macrosteles quadrilineatus* (GenBank: GCA_028750875.1) as references. Finally, to identify potential exogenous sequences, we screened all assembled genomes using Kraken2 (Wood et al., 2019).

Variant calling and identification of SNPs were performed using the Genome Analysis Toolkit (GATK v4.5.0.0) best practices pipeline (Van der Auwera & O'Connor, 2020). The cleaned reads, which were trimmed using fastp v0.23.4, were mapped to the reference genome (*A. roxana*, BUSCO completeness = 97.4%) using BWA-MEM2 v2.2.1 (Chen, 2023; Vasimuddin et al., 2019). Hard filtering was applied in GATK based on the following criteria: QD <2.0, QUAL <30.0, SOR >3.0, FS >60.0, MQ <40.0, MQRankSum < −12.5, ReadPosRankSum < −8.0. Further filtering with VCFtools (Danecek et al., 2011) retained only biallelic SNPs present in at least 30% of individuals, with a minimum site quality score of 30 and a mean site depth between 5× and 100×. Additionally, SNP sites potentially derived from exogenous sequences were removed. The resulting SNP VCF file was used for downstream gene flow analysis. Detailed scripts for this pipeline are publicly available on GitHub at https://github.com/WJJ-97/Phylogenomics/tree/main/Agnesiella_Phylogenomics/Scripts.

Marker extraction and datasets generation

Amino acid (aa) and nucleotide (nt) sequences corresponding to the longest transcripts were extracted from protein sequences annotated by GeMoMa. Single-copy orthologue sequences (SCOs) were identified using OrthoFinder v2.5.5 (Emms & Kelly, 2019). A probe set for UCEs was designed specifically for the genus *Agenesiella* using PHYLUCE v1.7.3 (Faircloth, 2016), using five Cicadellidae genomes: the three assemblies with BUSCO completeness values above 94%, as well as *Empoasca onukii* (GenBank GCA_018831715.1) and *Macrosteles quadrilineatus* (GenBank GCA_028750875.1). UCEs were extracted for each individual using PHYLUCE v1.7.3, with redundant sequences identified

and removed. A minimum representation of four taxa per locus was ensured for both datasets, and potential exogenous sequences were filtered out using Kraken2.

The SCOs and UCEs were aligned separately using MAFFT v7.525 with the L-INS-i method (Katoh & Standley, 2013). TrimAl v1.4.1 was used for trimming with the 'gappyout' option (Capella-Gutiérrez et al., 2009). The sequences were then filtered based on their properties to mitigate common confounding factors in phylogenomic inference (Steenwyk et al., 2023). We also filtered for data completeness, representing the minimum proportion of individuals per locus, to reduce the impact of missing data (Smith et al., 2020).

We applied several filtering strategies to mitigate common confounding factors in phylogenomic inference. These included sequence-based approaches, such as normalised relative compositional frequency variation (nRCFV) and the number of parsimony-informative sites (nPIS), as well as tree-based methods, including average bootstrap support (ABS) and clustering information distance (CID) (Fleming & Struck, 2023; Salichos & Rokas, 2013; Shen et al., 2016; Smith, 2020). These filtering strategies were used to subsample loci and generate datasets for further analyses. First, loci with low nRCFV values, indicating reduced compositional bias, were considered more suitable for phylogenetic analysis, as many phylogenetic models assume compositional homogeneity. Loci with nRCFV values below a specific threshold (median + 3.5 × Median Absolute Deviation (MAD)) were retained using RCFVReader_v1.pl. with default parameters (Fleming & Struck, 2023). Next, nPIS for each locus was calculated using PhyKIT v1.20 (Steenwyk et al., 2021), and loci with more than a certain threshold of parsimony-informative sites were retained. After reconstructing the gene trees, loci with higher ABS values across all internal branches of the gene tree were retained to improve phylogenetic signal (Salichos & Rokas, 2013). Finally, CID values between each pair of trees were computed using the 'TreeDist' R package, and gene trees with average CID values below a threshold (median + 3.5 × MAD) were retained, filtering out highly heterogeneous loci and trees.

Datasets were named according to their sequence type, completeness and final filtering strategy. Some datasets were generated for concatenation-based phylogenomic analyses using FASConCAT-G v1.06 (Kück & Longo, 2014). SCOs datasets were labelled as 'SCOs_AA_80_-CID' for aa sequences and 'SCOs_NT_80_CID' for nt sequences, both with 80% completeness and an ABS greater than 75. The UCEs dataset was named 'UCEs_NT_100_CID', with 100% completeness and an ABS greater than 80. To integrate phylogenetic signals from the different data types, redundant sequences were first removed from the SCOs and UCEs datasets. Then, the top 400 nucleotide positions from each dataset, selected based on their ABS, were combined to create a new dataset named 'SCOs_80_UCEs_100_NT_CID_800'. These datasets included gene trees and loci filtered using the CID strategy.

Phylogenetic analysis

We employed both coalescent-based and concatenation-based phylogenetic approaches to infer insect phylogeny. Specifically, we utilised the multi-species coalescent (MSC) model for coalescent-based analyses and partitioned and heterogeneous models for concatenation-based analyses. These analyses were conducted on two data types: SCOs (amino acid and nucleotide sequences) and UCEs (nucleotide sequences). For coalescent-based species tree reconstruction, we used ASTRAL-Hybrid under the MSC model (Zhang & Mirarab, 2022) to account for ILS. Branch support for these species trees was assessed using local posterior probability (LPP) from quartet frequencies, estimated from a set of gene trees (Sayyari & Mirarab, 2016). Gene trees were initially inferred using IQ-TREE v2.3.4 (Minh et al., 2020).

Concatenation-based phylogenetic analyses were performed using both maximum likelihood (ML) and Bayesian inference (BI) methods. ML analyses were conducted in IQ-TREE v2.3.4, and BI analyses were performed using MrBayes v3.2a (Ronquist et al., 2012). For partitioned analyses, the best partitioning schemes and substitution models were identified using ModelFinder (Kalyaanamoorthy et al., 2017). For ML analysis of amino acid datasets, we also implemented mixture models ('-m ELM+C20 + F + R') (Banos et al., 2024; Si Quang et al., 2008). To account for potential heterotachy across sites and lineages in both nucleotide and amino acid datasets, we applied the GHOST (General Heterogeneous evolution On a Single Topology) model in IQ-TREE (Crotty et al., 2020). For BI analyses, we performed two independent runs of 5 million generations, sampling every 500 generations. Each run consisted of six chains continuing until all effective sample sizes (ESS) exceeded 200 to ensure analysis convergence.

Branch support for ML analyses was evaluated using both 2000 SH-aLRT replicates and 2000 UFBoot2 bootstraps (Guindon et al., 2010; Hoang et al., 2018). Branches with SH-aLRT support ≥80% and UFBoot2 support ≥95% were considered strongly supported. For BI analyses, Bayesian Posterior Probabilities (BPP) were used to assess branch support, with values ≥95% indicating strong support. To further assess genealogical concordance within our phylogenomic datasets, we calculated gene concordance factors (gCF) and site concordance factors (sCF) using IQ-TREE (Minh et al., 2020; Mo et al., 2023). Phylogenetic trees were visualised and edited using iTOL v6 (Letunic & Bork, 2024) and Adobe Illustrator.

Divergence time estimation

Divergence times and 95% confidence intervals were estimated using a Bayesian relaxed clock approach implemented in MCMCtree (Álvarez-Carretero et al., 2019). We employed the species tree topology and loci from the 'SCOs_80_UCEs_100_NT_CID_800' dataset after the removal of significantly introgressed loci to mitigate the potential influence of introgression (Pang & Zhang, 2023). We employed an independent rates clock model (clock = 2) and a GTR substitution model (model = 7). To inform prior distributions for subsequent analyses, we performed a preliminary analysis. Based on the posterior distributions from this initial run, we set gamma priors for the mean substitution rate (rgene_gamma) with parameters (1.8, 20, 1)

and rate drift (sigma2_gamma) with parameters (1, 10, 1) for non-calibrated nodes. Two independent Markov chain Monte Carlo (MCMC) runs were performed, each for 1.5 million generations, with samples collected every 1000 generations. The initial 300,000 generations were discarded as burn-in. In the absence of robust fossil calibrations, we implemented a root age constraint, setting a maximum age of 47 Ma, consistent with previous research (Yan et al., 2022). Furthermore, we incorporated secondary calibrations derived from the fossil-calibrated Typhlocybinae timetree of Cao et al. (2023) to constrain the divergence times for *Amurta* and related genera (14.29–33.68 Ma).

Phylogenetic incongruence testing

To investigate potential sources of phylogenetic discordance, we employed Phytop v0.3 (Shang et al., 2024), a tool designed for visualising and quantifying gene tree discordance within species trees. Using gene trees and species trees inferred from ASTRAL-Hybrid analyses, Phytop calculates two indices to distinguish between ILS and introgression/hybridisation (IH): the ILS Index (ranging from 0% to 100%), which quantifies discordance attributable to ILS, and the IH Index (ranging from 0% to 50%), reflecting asymmetric discordance potentially caused by introgression. We assessed the statistical significance of observed discordance patterns using chi-square tests. Furthermore, Phytop's visualisation capabilities aided in the clear differentiation of discordance patterns resulting from ILS versus IH.

We also employed topology tests to compare alternative phylogenetic hypotheses derived from various datasets and phylogenetic methods. Specifically, we used the approximately unbiased (AU) test (Shimodaira, 2002), weighted Kishino-Hasegawa (WKH) test and weighted Shimodaira-Hasegawa (WSH) test, as implemented in IQ-TREE under various datasets and models.

Genome-wide evidence of gene flow and reticulate evolution

We employed site-pattern-based methods, specifically Patterson's *D* statistics, to detect introgression by analysing shared allele inconsistencies among species trios (Green et al., 2010). The *D*-statistic is a robust approach for inferring gene flow, even with limited sampling per individual (Hibbins & Hahn, 2022). We calculated genome-wide D-statistics using Dsuite v0.5 (Malinsky et al., 2021) from quality-filtered SNP data, evaluating all possible triplet topologies. To address potential false positives arising from variable evolutionary rates among divergent taxa, we employed the ABBA-clustering test, which is robust to mutation rate variation across the genome (Koppetsch et al., 2024). The Patterson's *D* and the f_4 admixture ratio were visualised as heatmaps using scripts (plot_d.rb and plot_f4ratio.rb) provided by Dsuite's developers. Furthermore, we applied the *f*-branch statistic to pinpoint specific branches exhibiting evidence of gene flow, leveraging the f_4-ratio in conjunction with established phylogenetic relationships (Malinsky et al., 2018, 2021).

