Received: 6 April 2025 | Accepted: 5 December 2025

DOI: 10.1002/aps3.70058

APPLICATION ARTICLE

Applications in Plant Sciences

Ancient hybridization and phylogenetic discordance: Exploring evolutionary complexity in Asteraceae

Paige A. Ellestad[1] | **Erika R. Moore-Pollard**[1] | **Carolina M. Siniscalchi**[1,2] | **Ramhari Thapa**[1] | **Linda E. Watson**[3] | **J. Mauricio Bonifacino**[4] | **Jennifer R. Mandel**[1]

[1]Department of Biological Sciences, University of Memphis, Memphis, Tennessee 38152, USA

[2]General Libraries, Mississippi State University, Mississippi State, Mississippi 39762, USA

[3]Plant Biology and Ecology Department, Oklahoma State University, Stillwater, Oklahoma 74074, USA

[4]Laboratorio de Botánica, Facultad de Agronomía, Universidad de la República, Av. Garzón 780, Montevideo 12900, Uruguay

Correspondence

Carolina M. Siniscalchi, General Libraries, Mississippi State University, Mississippi State, Mississippi 39762, USA.
Email: csiniscalchi@library.msstate.edu

This article is part of the special issue "Branching out: Resolving plant evolution through phylogenetic networks."

Abstract

Premise: Conflicting phylogenetic signals are common in plant phylogenomics and often reflect evolutionary histories shaped by processes like hybridization, incomplete lineage sorting, and whole-genome duplication (WGD). We aimed to identify and assess these complex processes in the hyper-diverse family Asteraceae to offer insight into the underlying causes of phylogenetic discordance.

Methods: We used new and existing Hyb-Seq and transcriptome data to explore phylogenetic discordance by testing for nuclear/plastid incongruences, WGD, and reticulation. We present a tutorial detailing the execution of complex bioinformatic analyses to increase transparency, facilitate reproducibility, and support advancements in the field of plant evolution (https://github.com/erika-r-moore/Ellestad_etal_2025_APPS_Hybridizations).

Results: We uncovered extensive discordance among nuclear gene trees and deep reticulation events, particularly among South American lineages. Signals of WGD were found across the family but were often difficult to interpret, likely due to variation in data completeness, the complexity of the events, and their ancient origins.

Discussion: Our study and tutorial, along with a growing body of phylogenomic research, emphasize the role of reticulation and WGD in the evolution of large, diverse clades, while also underscoring the challenges. We anticipate continued advancements in theoretical approaches that will further enhance empirical studies in reticulate evolution.

KEYWORDS

ancient hybridization, bioinformatics pipeline, Compositae, plant evolution, reticulation, whole-genome duplication

Resumen

Premisa: Las señales filogenéticas conflictivas son comunes en la filogenómica de plantas, y a menudo, reflejan historias evolutivas moldeadas por procesos como la hibridación, la clasificación incompleta de linajes y la duplicación del genoma completo. Nuestro objetivo fue identificar y evaluar estos procesos complejos en la diversa familia Asteraceae, para ofrecer una perspectiva sobre las causas subyacentes de la discordancia filogenética.

Métodos: Utilizamos datos nuevos y existentes de Hyb-Seq y transcriptomas para explorar la discordancia filogenética mediante pruebas de incongruencia entre los

Paige A. Ellestad and Erika R. Moore-Pollard contributed equally to this work.

genomas nucleares y plastidiales, duplicación completa del genoma (WGD) y reticulación. Presentamos un tutorial que detalla la ejecución de análisis bioinformáticos complejos para aumentar la transparencia, facilitar la reproducibilidad y apoyar los avances en el campo de la evolución de las plantas (https://github.com/erika-r-moore/Ellestad_etal_2025_APPS_Hybridizations).

Resultados: Descubrimos una discordancia extensa entre los árboles genéticos nucleares y eventos profundos de reticulación, particularmente entre linajes sudamericanos. Se detectaron señales de WGD en toda la familia, aunque a menudo resultaron difíciles de interpretar, probablemente debido a la variación en la integridad de los datos, la complejidad de los eventos y su origen antiguo.

Discusión: Nuestro estudio y tutorial, junto con un cuerpo creciente de investigaciones filogenómicas, destacan el papel de la reticulación y de los WGD en la evolución de clados grandes y diversos, al mismo tiempo que subrayan los desafíos asociados. Anticipamos avances continuos en enfoques teóricos que potenciarán aún más los estudios empíricos sobre la evolución reticulada.