We further utilised tree-based methods, applying QuIBL (Quantifying Introgression via Branch Lengths) to differentiate introgression from ILS and to estimate the fraction of introgression within species trios (Edelman et al., 2019). Given that QuIBL's accuracy is influenced by sequence alignment length (Koppetsch et al., 2024), we generated the dataset 'SCOs_UCEs_NT_nRCFV_41taxa_900', comprising nucleotide sequence data exceeding 900 bp for all ingroup taxa and the outgroup genus *Sannella*. From this dataset, we selected 3063 gene trees as input for QuIBL analysis. QuIBL was executed using default parameters, with the exception of increasing the number of expectation–maximisation (EM) steps from 10 to 50, a modification recommended for large datasets. Results were subsequently analysed and visualised using the 'quiblR' R package (https://github.com/nbedelman/quiblR).

Finally, to infer species hybridisation networks, we implemented quartet-based approaches under the network multispecies coalescent (NMSC) model. This analysis utilised 4532 gene trees derived from the 'SCOs_nPIS15_UCEs_nPIS50_NT_44taxa' dataset. This dataset, filtered based on nPIS thresholds, encompassed all sampled species. The NANUQ algorithm, implemented in the 'MSCquartets' R package (Allman et al., 2019), was employed to identify hybridisation signals, using an alpha level of 5e−12, and to resolve four-taxon networks, using a beta level of 0.05. To refine our network analysis, we also applied the T3 model and TINNiK algorithm, which categorise quartet concordance factors (qcCFs) into tree-like and reticulation signals within the simplex plot (Allman et al., 2022, 2024). A splits graph, visualising the NANUQ quartet distance matrix, was generated using SplitsTree6 (Huson & Bryant, 2024).

Functional annotation and enrichment analysis of introgression sequences

To identify introgressed genomic regions, we calculated *D*, f_d and f_{dM} statistics using the 'Dinvestigate' module of Dsuite. These statistics were computed for trios across 50-SNP windows with a 25-SNP step size, focusing on trios exhibiting significant introgression signals (Z-score >3 and f_4-ratio >0.1). Genomic regions with introgression signals (f_{dM} >0.01) were then extracted from the *A. roxana* reference genome using SeqKit v2.8.2 (Shen et al., 2024).

Functional annotation of the extracted sequences was performed using EggNOG-mapper v2.1.12 against the eggNOG database (v5.0) (Cantalapiedra et al., 2021; Huerta-Cepas et al., 2019). Subsequently, we conducted enrichment analysis and visualised the functional categories using the 'clusterProfiler' R package (Wu et al., 2021).

To assess the potential influence of introgression on phylogenetic inference, we identified putatively introgressed loci within the SCOs and UCEs datasets using BLAST v2.15.0 (Camacho et al., 2009). We excluded these sites and subsequently reconstructed phylogenetic trees using the remaining dataset to evaluate their impact on tree topology.

RESULTS

Genome assembly, annotation and variant analysis

Genome assemblies for 48 individuals achieved an average sequencing coverage ranging from 16.08x to 92.61x, with assembly sizes spanning 685 Mb to 1344 Mb (Table S2). BUSCO analysis indicated that genome completeness ranged from 57.5% to 97.3% (Figure S1). Gene prediction using GeMoMa identified 42,413 to 87,824 protein-coding sequences across these assemblies (Table S2). For variant detection, paired-end reads were aligned to the reference genome of *Agnesiella roxana*, resulting in average sample coverage from 3.6× to 38.7×. Initial base calling identified 339,476,540 SNP sites; subsequent filtering for quality yielded a final set of 11,117,624 high-confidence variants.

Marker extraction and datasets

From the assembled genomes of 48 individuals, we recovered 9265–10,949 loci, representing 70.12–82.86% of the targeted UCEs. Following the removal of potential exogenous and redundant sequences, and filtering loci retrieved from less than four samples, we obtained a final set of 10,014 UCE loci. The lengths of these loci ranged from 200 to 2518 bp. For SCOs, we identified 23,343 loci with sequence lengths ranging from 12 to 5782 amino acids (aa) and 39 to 17,349 bp for nucleotides (nt).

After alignment, trimming and rigorous filtering, we constructed four final datasets for concatenation-based phylogenomic analyses. These datasets were: 'SCOs_AA_80_CID' (261 loci, 156,356 amino acid sites), 'SCOs_NT_80_CID' (838 loci, 1,183,727 nucleotide sites), 'UCEs_NT_100_CID' (1067 loci, 1,012,589 nucleotide sites) and 'SCOs_80_UCEs_100_NT_CID_800' (800 loci, 1,087,968 nucleotide sites). The average locus lengths for these datasets were 599.07 aa, 1412.56 bp, 949.01 bp and 1359.96 bp, respectively. Further details regarding dataset composition following each filtering step, as well as the removal of loci identified as potentially introgressed, are provided in Table S3.

Phylogenetic relationships

Both coalescent-based and concatenation-based phylogenetic analyses provide robust support for the evolutionary relationships among closely related genera within the *Linnavuoriana* complex, and strongly support genera *Agnesiella* and *Vatana* as sister groups (SH-aLRT = 100, UFBoot2 = 100) (Figure 2), consistent with previous mitogenomic studies (Yan et al., 2022; Zhou et al., 2023). The monophyly of the genus *Agnesiella* and its two subgenera is strongly supported (Figure 2). Within the genus *Agnesiella*, six major clades (Clade I–VI) were identified, with the subgenus *Agnesiella* corresponding to Clade I and the subgenus *Draberiella* encompassing Clades II to VI. Phylogenetic relationships among Clades II, III and V are well resolved (Figure S2). However, there were inconsistencies regarding whether Clade III or Clade IV is closer to the base of the tree (Tables 1 and S3).

The coalescent-based phylogenetic reconstructions revealed 16 distinct topologies, with most nodes receiving strong support (Table S3). The topology of the subgenus *Agnesiella* (Clade I) was relatively stable, predominantly matching the Clade Ib topology (Figure S2; Table S3). In contrast, Clades IV and VI of the subgenus *Draberiella* displayed three and five different topologies, respectively (Figure S2; Table S3), with the variation primarily caused by the unstable positions of certain species, such as *A. ela* in Clade IV and *A. biprotrusa* and *A. xantha* in Clade VI (Figure S2). The positions of Clades III and IV varied depending on dataset filtering strategies. Species trees constructed from ABS-filtered datasets positioned Clade III closer to the base of the tree. The removal of loci showing significant introgression altered the topology of several major clades in a few datasets (Table S3).

The concatenation-based phylogenetic analyses generated eight distinct topologies, with most nodes receiving strong support (Table 1). However, these topologies differed from those obtained through coalescent-based analyses mainly within Clade I and Clade VI (Table 1). Clades I and VI showed four and five different topologies, respectively (Table 1; Figure S2). Partitioned analyses of datasets containing SCOs using both ML and BI methods also revealed topological incongruence of Clades I and VI (Table 1). The removal of loci showing significant introgression from the raw datasets produced topologies largely congruent with the original results. However, this removal led to topological shifts specifically within Clades IV and VI in phylogenetic trees derived from the datasets containing UCEs, slightly improving the stability of the overall topology (Table 1).

Phypop analyses attributed topological conflicts among gene trees primarily to ILS (Figure S3). However, we also detected signals of IH. For example, *A. hamaculeata* exhibited significant IH signals (30.6–46.1%) with (*A. recurva*, (*A. juglandis*, (*A.* sp.*1*, *A. protensa*))) in the subgenus *Agnesiella*. Similarly, *A. lidia* and *A. kamala* displayed substantial IH signals (24.8–44.0%) with other species in Clade III. In Clade VI, more complex IH signals were detected.

Topology tests were conducted for all major clades except Clade II. While no single topology received consistent support across all tests. The test results for each dataset consistently supported the phylogenetic tree topology constructed from that dataset (Figure S2). After comprehensive analyses, the species tree topology derived from the 'SCOs_80_UCEs_100_NT_CID_800' dataset following the removal of significantly introgressed loci was selected as the best representation of the true phylogenetic relationships (Figure 2). This dataset integrates phylogenetic signal from two distinct data types, and the ASTRAL-Hybrid species tree topology effectively mitigated the impact of ILS and demonstrating robustness to some gene flow (Jiao et al., 2020; Pang & Zhang, 2023; Zhang & Mirarab, 2022). Using this topology and

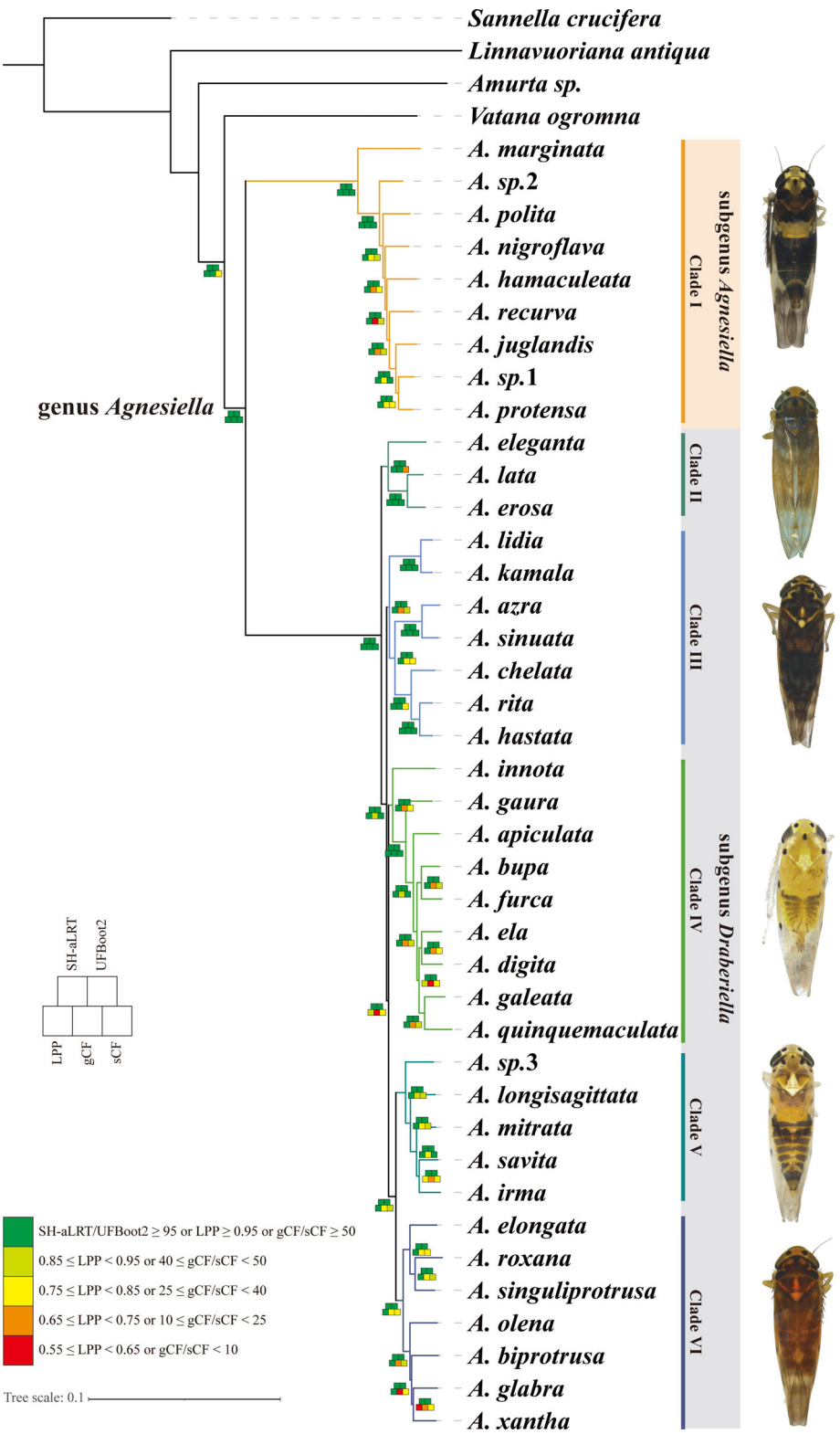

FIGURE 2 Maximum likelihood tree of the genus *Agnesiella*. Phylogeny inferred from the species tree topology and loci of the 'SCOs_80_UCEs_100_NT_CID_800' dataset following the removal of significantly introgressed loci. Node support from all analyses is represented by coloured squares.

dataset, constrained tree searches were performed with ML analysis in IQ-TREE under the partition model (Figure 2). Additionally, gCF and sCF were calculated to assess the level of support for individual loci and sites. Some nodes exhibited low concordance (gCF/sCF < 25), reflecting instability in both coalescent- and concatenation-based analyses.

TABLE 1 Phylogenetic topologies from concatenation-based analyses across different datasets, methods and substitution models.

Datasets	Methods	Substitution models	Raw topologies	Topologies after filtering significantly introgressed loci
SCOs_AA_80_CID	ML	Partition	Topo I: (Clade Ic, (Clade II, (Clade IVa, (Clade IIIa, (Clade Va, Clade VIf)))))	Topo I: (Clade Ic, (Clade II, (Clade IVa, (Clade IIIa, (Clade Va, Clade VIf)))))
		GHOST	Topo I: (Clade Ic, (Clade II, (Clade IVa, (Clade IIIa, (Clade Va, Clade VIf)))))	Topo I: (Clade Ic, (Clade II, (Clade IVa, (Clade IIIa, (Clade Va, Clade VIf)))))
		ELM+C20 + F + R	Topo I: (Clade Ic, (Clade II, (Clade IVa, (Clade IIIa, (Clade Va, Clade VIf)))))	Topo I: (Clade Ic, (Clade II, (Clade IVa, (Clade IIIa, (Clade Va, Clade VIf)))))
	BI	Partition	Topo I: (Clade Ic, (Clade II, (Clade IVa, (Clade IIIa, (Clade Va, Clade VIf)))))	Topo I: (Clade Ic, (Clade II, (Clade IVa, (Clade IIIa, (Clade Va, Clade VIf)))))
SCOs_NT_80_CID	ML	Partition	Topo II: (Clade Id, (Clade II, (Clade IVa, (Clade IIIa, (Clade Va, Clade VIg)))))	Topo II: (Clade Id, (Clade II, (Clade IVa, (Clade IIIa, (Clade Va, Clade VIg)))))
		GHOST	Topo II: (Clade Id, (Clade II, (Clade IVa, (Clade IIIa, (Clade Va, Clade VIg)))))	Topo II: (Clade Id, (Clade II, (Clade IVa, (Clade IIIa, (Clade Va, Clade VIg)))))
	BI	Partition	Topo III: (Clade Id, (Clade II, (Clade IVa, (Clade IIIa, (Clade Va, Clade VIh)))))	Topo III: (Clade Id, (Clade II, (Clade IVa, (Clade IIIa, (Clade Va, Clade VIh)))))
UCEs_NT_100_CID	ML	Partition	Topo IV: (Clade Ie, (Clade II, (Clade IIIa, (Clade IVa, (Clade Vb, Clade VIi)))))	Topo VII: (Clade Ie, (Clade II, (Clade IIIa, (Clade IVc, (Clade Vb, Clade VIi)))))
		GHOST	Topo IV: (Clade Ie, (Clade II, (Clade IIIa, (Clade IVa, (Clade Vb, Clade VIi)))))	Topo VII: (Clade Ie, (Clade II, (Clade IIIa, (Clade IVc, (Clade Vb, Clade VIi)))))
	BI	Partition	Topo IV: (Clade Ie, (Clade II, (Clade IIIa, (Clade IVa, (Clade Vb, Clade VIi)))))	Topo VII: (Clade Ie, (Clade II, (Clade IIIa, (Clade IVc, (Clade Vb, Clade VIi)))))
SCOs_80_UCEs_100_NT_CID_800	ML	Partition	Topo V: (Clade If, (Clade II, (Clade IIIa, (Clade IVa, (Clade Va, Clade VIi)))))	Topo V: (Clade If, (Clade II, (Clade IIIa, (Clade IVa, (Clade Va, Clade VIi)))))
		GHOST	Topo V: (Clade If, (Clade II, (Clade IIIa, (Clade IVa, (Clade Va, Clade VIi)))))	Topo VIII: (Clade If, (Clade II, (Clade IIIa, (Clade IVa, (Clade Va, Clade VIj)))))
	BI	Partition	Topo VI: (Clade Id, (Clade II, (Clade IIIa, (Clade IVa, (Clade Va, Clade VIi)))))	Topo VIII: (Clade If, (Clade II, (Clade IIIa, (Clade IVa, (Clade Va, Clade VIj)))))

Note: Details of the topologies of each major clade are provided in Figure S2.

Divergence times

MCMCTree analyses estimated that the split between the outgroups and the genus *Agnesiella* occurred during the early Oligocene to middle Miocene, with a mean estimate of 21.79 million years ago (Mya; 95% highest posterior density credible interval [HPD-CI]: 13.33–31.18 Mya) (Figure 3). The divergence between subgenera *Agnesiella* and *Draberiella* was estimated to have occurred during the early Miocene (95% HPD-CI: 12.61–30.08 Mya), and rapid speciation occurred mainly during the middle Miocene to late Pliocene.

Gene flow and species network

Analysis of gene flow and species relationships within the genus *Agnesiella* revealed several key findings. Patterson's *D* and f_4-ratio values were elevated among many *Agnesiella* species, indicating substantial gene flow between them (Figure S4). The *f*-branch statistic also provided evidence of gene flow between the outgroups and certain species of *Agnesiella* (Figure 4). Specifically, gene flow was detected between *L. antiqua* and *A. marginata* in the subgenus *Agnesiella*, as well as between *L. antiqua* and six species in the subgenus *Draberiella*. Additionally, *V. ogromna* showed gene flow with *A. lidia*. However, the estimated introgression fractions for these events were relatively low (Figure 4; Table S4). To evaluate the robustness of our introgression analyses against potential bias from 'ghost' lineage introgression (Pang & Zhang, 2023; Tricou et al., 2022), we intentionally removed seven species exhibiting introgression from several major clades and re-analysed the data. The results demonstrated that the removal of these species did not significantly change the original introgression findings (Figure S5).

Within *Agnesiella*, gene flow between the subgenera *Agnesiella* and *Draberiella* was limited. Within the subgenus *Agnesiella* (Clade I), gene flow among species was also relatively restricted. However, a small introgression fraction (2.11%) was detected between the most recent common ancestor (MRCA) of (*Agnesiella recurva*, (*Agnesiella juglandis*, (*Agnesiella* sp.1, *Agnesiella protensa*))) and *Agnesiella polita*. (Figure 4; Table S4). Additionally, *A. juglandis* and *A. protensa* exhibited a 4.73% introgression fraction.

In contrast, gene flow was more widespread and greater within subgenus *Draberiella*. For example, *A. ela* in Clade IV exhibited substantial introgression fractions (4.54%–9.67%) with multiple species

FIGURE 3 Divergence times of the genus *Agnesiella*. Nodes are annotated with average divergence time estimates, and blue bars indicate the 95% highest posterior density (HPD) intervals.

within Clade IV, as well as gene flow with species from other major clades (Figure 4, Table S4). Similarly, in Clade VI, the MRCA of (*A. elongata*, *A. singuliprotrusa*) exhibited high introgression fractions with species from Clades IV and V, up to 14.19% with *A. irma*. Additionally, *A. xantha* and *A. biprotrusa* in Clade VI also displayed gene flow with multiple species (Figure 4, Table S4).

The QuIBL analyses further supported widespread gene flow within *Agnesiella*, consistent with the *f*-branch statistic results (Figure 5a). The estimated average introgression fractions among species were relatively low. This likely reflects the inherent limitations of using short, single-sequence fragments for gene tree construction in the QuIBL analyses, which yielded fewer informative sites per gene tree. Consequently, this constraint on gene tree estimation may have reduced the sensitivity of the QuIBL method, potentially leading to conservative estimates of introgression (Ding et al., 2022).

Further analysis using the NANUQ algorithm, the T3 model, and the TINNiK algorithm revealed a significant number of quartet concordance factors (qcCFs) deviating from the expected three-line segment patterns under a strictly tree-like model (Figure S6). These deviations suggest the potential influence of reticulate evolutionary processes, such as hybridisation or introgression, within the analysed dataset. The resulting simplex plots highlighted multifurcations and robust signals of reticulation. The neighbour-net splits graph revealed high reticulation along the backbone of the phylogeny, particularly within the subgenus *Draberiella*, while less reticulation was observed at shallower nodes (Figure 5b).