Advances in the generation and analysis of genomic data have provided deep insights into the evolutionary histories that constitute the tree of life. Phylogenetic analyses of big datasets, or phylogenomics, have led to a wealth of studies that describe complex patterns of species diversification and trait evolution in plants and provide insight into their historical biogeography. Recent studies have employed hundreds to thousands of loci to discern and describe the evolutionary history of clades of organisms. As large-scale phylogenomic studies grow in number, it is becoming increasingly apparent that conflicting evolutionary histories and phylogenetic discordance are commonplace, if not pervasive (Roch and Warnow, 2015; Smith et al., 2015; Edwards et al., 2016; Hosner et al., 2016; Lanfear and Hahn, 2024). Thus, the acquisition of more data has often led to the discovery of even greater complexity, and it is often difficult to resolve patterns of evolution because these studies regularly report major differences in the topologies of individual loci (gene tree conflict) and cytonuclear discordance (Salichos and Rokas, 2013; Mendes and Hahn, 2016). Several biological processes, as well as methodological limitations, may play a role, including differential gene loss and gene retention following large-scale duplications (Wood et al., 2009; Barker et al., 2016a, 2016b), reticulate evolution (often ancient), gene flow, incomplete lineage sorting (ILS), difficulty in assigning orthology, and noise in data assembly and filtering. Given these limitations, an examination of phylogenomic data using various analytical tools across multiple phylogenetic scales to discern evolutionary patterns is necessary to tackle these complex histories. This is especially important in angiosperms, where studies have demonstrated that processes such as polyploidy and hybridization may be coupled and associated with the evolution of key traits and their evolutionary success (e.g., Schranz et al., 2012; Edger et al., 2015; Tank et al., 2015; Soltis and Soltis, 2016; Stull et al., 2023).

The genomes of angiosperms have repeatedly experienced duplication and subsequent re-diploidization, resulting in diploidized paleopolyploid genomes (Schranz et al., 2012; Edger et al., 2015; Tank et al., 2015; Soltis and Soltis, 2016; Landis et al., 2018; Li et al., 2021; Stull et al., 2023). In comparative studies of transcriptomes from plant lineages, large-scale gene duplications are often present in regions where the most phylogenetic discordance among datasets exists (One Thousand Plant Transcriptomes Initiative, 2019). Differential loss and retention of genes across lineages (i.e., fractionation) may lead to widespread discordance among phylogenetic analyses but may also be a driver of differential diversification among lineages. Empirical analyses indicate that increased taxon sampling (Lambert et al., 2015) and the use of full-length genes (Hosner et al., 2016) may improve phylogenetic resolution.

Gene flow has long been recognized as a fundamental driver of diversity, and hybridization and introgression can contribute to the evolution of novel phenotypes, commonly contributing to the speciation process in plants (Rieseberg, 1995; Arnold, 2007; Givnish, 2010; Ellstrand, 2014). The processes of hybridization and introgression may be particularly powerful forces in generating phylogenetic discordance, both among nuclear genes and between nuclear and plastid genomes (i.e., cytonuclear discordance). Interpreting these influences across ancient evolutionary scales has been especially difficult. However, recent methods that estimate phylogenetic networks, while simultaneously accounting for ILS and hybridization, have improved our ability to understand the extent of these two processes on plant evolution (Than et al., 2008; Solís-Lemus et al., 2017; Wen et al., 2018).

The sunflower family, Asteraceae or Compositae, is a hyper-diverse family of flowering plants comprising more than 30,000 described species (Figure 1). The family has a deep and widespread history of large-scale gene duplications (Barker et al., 2008, 2016b; Semple and Watanabe, 2009; Huang et al., 2016) and hybridization (Walter et al., 2020; White et al., 2020; Hernández et al., 2024). Hybridization events often produce intermediate or mixed morphological traits, leaving signatures that may persist over long evolutionary timescales. Such ancient reticulation may be tied to whole-genome duplication (WGD) events; however, subsequent diploidization makes the signals difficult to trace across ancient timescales.