Functional annotation and enrichment analysis

Our analysis identified 1455 sequences exhibiting significant introgression signals. Of these introgressed sequences, 226 loci were used in the initial phylogenetic analyses, including 60 SCOs loci and 166 UCEs loci. Functional annotation of the 1455 introgressed sequences revealed that 206 were associated with Gene Ontology (GO) terms, and 254 were mapped to KEGG pathways (Table S5). GO enrichment analysis of genes mapped to KEGG pathways demonstrated a significant overrepresentation of terms related to metabolic processes (Figure 6). Specifically, these significantly overrepresented GO terms included cellular metabolic process (GO:0044237), nitrogen compound metabolic process (GO:0006807), and organic substance metabolic process (GO:0071704). Significant enrichment was also observed for terms associated with organelle lumen organisation and function (e.g. membrane-enclosed lumen [GO:0031974], organelle lumen

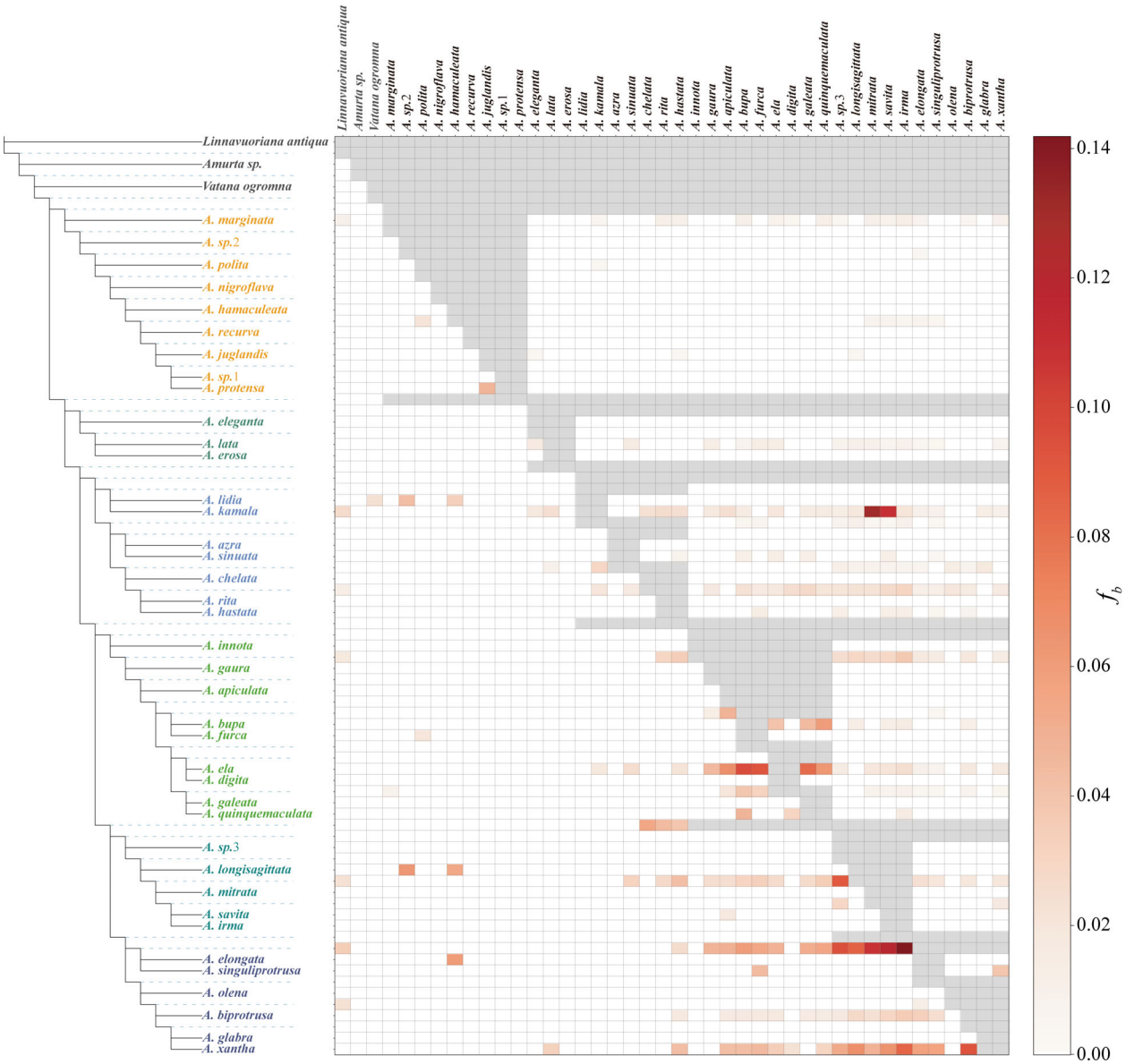

FIGURE 4 Pairwise *f*-branch quantification of introgression between branches. Pairwise *f*-branch statistic calculated using Dsuite to quantify the fraction of introgression between populations or species. Analysis is based on quality-filtered SNP data and the species tree topology from the 'SCOs_80_UCEs_100_NT_CID_800' dataset following the removal of significant introgression loci.

[GO:0043233], intracellular organelle lumen [GO:0070013]) and signalling/communication pathways (e.g. cell–cell signalling [GO:0007154] and protein binding [GO:0005515]). These enriched metabolic processes encompass functions in the digestion, detoxification and metabolism of plant secondary metabolites, exemplified by terms such as heterocycle metabolic process (GO:0046483), cellular aromatic compound metabolic process (GO:0006725), and organic acid metabolic process (GO:0006082) (Table S5). Notably, the detection of strong introgression signals in several Cytochrome P450 genes, known to be involved in detoxifying plant secondary metabolites (Lu, et al., 2021), suggests that interspecific gene flow may have facilitated adaptation to host plant defences in *Agnesiella*.

DISCUSSION

Phylogenetic relationships

Relationships among genera of the *Linnavuoriana* complex were well resolved by our phylogenetic analyses. Genus *Agnesiella* is monophyletic and sister to *Vatana*. Within *Agnesiella*, we confirmed the monophyly of its two subgenera, and inter-specific relationships were also well resolved overall despite extensive gene tree conflict. Five main clades were recovered within the subgenus *Draberiella*. These results are also consistent with some morphological characters that are apparently synapomorphies, such as the dorsally extended pygofer processes in Clade II and the predominantly bifurcate pygofer

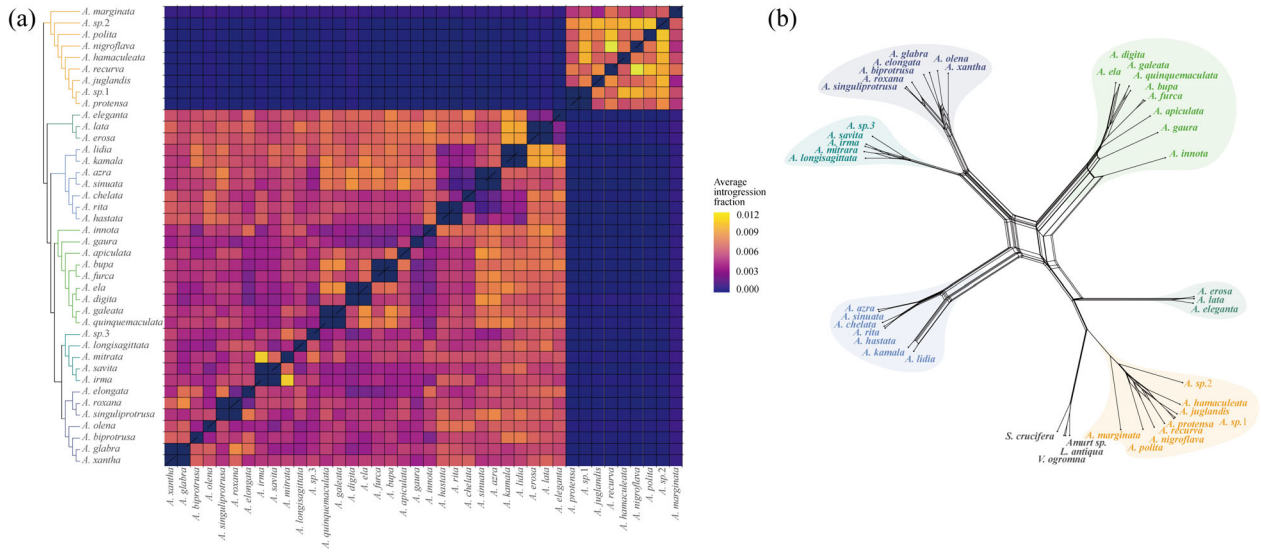

FIGURE 5 Estimating and visualising reticulate evolution: QuIBL and Neighbour-Net analyses. (a) Average introgression fractions inferred for each species pair using the QuIBL analysis; (b) Phylogenetic network (Neighbour-Net splits graph) based on the NANUQ quartet distance matrix.

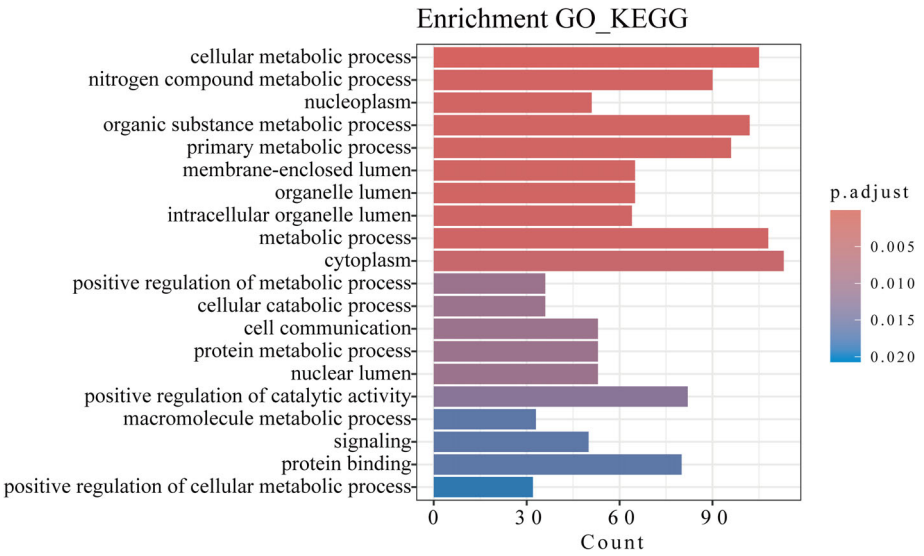

FIGURE 6 Functional enrichment of the significant introgression sequences. Detailed functional annotation results of the significant introgression sequences, performed using EggNOG-mapper, are provided in Table S5.

processes in Clade III (Wang et al., 2022, 2024; Yan & Yang, 2016, 2019). However, our results also revealed widespread phylogenetic incongruence, which we attribute to a variety of biological and non-biological factors (Steenwyk et al., 2023). This highlights the need to consider more complex evolutionary scenarios beyond strict bifurcation.

The sensitivity of the resulting tree topologies obtained when using different dataset filtering strategies, phylogenetic analysis methods and substitution models highlights the significant impact of confounding factors on phylogenetic inference in *Agnesiella* (Tables 1 and S3). The ABS filtering strategy, in particular, significantly influenced the topologies of major clades. This effect is likely attributable to the removal of gene trees derived from loci exhibiting high heterogeneity or weak phylogenetic signal, which persisted despite prior filtering steps. The presence of such loci can compromise the robustness of gene tree inference (Singhal et al., 2021), and summary coalescent methods are known to be particularly sensitive to errors in gene tree estimation (Springer & Gatesy, 2016). Additionally, the use of different phylogenetic methods (ML and BI) and substitution models on the same dataset resulted in varying tree topologies, likely due to inherent differences in the approaches and biases in model selection (Hoff et al., 2016; Kapli et al., 2020; Susko & Roger, 2021).

In rapidly diversifying groups like *Agnesiella*, biological factors such as ILS may be particularly prevalent and pose challenges for molecular phylogenetic inference (Meleshko et al., 2021; Suh et al., 2015). The Phytop results suggest that ILS largely explains the

observed phylogenetic incongruence among gene trees (Figure S3). Notably, the phylogenetic incongruences observed between datasets derived from (UCEs) and single-copy orthologs (SCOs) in this investigation are most parsimoniously explained by ILS. UCEs, being primarily located in non-coding regions and under strong purifying selection (Bejerano et al., 2004; Snetkova et al., 2022), exhibit lower recombination rates and reduced copy number variation (Derti et al., 2006). The UCE sequences we used include flanking regions with higher variability, making them more suitable for resolving relationships among closely related species (Faircloth et al., 2012). In contrast, SCOs are less conserved and subject to diverse selective pressures, including recombination, which can introduce additional complexities.

Despite employing a coalescent-based approach to mitigate the effects of ILS (Zhang & Mirarab, 2022), some topological incongruence remained, notably within Clades IV and VI. While ILS is a plausible explanation for some of the observed phylogenetic discordance, we cannot exclude the potential contribution of other biological processes, such as hybridisation and introgression. Indeed, in groups experiencing recent and rapid diversification, such as the *Agnesiella*, ILS and introgression are likely to occur simultaneously (Edelman & Mallet, 2021). When ILS is prevalent, even low levels of interspecific gene flow can produce anomalous gene trees, resulting in discordance between gene trees and species trees (Jiao et al., 2020; Pang & Zhang, 2023).

Indeed, our analysis revealed widespread hybridisation and introgression among *Agnesiella* species, which further contributed to phylogenetic incongruence. The patterns we observed in these leafhoppers are similar to those observed in mosquitoes of the *Anopheles gambiae* complex, which hindered coalescent-based methods from accurately reconstructing their true evolutionary history (Fontaine et al., 2015). The *f*-branch statistic clearly demonstrated hybridisation and estimated varying levels of introgression, which appear to correlate with the degree of topological discordance. For example, the unstable placement of *A. hamaculeata* within the subgenus *Agnesiella* (Clade I) may be attributed to ancestral gene flow between *A. polita* and the MRCA of (*A. recurva*, *A. juglandis*, *A.* sp.*1* and *A. protensa*) (Figure 4). Similarly, the clustering of *A. juglandis* and *A. protensa* as sister species may be influenced by introgression between the two. The unstable placement of Clade III in the overall phylogeny is probably a consequence of extensive gene flow across Clades III, IV and V. Notably high introgression fractions involving *A. kamala* and *A. ela* (Figure 4) further support this, potentially also explaining the observed subdivision within Clade III (Figure S2). The unstable position of *A. ela* within Clade IV is likely driven by its high introgression fraction with multiple species. Clade VI exhibits the most pronounced topological instability (Figure S2), which is likely attributable to extensive gene flow, as indicated by substantial introgression fractions.

The potential influence of gene flow from unsampled 'ghost' lineages merits further consideration. Such cryptic introgression can significantly bias species tree inference and divergence time estimations (Pang & Zhang, 2023). Indeed, neglecting the impact of ghost lineages may lead to misinterpretations of significant introgression signals, a risk that is compounded when employing distantly related outgroups, which inherently increases the probability of error (Tricou et al., 2022). In this study, however, we mitigated this concern by selecting relatively closely related outgroups. Furthermore, our evaluation of the potential impact of ghost lineages suggests that their influence on our introgression inferences is likely limited. While heuristic approaches such as D-statistics, QuIBL and NANUQ are known to be insensitive to ghost introgression, Pang and Zhang (2024) advocate for the use of full-likelihood methods like Bayesian Phylogenetics and Phylogeography (BPP) to rigorously validate introgression events initially detected by heuristic methods. Nevertheless, the substantial computational demands of these full-likelihood approaches currently constrain their practical application to large-scale datasets.

Interspecific introgression and evolution of Agnesiella

The findings of this study do not conclusively demonstrate that hybridisation led to reproductive isolation and the formation of new hybrid species (Schumer et al., 2014), as was proposed for some Nearctic microleafhoppers (Ross, 1958). However, the evidence of phylogenetic incongruences and gene flow between species within the genus *Agnesiella* and closely related genera strongly indicates that introgression among morphologically well-characterised species has been common in this group (Schumer et al., 2018; Wang & Liu, 2025).

Under natural selection, beneficial genes introduced through introgression are more likely to be retained, leading to adaptive introgression that can enhance species fitness, facilitate the evolution of novel traits, and promote speciation (Martin & Jiggins, 2017). For instance, the hybrid butterfly species *Heliconius elevatus* exhibits widespread and ongoing gene flow with *H. pardalinus*, with the remaining 1% of its genome introgressed from *H. melpomene*, including genes responsible for wing colour patterns that play a key role in sexual selection and strengthen reproductive isolation (Rosser et al., 2019, 2024).

Functional annotation and enrichment analyses revealed that introgressed sequences in *Agnesiella* are associated with important metabolic and detoxification processes, as well as organelle function and signalling pathways (Figure 6). Notably, nitrogen compound metabolism and organic substance metabolism are likely to significantly influence the adaptability of *Agnesiella* species. Nitrogen metabolism is particularly critical for *Agnesiella* because their primary host plants, *Alnus* species, are the only actinorhizal genus within the Betulaceae family (Benson & Dawson, 2007). *Alnus* species form nodules housing symbiotic nitrogen-fixing *Frankia* bacteria, leading to the enrichment of parenchyma cells with nitrogen compounds (Ballhorn et al., 2017). Consequently, enhanced nitrogen metabolism, a trait that *Agnesiella* species may have acquired through adaptive introgression of relevant genes, is crucial for these *Alnus*-feeding herbivores. Other enriched metabolic processes are associated with the insect's capacity to process and detoxify plant secondary metabolites, such as heterocyclic and aromatic compounds (Table S5). These detoxification mechanisms are often mediated by enzymes like cytochrome P450s (Heidel-Fischer & Vogel, 2015; Lu et al., 2021). Furthermore, organelle

lumen organisation and function, along with signalling and communication processes, are also likely to influence nutrient acquisition from plants and energy metabolism (Jain & Zoncu, 2022). Such introgressed adaptations could enable *Agnesiella* species to rapidly adapt to their host plants or other environmental challenges, potentially contributing to speciation (Edelman & Mallet, 2021; Forbes et al., 2017).

The diversification of *Agnesiella* primarily occurred during the middle Miocene to late Pliocene, a period when the Hengduan Mountains, the main distribution area of the genus, experienced significant orogenic activity, river incision, sediment infilling and substantial climatic changes (Meng et al., 2016; Spicer et al., 2020; Sun et al., 2011; Wang et al., 2012). Previous studies indicate that orogenesis in the Hengduan Mountains during this epoch markedly accelerated diversification processes (Ding et al., 2020; Xing & Ree, 2017). These complex geological and climatic transformations continually generated novel ecological niches, fostering the formation of distinct 'sky island' environments (Ding et al., 2020; He & Jiang, 2014; Wang et al., 2012; Zhang, Ding, et al., 2019). Such conditions would have promoted the accumulation of lineage-specific genetic mutations (Spicer et al., 2020; Wu et al., 2022). Concurrently, climatic oscillations likely induced range expansions of previously isolated populations, leading to secondary contact and interspecific hybridisation (Mosbrugger et al., 2018; Wu et al., 2022). A recently proposed speciation mechanism, the mixing-isolation-mixing (MIM) cycle, posits that repeated phases of isolation and hybridisation can significantly accelerate species diversification (He et al., 2019). These MIM patterns are not unique to the Hengduan Mountains and are likely prevalent in similar mountainous regions worldwide (Igea & Tanentzap, 2021; Li et al., 2023; Ma et al., 2022; Xu & Hausdorf, 2021), suggesting that the diversification of *Agnesiella* may have been driven, in part, by the MIM mechanism. Intriguingly, *Alnus* species, the primary host plants of *Agnesiella*, are estimated to have diversified from the early Oligocene to the middle Miocene, with both genera exhibiting overlapping geographic distributions during this period (Chen & Li, 2004; Wu et al., 2023; Yang et al., 2022). This temporal and spatial overlap suggests that the complex orogenic and climatic changes in the Hengduan Mountains may have also promoted the diversification and dispersal of *Alnus* plants, further contributing to the species diversification of *Agnesiella*.

CONCLUSIONS

This study provides the first genome-wide analysis of hybridisation and introgression in leafhoppers (Cicadellidae). Integrating comprehensive genomic datasets with diverse analytical methods, we demonstrated significant gene flow among species within the genus *Agnesiella*. Phylogenetic analyses reveal a complex evolutionary history for *Agnesiella*, shaped by ILS and introgression. Species diversification within *Agnesiella* occurred primarily during the middle Miocene to late Pliocene, potentially driven by the complex orogenic processes and climatic changes associated with the Hengduan Mountains. Adaptive introgression likely played a crucial role in *Agnesiella* diversification, potentially facilitating adaptation to novel environmental pressures or host plant defences.

To further resolve *Agnesiella* phylogeny, future research should prioritise expanded taxon sampling, utilise higher-quality genomic data and the application of full-likelihood methods. This will more precisely elucidate the evolutionary history of *Agnesiella* and deepen our understanding of gene flow's role in insect adaptation, divergence and speciation.

AUTHOR CONTRIBUTIONS

Junjie Wang: Methodology; investigation; data curation; software; formal analysis; visualisation; validation; writing – original draft. Xian Zhou: Resources; validation; writing – review and editing. Christopher H. Dietrich: Resources; validation; writing – review and editing. Yanghui Cao: Conceptualisation; resources; validation; writing – review and editing. Min Huang: Conceptualisation; funding acquisition; resources; project administration; supervision; writing – review and editing.