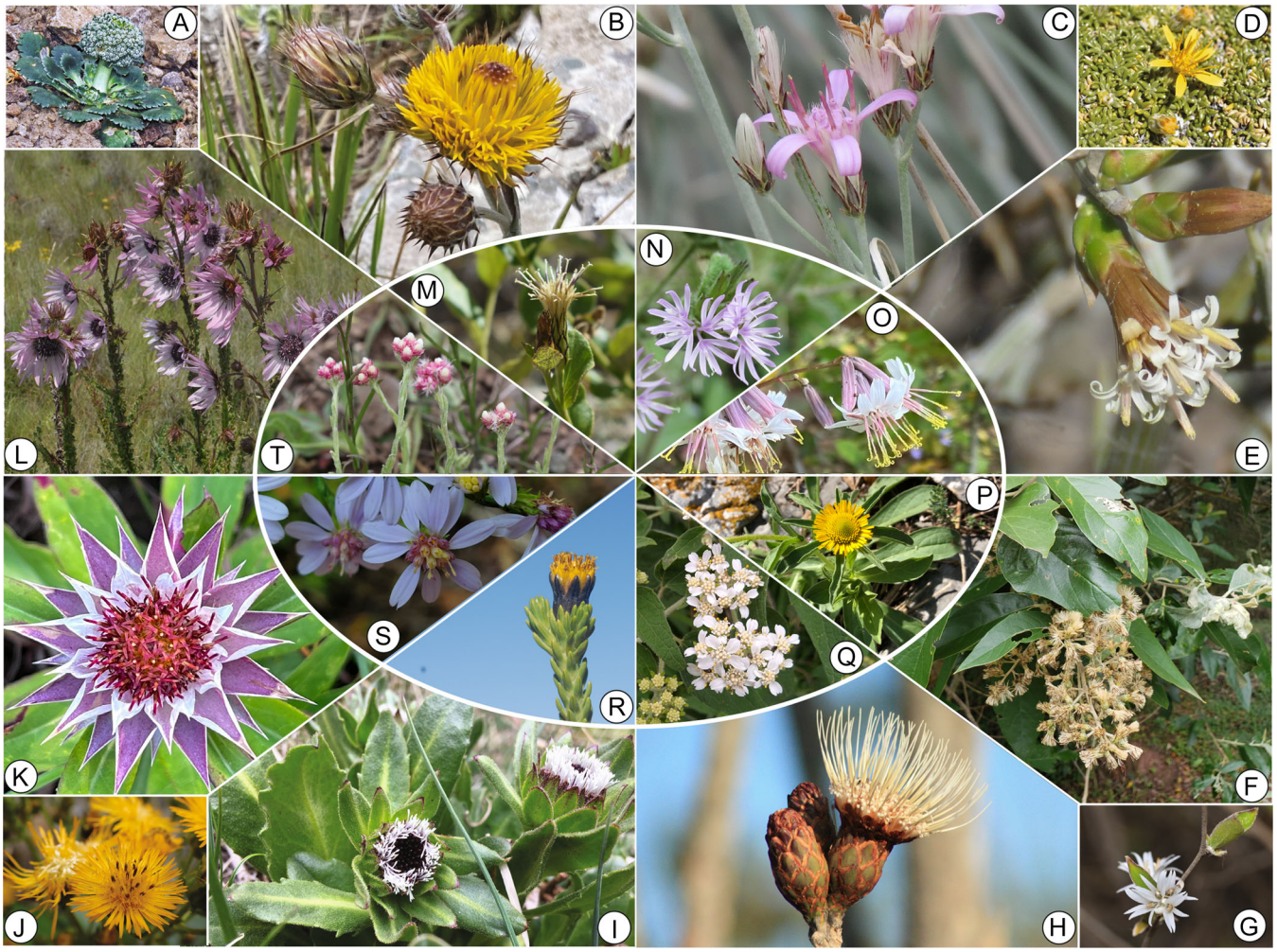

FIGURE 1 Diversity of Asterales (photo credits in square brackets). (A) Calyceraceae, *Gamocarpha macrocephala* [Hugo Hulsberg, CC-BY]. (B–T) Asteraceae. (B) Barnadesieae, *Schlechtendalia luzulifolia* [Martin Coronel Varela, CC-BY-NC]; (C) Hyalideae, *Hyalis argentea* [Gonzalo Roget, CC-BY]; (D) Mutisieae, *Brachyclados caespitosus* [E. Santos Ortega, CC-BY]; (E) *Cyclolepis genistoides* [Ariadna Tripaldi, CC-BY]; (F) Gochnatieae, *Moquiniastrum polymorphum* [C. M. Siniscalchi]; (G) Pertyeae, *Myripnois dioica* [koubai, CC-BY-NC]; (H) Wunderlichieae, *Wunderlichia crulsiana* [C. M. Siniscalchi]; (I) Platycarpheae, *Cavea tanguensis* [Joe Black, CC-BY-NC]; (J) Liabeae, *Erato polymnioides* [Skjold Søndergaard, CC-BY]; (K) Dicomeae, *Macledium zeyheri* [Heinrich Human, CC-BY-NC]; (L) Arctotideae, *Berkheya purpurea* [Vicki Funk]; (M) Cardueae, *Shangwua jacea* [Arun N, CC-BY-NC]; (N) Vernonieae, *Elephantopus tomentosus* [jfox16, CC-BY]; (O) Cichorieae, *Nabalus albus* [biomania, CC-BY-NC]; (P) Inuleae, *Pallenis spinosa* [Pavel Kacl, CC-BY]; (Q) Heliantheae, *Montanoa tomentosa* [Jordi Martínez Martínez, CC-BY]; (R) Senecioneae, *Xenophyllum lycopodioides* [Vicki Funk]; (S) Asteraeae, *Symphyotrichum drummondii* [C. M. Siniscalchi]; (T) Gnaphalieae, *Antennaria rosea* [R. Thapa].