ACKNOWLEDGEMENTS

This research was supported by the National Natural Science Foundation of China (32070478 and 32370493) and the Ministry of Science and Technology of the People's Republic of China (2015FY210300). We express our sincere appreciation to the faculty and students of the Entomological Museum of Northwest A&F University, with particular gratitude to Lu Lin and Xue Qingquan for their invaluable assistance in specimen acquisition, and acknowledge Hao Xiangyu, Ma Wanruo, Xiao Jintian, and Zhang Xuan for their insightful contributions to the conceptual framework of this research. Furthermore, we extend our gratitude to Prof. Zhang Feng and Dr. Du Shiyu of Nanjing Agricultural University for their methodological guidance.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Raw sequence data generated by this study have been deposited in the Sequence Read Archive (SRA) database of GenBank under the BioProjects PRJNA1226749. SRA accession numbers are provided in Table S1. Phylogenetic trees generated from the study are available on GitHub at https://github.com/WJJ-97/Phylogenomics/tree/main/Agnesiella_Phylogenomics/Phylogenetic%20Trees.

ORCID

Junjie Wang https://orcid.org/0000-0002-6611-1075
Xian Zhou https://orcid.org/0000-0001-5907-3464
Christopher H. Dietrich https://orcid.org/0000-0003-4005-4305
Yanghui Cao https://orcid.org/0000-0002-0515-0767
Min Huang https://orcid.org/0000-0001-7621-4863

REFERENCES

Aguillon, S.M., Dodge, T.O., Preising, G.A. & Schumer, M. (2022) Introgression. *Current Biology*, 32(16), R865–R868.

Allman, E.S., Baños, H., Mitchell, J.D. & Rhodes, J.A. (2024) TINNiK: inference of the tree of blobs of a species network under the coalescent model. *Algorithms for Molecular Biology*, 19(1), 23.

Allman, E.S., Baños, H. & Rhodes, J.A. (2019) NANUQ: a method for inferring species networks from gene trees under the coalescent model. *Algorithms for Molecular Biology*, 14, 1–25.

Allman, E.S., Mitchell, J.D. & Rhodes, J.A. (2022) Gene tree discord, simplex plots, and statistical tests under the coalescent. *Systematic Biology*, 71(4), 929–942.

Álvarez-Carretero, S., Goswami, A., Yang, Z. & Dos Reis, M. (2019) Bayesian estimation of species divergence times using correlated quantitative characters. *Systematic Biology*, 68(6), 967–986.

Ballhorn, D.J., Elias, J.D., Balkan, M.A., Fordyce, R.F. & Kennedy, P.G. (2017) Colonization by nitrogen-fixing *Frankia* bacteria causes short-term increases in herbivore susceptibility in red alder (*Alnus rubra*) seedlings. *Oecologia*, 184, 497–506.

Banos, H., Wong, T.K., Daneau, J., Susko, E., Minh, B.Q., Lanfear, R. et al. (2024) GTRpmix: a linked general time-reversible model for profile mixture models. *Molecular Biology and Evolution*, 41(9), msae174.

Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S. et al. (2004) Ultraconserved elements in the human genome. *Science*, 304(5675), 1321–1325.

Benson, D.R. & Dawson, J.O. (2007) Recent advances in the biogeography and genecology of symbiotic *Frankia* and its host plants. *Physiologia Plantarum*, 130(3), 318–330.

Brůna, T., Hoff, K.J., Lomsadze, A., Stanke, M. & Borodovsky, M. (2021) BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics and Bioinformatics*, 3(1), lqaa108.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. et al. (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, 10, 1–9.

Cantalapiedra, C.P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. (2021) eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Molecular Biology and Evolution*, 38(12), 5825–5829.

Cao, Y., Dietrich, C.H., Kits, J.H., Dmitriev, D.A., Richter, R., Eyres, J. et al. (2023) Phylogenomics of microleafhoppers (Hemiptera: Cicadellidae: Typhlocybinae): morphological evolution, divergence times, and biogeography. *Insect Systematics and Diversity*, 7(4), 1. Available from: https://doi.org/10.1093/isd/ixad010

Capella-Gutiérrez, S., Silla-Martínez, J.M. & Gabaldón, T. (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, 25(15), 1972–1973.

Chen, S. (2023) Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta*, 2(2), e107.

Chen, Z. & Li, J. (2004) Phylogenetics and biogeography of *Alnus* (Betulaceae) inferred from sequences of nuclear ribosomal DNA ITS region. *International Journal of Plant Sciences*, 165(2), 325–335.

Crotty, S.M., Minh, B.Q., Bean, N.G., Holland, B.R., Tuke, J., Jermiin, L.S. et al. (2020) GHOST: recovering historical signal from heterotachously evolved sequence alignments. *Systematic Biology*, 69(2), 249–264.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A. et al. (2011) The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158. Available from: https://doi.org/10.1093/bioinformatics/btr330

Derti, A., Roth, F.P., Church, G.M. & Wu, C.T. (2006) Mammalian ultraconserved elements are strongly depleted among segmental duplications and copy number variants. *Nature Genetics*, 38(10), 1216–1220.

Dietrich, C.H. (2013) South American leafhoppers of the tribe Typhlocybini (Hemiptera: Cicadellidae: Typhlocybinae). *Zoologia (Curitiba)*, 30, 519–568.

Ding, W.N., Ree, R.H., Spicer, R.A. & Xing, Y.W. (2020) Ancient orogenic and monsoon-driven assembly of the world's richest temperate alpine flora. *Science*, 369(6503), 578–581.

Ding, Y., Ji, F.D. & Huang, H.T. (2022) Methods of Detecting Introgression after Species-tree Construction. *Bio*, 101, e1010678. Available from: https://doi.org/10.21769/BioProtoc.1010678

Dmitriev, D.A., Angelova, R., Anufriev, G.A., Bartlett, C.R., Blanco-Rodríguez, E., Borodin, O.I. et al. (2022) World Auchenorrhyncha Database. TaxonPages https://hoppers.speciesfile.org (Retrieved on Mar. 12, 2025).

Dmitriev, D.A. & Dietrich, C.H. (2010) Review of the New World Erythroneurini (Hemiptera: Cicadellidae: Typhlocybinae). IV. Genus *Eratoneura*. *Illinois Natural History Survey Bulletin*, 39(3), 79–258.

Edelman, N.B., Frandsen, P.B., Miyagi, M., Clavijo, B., Davey, J., Dikow, R.B. et al. (2019) Genomic architecture and introgression shape a butterfly radiation. *Science*, 366(6465), 594–599. Available from: https://doi.org/10.1126/science.aaw2090

Edelman, N.B. & Mallet, J. (2021) Prevalence and adaptive impact of introgression. *Annual Review of Genetics*, 55(1), 265–283.

Emms, D.M. & Kelly, S. (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology*, 20, 1–14.

Faircloth, B.C. (2016) PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics*, 32(5), 786–788.

Faircloth, B.C., McCormack, J.E., Crawford, N.G., Harvey, M.G., Brumfield, R.T. & Glenn, T.C. (2012) Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology*, 61(5), 717–726.

Fleming, J.F. & Struck, T.H. (2023) nRCFV: a new, dataset-size-independent metric to quantify compositional heterogeneity in nucleotide and amino acid datasets. *BMC Bioinformatics*, 24(1), 145.

Fontaine, M.C., Pease, J.B., Steele, A., Waterhouse, R.M., Neafsey, D.E., Sharakhov, I.V. et al. (2015) Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science*, 347(6217), 1258524.

Forbes, A.A., Devine, S.N., Hippee, A.C., Tvedte, E.S., Ward, A.K., Widmayer, H.A. et al. (2017) Revisiting the particular role of host shifts in initiating insect speciation. *Evolution*, 71(5), 1126–1137.

Fu, J.Y., Han, B.Y. & Xiao, Q. (2014) Mitochondrial COI and 16sRNA evidence for a single species hypothesis of *E. Vitis*, *J. Formosana* and *E. Onukii* in East Asia. *PLoS One*, 9(12), e115259.

Goodchild, A.J.P. (1966) Evolution of the alimentary canal in the Hemiptera. *Biological Reviews*, 41(1), 97–139.

Grant, P.R. & Grant, B.R. (2019) Hybridization increases population variation during adaptive radiation. *Proceedings of the National Academy of Sciences*, 116(46), 23216–23224.

Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M. et al. (2010) A draft sequence of the Neandertal genome. *Science*, 328(5979), 710–722. Available from: https://doi.org/10.1126/science.1188021

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.

He, K. & Jiang, X. (2014) Sky islands of southwest China. I: an overview of phylogeographic patterns. *Chinese Science Bulletin*, 59, 585–597.

He, Z., Li, X., Yang, M., Wang, X., Zhong, C., Duke, N.C. et al. (2019) Speciation with gene flow via cycles of isolation and migration: insights from multiple mangrove taxa. *National Science Review*, 6(2), 275–288. Available from: https://doi.org/10.1093/nsr/nwy078

Heidel-Fischer, H.M. & Vogel, H. (2015) Molecular mechanisms of insect adaptation to plant secondary compounds. *Current Opinion in Insect Science*, 8, 8–14.

Hibbins, M.S. & Hahn, M.W. (2022) Phylogenomic approaches to detecting and characterizing introgression. *Genetics*, 220(2), iyab173.

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.

Hoff, M., Orf, S., Riehm, B., Darriba, D. & Stamatakis, A. (2016) Does the choice of nucleotide substitution models matter topologically? *BMC Bioinformatics*, 17, 1–13.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H. et al. (2019) eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, 47(D1), D309–D314. Available from: https://doi.org/10.1093/nar/gky1085

Huson, D.H. & Bryant, D. (2024) The SplitsTree app: interactive analysis and visualization using phylogenetic trees and networks. *Nature Methods*, 21(10), 1773–1774.

Igea, J. & Tanentzap, A.J. (2021) Global topographic uplift has elevated speciation in mammals and birds over the last 3 million years. *Nature Ecology & Evolution*, 5(11), 1530–1535.

Jain, A. & Zoncu, R. (2022) Organelle transporters and inter-organelle communication as drivers of metabolic regulation and cellular homeostasis. *Molecular Metabolism*, 60, 101481.

Jiao, X., Flouri, T., Rannala, B. & Yang, Z. (2020) The impact of cross-species gene flow on species tree estimation. *Systematic Biology*, 69(5), 830–847.

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

Kapli, P., Yang, Z. & Telford, M.J. (2020) Phylogenetic tree building in the genomic age. *Nature Reviews Genetics*, 21(7), 428–444.

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.

Keilwagen, J., Hartung, F. & Grau, J. (2019) GeMoMa: homology-based gene prediction utilizing intron position conservation and RNA-seq data. *Gene Prediction: Methods and Protocols*, 1962, 161–177.

Koppetsch, T., Malinsky, M. & Matschiner, M. (2024) Towards reliable detection of introgression in the presence of among-species rate variation. *Systematic Biology*, 73(5), syae028. Available from: https://doi.org/10.1093/sysbio/syae028

Krishnankutty, S.M., Rakitov, R. & Dietrich, C.H. (2015) Taxonomy and phylogeny of the north American leafhopper genus *Cuerna* (Hemiptera: Cicadellidae). *Annals of the Entomological Society of America*, 108(3), 339–371.