Based on genetic and morphological evidence, numerous hypotheses have been proposed for ancient hybridization processes in Asteraceae, including in the sunflower genus *Helianthus* L. (Rieseberg, 2006), intergeneric hybridization in Senecioneae (Calvo et al., 2013), and within tribe Pertyeae (Freire, 2009). In addition to these well-studied cases, other lineages within the family exhibit unusual combinations of morphological traits that may also reflect a hybrid origin. For example, *Famatinanthus decussatus* (Hieron.) Ariza & S.E.Freire (Famatinantheae) exhibits unique morphological features that distinguish it from other Asteraceae members, including a cobblestone-like style surface (Erbar and Leins, 2025); decussate, opposite leaves; solitary capitula; broadly campanulate involucres; and corollas with extremely long lobes (Panero et al., 2014). These morphological changes could be indicative of hybridization, as such events can drive the emergence of novel traits and ecological adaptations (Arnold, 1992), potentially aiding the establishment of unique lineages. The unusual pollen morphology observed in some species of *Ainsliaea* DC. (Pertyeae) and the monotypic genus *Hecastocleis* A.Gray (Hecastocleideae), which deviates from the typical trizonocolporate structure (with three equatorial, compound apertures) of Compositae, supports the possibility of complex evolutionary processes. Wodehouse (1929) noted that *Hecastocleis* showed the greatest pollen similarity to *Ainsliaea*, suggesting a shared relationship. However, Tellería and Katinas (2005) pointed out that while the shared tricolpate pollen supports this connection, the psilate, regularly thickened exine aligns *Hecastocleis* more closely with lineages like Mutisieae and Gochnatieae. Additionally, while

Hecastocleis and *Ainsliaea* were historically grouped together with Mutisieae (Funk et al., 2009), molecular phylogenies based on plastid DNA (Panero and Funk, 2002, 2008) separated them entirely, placing *Hecastocleis* apart from *Ainsliaea* with intervening Carduoideae. Disentangling ancient patterns within complex evolutionary systems remains challenging; however, by implementing robust phylogenetic and network-based tools, Asteraceae offers a rich framework for testing reticulation hypotheses.

We previously published a robust backbone phylogeny for the Compositae that resolved many evolutionary relationships among its major lineages (Mandel et al., 2019). This study placed the origin of the sunflower family at ~83 mya in the late Cretaceous, and this phylogenetic backbone continues to serve as a framework for researchers studying the systematics and evolution of the major lineages that comprise the family. This robust phylogeny has led to a revised classification and new descriptions of several tribes and subfamilies within the family (Susanna et al., 2020). While the backbone inferred from the concatenated, maximum likelihood (ML) approach of Mandel et al. (2019) was more fully supported than earlier studies (Funk et al., 2009), numerous internal nodes received low support. Among these were the unresolved relationships among the closely related tribes Anthemideae, Astereae, Calenduleae, Gnaphalieae, and Senecioneae, which are members of a clade that has experienced several WGD events and rapid radiations (Watson et al., 2020; Criado-Ruiz et al., 2024). Watson et al. (2020) reported conflicting support in nuclear and plastid datasets and identified four potential ancient hybridization events within and across these tribes. Thus, clades of the large sunflower family phylogeny with high levels of conflict serve as good candidates to test major duplications or WGD events and reticulate evolution.