Kück, P. & Longo, G.C. (2014) FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. *Frontiers in Zoology*, 11, 1–8.

Lamichhaney, S., Berglund, J., Almén, M.S., Maqbool, K., Grabherr, M., Martinez-Barrio, A. et al. (2015) Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*, 518(7539), 371–375.

Letunic, I. & Bork, P. (2024) Interactive tree of life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Research*, 52(W1), gkae268. Available from: https://doi.org/10.1093/nar/gkae268

Li, J., Li, Q., Wu, Y., Ye, L., Liu, H., Wei, J. et al. (2021) Mountains act as museums and cradles for hemipteran insects in China: evidence from patterns of richness and phylogenetic structure. *Global Ecology and Biogeography*, 30(5), 1070–1085.

Li, J., Shi, L., Vasseur, L., Zhao, Q., Chen, J., You, M. et al. (2022) Genetic analyses reveal regional structure and demographic expansion of the predominant tea pest *Empoasca onukii* (Hemiptera: Cicadellidae) in China. *Pest Management Science*, 78(7), 2838–2850.

Li, Y.F., Wang, S.J., Zhou, J.Y., Gao, C.Q., Zheng, C.G., Xue, H.J. et al. (2023) Integrative taxonomy of the stalk-eyed bug genus *Chauliops* (Heteroptera: Malcidae: Chauliopinae) reveals orogeny-driven speciation. *Journal of Systematics and Evolution*, 61(5), 932–947.

Lu, K., Song, Y. & Zeng, R. (2021) The role of cytochrome P450-mediated detoxification in insect adaptation to xenobiotics. *Current Opinion in Insect Science*, 43, 103–107.

Lu, L., Dietrich, C.H., Cao, Y. & Zhang, Y. (2021) A multi-gene phylogenetic analysis of the leafhopper subfamily Typhlocybinae (Hemiptera: Cicadellidae) challenges the traditional view of the evolution of wing venation. *Molecular Phylogenetics and Evolution*, 165, 107299.

Ma, Y., Mao, X., Wang, J., Zhang, L., Jiang, Y., Geng, Y. et al. (2022) Pervasive hybridization during evolutionary radiation of *rhododendron* subgenus *Hymenanthes* in mountains of southwest China. *National Science Review*, 9(12), nwac276. Available from: https://doi.org/10.1093/nsr/nwac276

Malinsky, M., Matschiner, M. & Svardal, H. (2021) Dsuite-fast D-statistics and related admixture evidence from VCF files. *Molecular Ecology Resources*, 21(2), 584–595.

Malinsky, M., Svardal, H., Tyers, A.M., Miska, E.A., Genner, M.J., Turner, G.F. et al. (2018) Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nature Ecology & Evolution*, 2(12), 1940–1955.

Mallet, J. (2005) Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, 20(5), 229–237.

Mallet, J., Besansky, N. & Hahn, M.W. (2016) How reticulated are species? *BioEssays*, 38(2), 140–149.

Manni, M., Berkeley, M.R., Seppey, M., Simão, F.A. & Zdobnov, E.M. (2021) BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Molecular Biology and Evolution*, 38(10), 4647–4654.

Marder, E., Smiley, T.M., Yanites, B.J. & Kravitz, K. (2025) Direct effects of mountain uplift and topography on biodiversity. *Science*, 387(6740), 1287–1291.

Martin, S.H. & Jiggins, C.D. (2017) Interpreting the genomic landscape of introgression. *Current Opinion in Genetics & Development*, 47, 69–74.

Mavárez, J., Salazar, C.A., Bermingham, E., Salcedo, C., Jiggins, C.D. & Linares, M. (2006) Speciation by hybridization in *Heliconius* butterflies. *Nature*, 441(7095), 868–871.

Meier, J.I., Marques, D.A., Mwaiko, S., Wagner, C.E., Excoffier, L. & Seehausen, O. (2017) Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature Communications*, 8(1), 14363.

Meleshko, O., Martin, M.D., Korneliussen, T.S., Schröck, C., Lamkowski, P., Schmutz, J. et al. (2021) Extensive genome-wide phylogenetic discordance is due to incomplete lineage sorting and not ongoing introgression in a rapidly radiated bryophyte genus. *Molecular Biology and Evolution*, 38(7), 2750–2766. Available from: https://doi.org/10.1093/molbev/msab063

Meng, K., Wang, E. & Wang, G. (2016) Uplift of the Emei Shan, western Sichuan basin: implication for eastward propagation of the Tibetan plateau in early Miocene. *Journal of Asian Earth Sciences*, 115, 29–39.

Merrill, R.M., Naisbit, R.E., Mallet, J. & Jiggins, C.D. (2013) Ecological and genetic factors influencing the transition between host-use strategies in sympatric *Heliconius* butterflies. *Journal of Evolutionary Biology*, 26(9), 1959–1967.

Mi, X., Feng, G., Hu, Y., Zhang, J., Chen, L., Corlett, R.T. et al. (2021) The global significance of biodiversity science in China: an overview. *National Science Review*, 8(7), nwab032. Available from: https://doi.org/10.1093/nsr/nwab032

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.

Mo, Y.K., Lanfear, R., Hahn, M.W. & Minh, B.Q. (2023) Updated site concordance factors minimize effects of homoplasy and taxon sampling.

Bioinformatics, 39(1), btac741. Available from: https://doi.org/10.1093/bioinformatics/btac741

Morgulis, A., Gertz, E.M., Schäffer, A.A. & Agarwala, R. (2006) WindowMasker: window-based masker for sequenced genomes. *Bioinformatics*, 22(2), 134–141.

Mosbrugger, V., Favre, A., Muellner-Riehl, A.N., Päckert, M. & Mulch, A. (2018) Cenozoic evolution of geo-biodiversity in the Tibeto-Himalayan region. *Mountains, Climate, and Biodiversity*, 429, 448.

Nickel, H. & Remane, R. (2002) Artenliste der Zikaden Deutschlands, mit Angabe von Nährpflanzen, Nahrungsbreite, Lebenszyklus, Areal und Gefährdung. *Beiträge Zur Zikadenkunde*, 5, 27–64.

Pang, X.X. & Zhang, D.Y. (2023) Impact of ghost introgression on coalescent-based species tree inference and estimation of divergence time. *Systematic Biology*, 72(1), 35–49.

Pang, X.X. & Zhang, D.Y. (2024) Detection of ghost introgression requires exploiting topological and branch length information. *Systematic Biology*, 73(1), 207–222.

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D.L., Darling, A., Höhna, S. et al. (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, 61(3), 539–542. Available from: https://doi.org/10.1093/sysbio/sys029

Ross, H.H. (1958) Evidence suggesting a hybrid origin for certain leafhopper species. *Evolution*, 12(3), 337–346. Available from: https://doi.org/10.2307/2405855

Rosser, N., Queste, L.M., Cama, B., Edelman, N.B., Mann, F., Mori Pezo, R. et al. (2019) Geographic contrasts between pre-and postzygotic barriers are consistent with reinforcement in *Heliconius* butterflies. *Evolution*, 73(9), 1821–1838. Available from: https://doi.org/10.1111/evo.13804

Rosser, N., Seixas, F., Queste, L.M., Cama, B., Mori-Pezo, R., Kryvokhyzha, D. et al. (2024) Hybrid speciation driven by multilocus introgression of ecological traits. *Nature*, 628(8009), 1–7.

Rossi, M., Hausmann, A.E., Alcami, P., Moest, M., Roussou, R., Van Belleghem, S.M. et al. (2024) Adaptive introgression of a visual preference gene. *Science*, 383(6689), 1368–1373. Available from: https://doi.org/10.1126/science.adj9201

Salichos, L. & Rokas, A. (2013) Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature*, 497(7449), 327–331.

Sánchez, A.P., Pardo-Diaz, C., Enciso-Romero, J., Muñoz, A., Jiggins, C.D., Salazar, C. et al. (2015) An introgressed wing pattern acts as a mating cue. *Evolution*, 69(6), 1619–1629.

Sayyari, E. & Mirarab, S. (2016) Fast coalescent-based computation of local branch support from quartet frequencies. *Molecular Biology and Evolution*, 33(7), 1654–1668.

Schumer, M., Rosenthal, G.G. & Andolfatto, P. (2014) How common is homoploid hybrid speciation? *Evolution*, 68(6), 1553–1560.

Schumer, M., Rosenthal, G.G. & Andolfatto, P. (2018) What do we mean when we talk about hybrid speciation? *Heredity*, 120(4), 379–382.

Shang, H.Y., Jia, K.H., Li, N.W., Zhou, M.J., Yang, H., Tian, X.L. et al. (2024) Phytop: a tool for visualizing and recognizing signals of incomplete lineage sorting and hybridization using species trees output from ASTRAL. *Horticulture Research*, 12(3), uhae330. Available from: https://doi.org/10.1093/hr/uhae330

Shen, W., Sipos, B. & Zhao, L. (2024) SeqKit2: a swiss army knife for sequence and alignment processing. *iMeta*, 3(3), e191. Available from: https://doi.org/10.1002/imt2.191

Shen, X.X., Salichos, L. & Rokas, A. (2016) A genome-scale investigation of how sequence, function, and tree-based gene properties influence phylogenetic inference. *Genome Biology and Evolution*, 8(8), 2565–2580.

Shimodaira, H. (2002) An approximately unbiased test of phylogenetic tree selection. *Systematic Biology*, 51(3), 492–508.

Si Quang, L., Gascuel, O. & Lartillot, N. (2008) Empirical profile mixture models for phylogenetic reconstruction. *Bioinformatics*, 24(20), 2317–2323.

Singhal, S., Colston, T.J., Grundler, M.R., Smith, S.A., Costa, G.C., Colli, G.R. et al. (2021) Congruence and conflict in the higher-level phylogenetics of squamate reptiles: an expanded phylogenomic perspective. *Systematic Biology*, 70(3), 542–557. Available from: https://doi.org/10.1093/sysbio/syaa054

Smith, B.T., Mauck, W.M., III, Benz, B.W. & Andersen, M.J. (2020) Uneven missing data skew phylogenomic relationships within the lories and lorikeets. *Genome Biology and Evolution*, 12(7), 1131–1147.

Smith, M.R. (2020) Information theoretic generalized Robinson–Foulds metrics for comparing phylogenetic trees. *Bioinformatics*, 36(20), 5007–5013.

Snetkova, V., Pennacchio, L.A., Visel, A. & Dickel, D.E. (2022) Perfect and imperfect views of ultraconserved sequences. *Nature Reviews Genetics*, 23(3), 182–194.