Mandel et al. (2019) also presented a phylogeny based on a multispecies coalescent approach that showed disagreement (albeit with very low local posterior probability) in key areas, notably with the placement of its sister family, Calyceraceae, which was nested within the Asteraceae. A more recent angiosperm phylogeny based on a different suite of orthologous genes also reported (with low support) a similar finding for the placement of Calyceraceae nested within the Asteraceae (Baker et al., 2022; Zuntini et al., 2024). Discrepancies in resulting phylogenetic trees inferred by different analytical approaches (i.e., concatenated vs. coalescent-based) highlight challenges to resolving evolutionary relationships. Integrating plastid genomic data may provide additional insights to clarify these conflicts. To this end, we used DNA sequence data of over 1000 nuclear loci and plastomes (obtained via Hyb-Seq methods; Weitemier et al., 2014; Mandel et al., 2014, 2015, 2017), as well as transcriptome data with targeted sampling of the understudied South American lineages of Asteraceae. Using these data, we investigated gene tree conflict, cytonuclear discordance, WGDs, and evidence for reticulate evolution across the sunflower family to investigate the influence of polyploidy and gene flow on the evolution of Asteraceae. Moreover, we integrated multiple streams of phylogenomic and transcriptomic data to investigate sources of phylogenetic discordance. A step-by-step guide for researchers interested in applying our methods is provided through GitHub (https://erika-r-moore.github.io/Ellestad_etal_2025_APPS_Hybridizations/; Figure 2).

METHODS

Taxon sampling and sequencing

For phylogenetic reconstructions, we sampled 267 total taxa from 47 tribes (out of the 50 tribes as recognized by Susanna et al., 2020) within Asteraceae, one species from the sister family Calyceraceae, and two from the Goodeniaceae (Appendix S1; see Supporting Information with this article). Genomic data for 265 of the 267 taxa were previously available from Mandel et al. (2019). This study added two new samples from the tribes Feddeeae and Polymnieae. For these two samples, DNA was extracted from silica gel–dried leaf tissue using the E.Z.N.A. SQ Plant DNA Kit (Omega Bio-Tek, Atlanta, Georgia, USA) following the manufacturer's instructions, with the addition of polyvinylpyrrolidone (PVP) and ascorbic acid to the SQ1 buffer (10 mL SQ1 buffer, 100 mg PVP, 90 mg ascorbic acid).

Up to 1 μg of DNA per sample was used to prepare genomic libraries using a NEBNext Ultra or NEBNext Ultra II Library Prep Kit (New England Biolabs, Ipswich, Massachusetts, USA) following the manufacturer's protocol. Libraries were used for target sequence capture using the myBaits Expert Compositae-1061 (Arbor Biosciences, Ann Arbor, Michigan, USA) following the manufacturer's protocol (v.5), with the following modifications: up to 500 ng of DNA library was used in a pooled capture with up to four samples per pool. For post-capture PCR amplification, the following conditions were set: 10 mL of DNA template from recovered capture, annealing temperature of 60°C, an elongation time of 45 s, and 15 cycles of amplification. Target-enriched libraries were quantified using a KAPA Library Quantification Kit (Kapa Biosystems, Boston, Massachusetts, USA) and pooled for sequencing. Enriched libraries were sequenced on the Illumina platform (paired end, 150 bp) by NovoGene (Sacramento, California, USA).

Eighteen taxa from 14 tribes within Asteraceae (Appendix S1) were collected for RNA-Seq transcriptome sequencing from fresh leaves in the field and submerged in 5 mL of RNA*later* Stabilization Solution (Invitrogen by Thermo Fisher Scientific, Carlsbad, California, USA) and stored on ice during shipment. RNA was isolated using a phosphate-buffered saline (PBS) wash (50 mM potassium phosphate, 150 mM NaCl; pH 7.2) prior to extraction with the E.Z.N.A. Plant RNA Kit (Omega Bio-Tek). RNA libraries were prepared by NovoGene from enriched mRNA

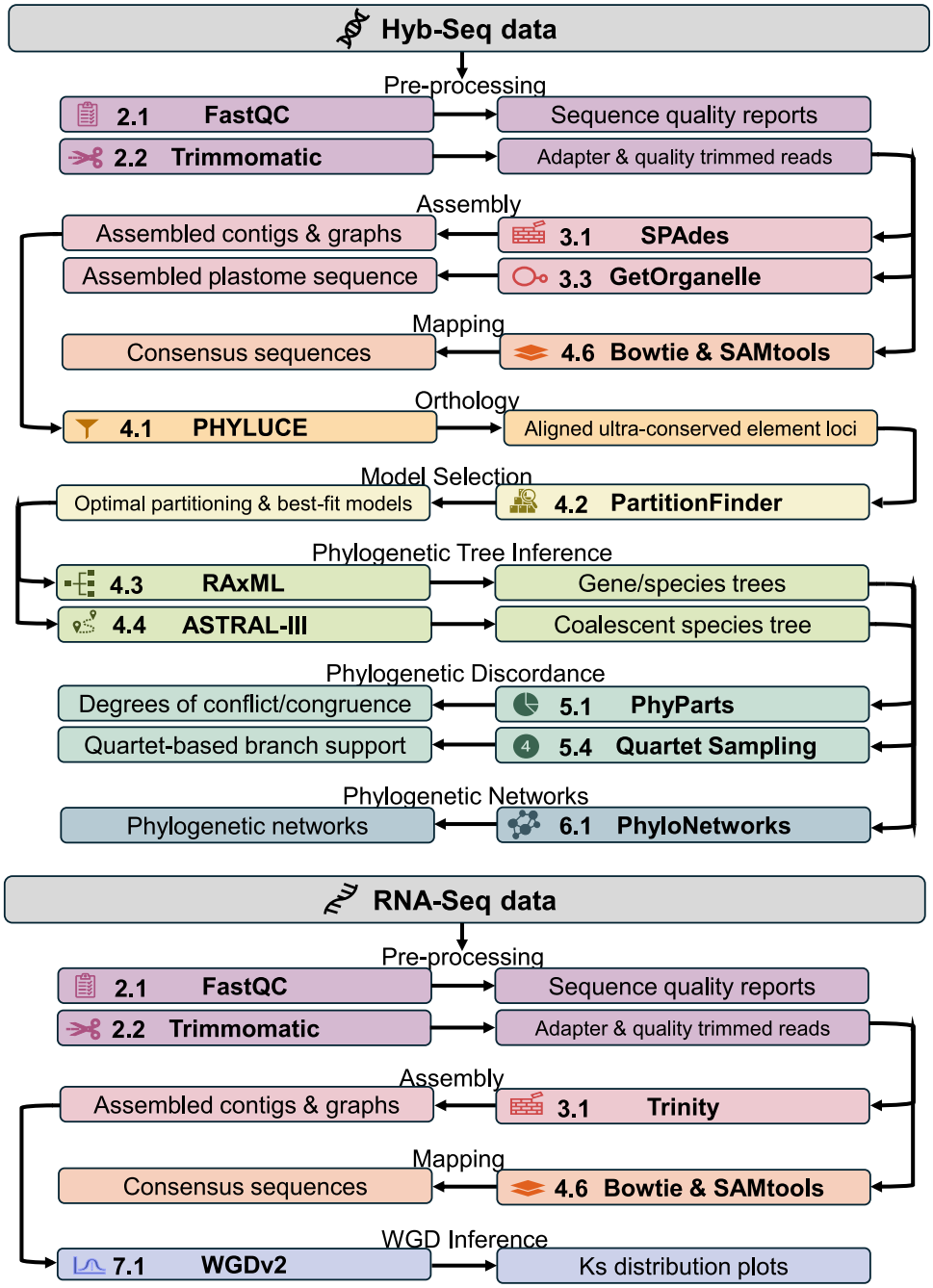

FIGURE 2 Simplified workflow of the bioinformatic pipeline of the main analyses involved in genomic investigations, from the raw input data to chosen software to general output. Numbers refer to sections of the bioinformatic tutorial available at https://erika-r-moore.github.io/Ellestad_etal_2025_APPS_Hybridizations/. The software programs used in each step are bolded.

using oligo(dT) beads with a 250–300 bp insert cDNA library using the TruSeq library preparation kit (Illumina, San Diego, California, USA). Libraries were sequenced on the Illumina platform (paired end, 150 bp) by Novo-Gene. Additionally, raw data from 51 transcriptomes were downloaded from the National Center for Biotechnology Information (NCBI), including 47 transcriptomes from the 1KP project (One Thousand Plant Transcriptomes Initiative, 2019) and Zhang et al. (2021) (see Appendix S1 for details).