Spicer, R.A., Farnsworth, A. & Su, T. (2020) Cenozoic topography, monsoons and biodiversity conservation within the Tibetan region: an evolving story. *Plant Diversity*, 42(4), 229–254.

Springer, M.S. & Gatesy, J. (2016) The gene tree delusion. *Molecular Phylogenetics and Evolution*, 94, 1–33.

Steenwyk, J.L., Buida, T.J., III, Labella, A.L., Li, Y., Shen, X.X. & Rokas, A. (2021) PhyKIT: a broadly applicable UNIX shell toolkit for processing and analyzing phylogenomic data. *Bioinformatics*, 37(16), 2325–2331.

Steenwyk, J.L., Li, Y., Zhou, X., Shen, X.X. & Rokas, A. (2023) Incongruence in the phylogenomics era. *Nature Reviews Genetics*, 24(12), 834–850.

Suh, A., Smeds, L. & Ellegren, H. (2015) The dynamics of incomplete lineage sorting across the ancient adaptive radiation of neoavian birds. *PLoS Biology*, 13(8), e1002224.

Sun, B.N., Wu, J.Y., Liu, Y.S.C., Ding, S.T., Li, X.C., Xie, S.P. et al. (2011) Reconstructing Neogene vegetation and climates to infer tectonic uplift in western Yunnan, China. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 304(3–4), 328–336.

Susko, E. & Roger, A.J. (2021) Long branch attraction biases in phylogenetics. *Systematic Biology*, 70(4), 838–843.

Suvorov, A., Kim, B.Y., Wang, J., Armstrong, E.E., Peede, D., D'agostino, E.R. et al. (2022) Widespread introgression across a phylogeny of 155 *drosophila* genomes. *Current Biology*, 32(1), 111–123.

Thawornwattana, Y., Seixas, F., Yang, Z. & Mallet, J. (2023) Major patterns in the introgression history of *Heliconius* butterflies. *eLife*, 12, RP90656.

Tricou, T., Tannier, E. & de Vienne, D.M. (2022) Ghost lineages highly influence the interpretation of introgression tests. *Systematic Biology*, 71(5), 1147–1158.

Van der Auwera, G.A. & O'Connor, B.D. (2020) *Genomics in the cloud: using docker, GATK, and WDL in Terra*. Sebastopol, CA: O'Reilly Media.

Vasimuddin, M., Misra, S., Li, H. & Aluru, S. (2019) Efficient architecture-aware acceleration of BWA-MEM for multicore systems. In: *2019 IEEE international parallel and distributed processing symposium (IPDPS)*. Rio de Janeiro, Brazil: IEEE, pp. 314–324.

Wang, E., Kirby, E., Furlong, K.P., Van Soest, M., Xu, G., Shi, X. et al. (2012) Two-phase growth of high topography in eastern Tibet during the Cenozoic. *Nature Geoscience*, 5(9), 640–645. Available from: https://doi.org/10.1038/ngeo1538

Wang, J., Liu, J., Zhang, Y. & Huang, M. (2024) Eight new species and four new records of the leafhopper genus *Agnesiella* Dworakowska (Hemiptera: Cicadellidae: Typhlocybinae) from China. *Zootaxa*, 5537(1), 49–75.

Wang, J., Zhang, Y. & Huang, M. (2022) Taxonomic study of the leafhopper genus *Agnesiella* Dworakowska (Hemiptera: Cicadellidae: Typhlocybinae) from China, with descriptions of 13 new species. *Zootaxa*, 5094(2), 201–233.

Wang, Z.F. & Liu, J.Q. (2025) Speciation studies in the genomic era. *Hereditas*, 47(1), 71–100.
Wood, D.E., Lu, J. & Langmead, B. (2019) Improved metagenomic analysis with kraken 2. *Genome Biology*, 20, 1–13.
Wu, S., Jun, W., Meng, Y. & Nie, Z.L. (2023) Fossil diversity of Betulaceae and its evolutionary history in the northern hemisphere. *Journal of Plant Resources and Environment*, 32(1), 77–88.
Wu, S., Wang, Y., Wang, Z., Shrestha, N. & Liu, J. (2022) Species divergence with gene flow and hybrid speciation on the Qinghai–Tibet plateau. *New Phytologist*, 234(2), 392–404.
Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z. et al. (2021) clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *The Innovation*, 1(2), 100141.
Xing, Y. & Ree, R.H. (2017) Uplift-driven diversification in the Hengduan Mountains, a temperate biodiversity hotspot. *Proceedings of the National Academy of Sciences*, 114(17), E3444–E3451.
Xu, J. & Hausdorf, B. (2021) Repeated hybridization increased diversity in the door snail complex *Charpentieria itala* in the southern Alps. *Molecular Phylogenetics and Evolution*, 155, 106982.
Xu, Y., Dietrich, C.H., Zhang, Y.L., Dmitriev, D.A., Zhang, L., Wang, Y.M. et al. (2021) Phylogeny of the tribe Empoascini (Hemiptera: Cicadellidae: Typhlocybinae) based on morphological characteristics, with reclassification of the *Empoasca* generic group. *Systematic Entomology*, 46(1), 266–286.
Yan, B., Dietrich, C.H., Yu, X., Jiao, M., Dai, R. & Yang, M. (2022) Mitogenomic phylogeny of Typhlocybinae (Hemiptera: Cicadellidae) reveals homoplasy in tribal diagnostic morphological traits. *Ecology and Evolution*, 12(6), e8982.
Yan, B. & Yang, M.F. (2016) Three new species and one new record of the leafhopper genus *Agnesiella* Dworakowska, 1970 (Hemiptera: Cicadellidae: Typhlocybinae: Typhlocybini) from China. *Zootaxa*, 4184(3), 529–540.
Yan, B. & Yang, M.F. (2019) Chinese species of the genus of *Agnesiella* Dworakowska (Hemiptera: Cicadellidae: Typhlocybini), with four new species and four new records. *Zootaxa*, 4565(2), 151–170.
Yang, Z., Ma, W., Yang, X., Wang, L., Zhao, T., Liang, L. et al. (2022) Plastome phylogenomics provide new perspective into the phylogeny and evolution of Betulaceae (Fagales). *BMC Plant Biology*, 22(1), 611. Available from: https://doi.org/10.1186/s12870-022-03991-1
Zhang, C. & Mirarab, S. (2022) Weighting by gene tree uncertainty improves accuracy of quartet-based species trees. *Molecular Biology and Evolution*, 39(12), msac215.
Zhang, D., Hao, G.Q., Guo, X.Y., Hu, Q.J. & Liu, J.Q. (2019) Genomic insight into "sky Island" species diversification in a mountainous biodiversity hotspot. *Journal of Systematics and Evolution*, 57(6), 633–645.
Zhang, F., Ding, Y., Zhu, C.D., Zhou, X., Orr, M.C., Scheu, S. et al. (2019) Phylogenomics from low-coverage whole-genome sequencing. *Methods in Ecology and Evolution*, 10(4), 507–517.
Zhao, Q., Shi, L., He, W., Li, J., You, S., Chen, S. et al. (2022) Genomic variations in the tea leafhopper reveal the basis of its adaptive evolution. *Genomics, Proteomics & Bioinformatics*, 20(6), 1092–1105. Available from: https://doi.org/10.1016/j.gpb.2022.05.011
Zhou, X., Lei, Y., Dietrich, C.H. & Huang, M. (2023) Investigating monophyly of Typhlocybini based on complete mitochondrial genomes with characterization and comparative analysis of 19 species (Hemiptera: Cicadellidae: Typhlocybinae). *Insects*, 14(11), 842.

SUPPORTING INFORMATION

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

Figure S1. Genome completeness assessments using BUSCO.

Figure S2. Phylogenetic topology hypotheses within each major clade. Squares represent distinct combinations of datasets and substitution models. Square colour indicates the *p*-value from the AU test. Details of the topologies of phylogenetic trees reconstructed from various datasets and methods are provided in Tables 1 and S3.

Figure S3. Phytop analysis of topological discordance. (a) 'SCOs_AA_80_CID' dataset; (b) 'SCOs_NT_80_CID' dataset; (c) 'UCEs_NT_100_CID' dataset; (d) 'SCOs_80_UCEs_100_NT_CID_800' dataset. *n* = number of gene trees; P = *p*-value of χ2 test (equality of topologies q2 and q3); ILS-i = ILS index; IH-i = IH index; ILS-e = proportion of incongruence explained by ILS; IH-e = proportion explained by IH.

Figure S4. Pairwise Patterson's *D* and f_4-ratios. (a) Pairwise Patterson's *D* statistic values, indicating the extent of gene flow. (b) Pairwise f_4-ratio values, quantifying relative admixture proportions. Colours represent the statistical support (light to dark) and the magnitude of both the *D*-statistic and f_4-Ratio (blue to red).

Figure S5. Pairwise *f*-branch quantification of introgression following lineages removal. Pairwise *f*-branch statistics were recalculated, using the same SNP data and species tree topology as in the original analysis, but with the modification of removing specific lineages. We removed one species each from the outgroups and each major clade—specifically, *Linnavuoriana antiqua*, *A. marginata*, *A. lata*, *A. sinuata*, A. furca, A. longisagittata, and *A. biprotrusa*—to assess the potential influence of 'ghost' lineage introgression on introgression analyses.

Figure S6. Simplex plots of qcCFs classification. (a) NANUQ simplex plot: qcCF classification into tree-like (blue circles), hybridisation-supporting (red triangles), and unresolved quartets (brown squares). Alpha (5e−12) and beta (0.05) thresholds used; (b) T3 model simplex plot: Refined distinction of tree-like (blue circles) and hybridisation-supporting qcCFs (red triangles) based on hypothesis testing; (c) TINNiK simplex plot: Depiction of B-quartets (brown symbols, blobs/reticulate regions) and T-quartets (green symbols, tree-like relationships). These simplex plots collectively visualise hybridisation signals, unresolved phylogenetic relationships, and tree-like structures within the species network, providing a comprehensive perspective on phylogenetic discordance.

Table S1. Specimens information.

Table S2. Genome assembly statistics and gene prediction summary.

Table S3. Dataset characteristics and topologies of coalescent-based species tree. Dataset characteristics include the number of loci, number of sites and average locus length. Species trees were constructed using ASTRAL-Hybrid under the MSC model.

Table S4. Pairwise *f*-branch statistic results.

Table S5. Functional annotation of the significant introgression sequences.

How to cite this article: Wang, J., Zhou, X., Dietrich, C.H., Cao, Y. & Huang, M. (2025) Extensive hybridisation and complex evolutionary history in the leafhopper genus *Agnesiella* (Hemiptera: Cicadellidae: Typhlocybinae). *Systematic Entomology*, 50(4), 903–919. Available from: https://doi.org/10.1111/syen.12686