Plastid sequence assembly and phylogenetic reconstruction

Chloroplast sequences were extracted from trimmed paired-end reads using two approaches. First, the GetOrganelle toolkit (Jin et al., 2020) was used to assemble circularized plastomes for all sequences with sufficient off-target Hyb-Seq data. The recommended parameters for embryophyte plant plastomes were used, but the number of rounds and the disentangle time limit were increased

(-R 75 --disentangle-time-limit 7200). Complete plastome assemblies were assessed for quality using Bandage (Wick et al., 2015), annotated using GeSeq (Tillich et al., 2017), then assessed for structural rearrangements using Mauve (Darling et al., 2004). A complete plastome representative was chosen for each Asteraceae subfamily and used as references for the second approach to assemble partial genomes using mapped reads. For subfamilies without a complete plastome representative, supplementary assemblies were downloaded from GenBank. All complete plastomes were restructured to ensure analogous quadripartite order in Geneious 11.1.5 (https://www.geneious.com). Then, for each subfamily, trimmed paired-end reads from both Hyb-Seq and transcriptome data were mapped to the respective reference plastome using Bowtie 2 (Langmead and Salzberg, 2012). Variants were called using SAMtools and BCFtools (Danecek et al., 2021), consensus files were created by applying variants to the reference genome, and files were then converted to FASTA format using seqtk (https://github.com/lh3/seqtk). Percentages of missing data were calculated for each mapped consensus sequence. Two alignments were created using MAFFT v.7.029b (Katoh and Standley, 2013): one of all mapped consensus sequences (including sequences with low mapped reads), and one including only tribal representatives exhibiting the highest amount of mapped data. The latter alignment was generated to minimize the potential effect of missing data on phylogenetic inference. For each alignment generated, model selection and maximum likelihood phylogenetic analysis were conducted using IQ-TREE (Nguyen et al., 2015; Minh et al., 2020), and trees were visualized using FigTree (http://tree.bio.ed.ac.uk/software/figtree/). Plastome and nuclear trees were compared using the function *cophylo* from the R package phytools (Revell, 2012) in R v.4.0.5 (R Core Team, 2016; Posit Team, 2025).

Nuclear phylogenetic analyses

We conducted concatenated and coalescent-based approaches for phylogenetic reconstruction with the genomic data. The concatenated tree combined all gene sequences into a single alignment for analysis as one dataset, while the coalescent-based tree was inferred with ASTRAL-III (Zhang et al., 2018) and used individual gene trees to estimate a species tree that accounted for gene tree discordance due to processes like ILS. For the genomic data, sequence reads from the Compositae-1061 enrichment were cleaned and trimmed using Trimmomatic v0.36 (Bolger et al., 2014) with a sliding window of 20 bp, retained only if they had a quality score of Q20 or greater, and assembled into contigs de novo using the SPAdes assembler v.3 (Prjibelski et al., 2020). Contigs of each taxon were analyzed using the PHYLUCE pipeline v.1.5.0 (Faircloth, 2016), which generates orthology predictions across loci and taxa in a conservative manner, specifically, when two or more contigs from a single taxon match a targeted locus, that locus is removed for that taxon. Nucleotide alignments on individual loci were carried out in MAFFT, and the resulting alignments were implemented in PartitionFinder v.2.1.1 (Lanfear et al., 2012) to determine the most appropriate nucleotide substitution model for each gene. The overall best-fitting model was used to build a concatenated maximum likelihood tree in RAxML v.8.1.3 (Stamatakis, 2014) with 1000 bootstrap (BS) replicates. For the coalescent-based analysis, individual gene trees were generated using RAxML with 100 BS replicates, utilizing the best-fitting model for each aligned, single-copy gene. The resulting gene matrices were used in ASTRAL-III v.5.7.3 to generate species trees with local posterior probability (LPP) values. The normalized quartet score was calculated using the -q option in ASTRAL-III. The concatenated and coalescent-based trees were then compared with the adjusted Robinson–Foulds (RF_{adj}) distance as outlined by Mitchell et al. (2017) using the *RF.dist()* function in the phangorn package (Schliep, 2011) in R with unrooted trees as input.

For transcriptome analyses, reads were trimmed with Trimmomatic using the parameters SLIDINGWINDOW:4:5, LEADING:5, TRAILING:5, and MINLEN:25, followed by assembly with Trinity v.2.2 (Grabherr et al., 2011; Haas et al., 2013). The resulting de novo transcriptomes were analyzed for completeness in BUSCO v.5.2.2 (Manni et al., 2021) using the embryophyta_odb10 database. We generated unigenes by selecting the longest isoform from each Trinity assembly using the script Trinity2Unigene.pl (Feng et al., 2020). We reduced redundant unigenes using CD-HIT-EST v4.7 (with the parameter -c 0.98) (Fu et al., 2012). Transcriptomes were translated to protein sequences by extracting the longest open reading frame, and candidate coding regions were then identified using TransDecoder v.3.0 (https://github.com/TransDecoder/TransDecoder).

Identifying phylogenetic discordance

We implemented the Quartet Sampling method of Pease et al. (2018) to characterize discordance in the resulting concatenated and coalescent-based nuclear phylogenies to assess the confidence, consistency, and informativeness of internal tree relationships, plus the reliability for each terminal taxon in the phylogeny. Briefly, for a given tree, the Quartet Sampling method separates internal tree branches into four non-overlapping subsets of taxa where the four sets (=quartets) present three relationships: a concordant relationship that matches the relationships in the input species tree and two alternative discordant relationships. We ran an unpartitioned analysis, using RaxML-ng v.0.9.0 for inference. The number of replicates per branch was set to 100 and the log-likelihood cutoff was set to 2, as suggested by the authors (Pease et al., 2018). Four values were calculated in this analysis: Quartet Concordance (QC), Quartet Differential (QD), Quartet Informativeness (QI), and Quartet Fidelity (QF). QC quantifies the relative support among the three possible topologies of four taxa, with a value of 1 representing full support; QD measures the disparity between the sampled proportions of the two discordant topologies, with 1

indicating no alternative history favored; QI measures branch informativeness, with a value of 1 indicating all quartets informative; and QF reports the total replicates that result in a concordant topology, with 1 indicating the particular taxon produced concordant topologies across all internal branches in which it was sampled. Results are shown as QC/QD/QI at each node and QF at each tip of the tree.

Additionally, PhyParts v.0.0.1 (Smith et al., 2015) was used with default settings to quantify and visualize gene-tree discordance in the concatenated phylogeny. Gene trees and the concatenated species tree were rooted to *Brunonia australis* R.Br. (Goodeniaceae) using the *pxrr* function in phyx v.1.3.1 (Brown et al., 2017). Plots containing pie charts at each node (indicating whether gene trees were supported, conflicted, missing, or uninformative at each node) were generated using a custom pipeline (https://bitbucket.org/dfmoralesb/target_enrichment_orthology/src/master/).

Inferring whole-genome duplications

The timing of ancient WGD events was inferred using intraspecific synonymous substitutions per site (K_S) age distributions with the *dmd*, *ksd*, and *viz* functions in wgd v.2.0.38 (Chen and Zwaenepoel, 2023; Chen et al., 2024), with the coding sequence (CDS) and general feature format (GFF) files of the transcriptome data as input. Using those commands, wgd fits an exponential-lognormal mixture model (ELMM) from ksrates (Sensalari et al., 2022) on the whole paranome K_S age distribution to more accurately estimate the location of potential WGDs. This analysis results in a node-average K_S distribution with up to five lognormal components (a–e) plotted along the whole paranome. Additionally, rate correction was performed using a five-taxon species tree including representatives from Campanulaceae (*Platycodon grandiflorus* (Jacq.) A.DC.), Goodeniaceae (*Scaevola aemula* R.Br.), Calyceraceae (*Acicarpha procumbens* Less.), and Asteraceae (tribes Barnadesieae [*Arnaldoa macbrideana* Ferreyra] and Mutiseae [*Gerbera delavayi* Franch.]). This tree was generated by pruning the rooted, nuclear concatenated species tree to include one representative per focal lineage using the *pxrmt* function in Phyx. The components of each K_S plot can be used to estimate the evolutionary age of a WGD event. Boxplots of the lognormal components for each tribe were visualized with the ggplot2 package (Wickham, 2016) in R.

Testing for evidence of reticulate evolution

To discover whether our data indicate evidence for reticulate evolution in key areas of the phylogeny, we implemented the function Species Networks applying Quartets (SNaQ) in PhyloNetworks (Solís-Lemus and Ané, 2016; Solís-Lemus et al., 2017), focusing on key backbone nodes at the base of the phylogeny where evidence for discordant topologies arose (see below). The SNaQ/PhyloNetworks approach incorporates uncertainty in estimated gene trees and gene tree discordance due to ILS. PhyloNetworks infers phylogenetic networks from multi-locus genetic data in a pseudolikelihood framework. The model considers both ILS, through the coalescent model, and horizontal inheritance of genes through reticulation.

We aimed to address reticulate evolution at the base of the family focusing on areas with nuclear vs. plastid data discordance as described in recent work in the family (Mandel et al., 2019; Palazzesi et al., 2022). We conducted six analyses: the first and second (South America [SA] and South America–alternative [SA-alt], respectively) focused on the sister family Calyceraceae and South American Asteraceae lineages but with varied taxa; the third and fourth (Hecastocleis and Hecastocleis-African, respectively) focused on the subfamily Hecastocleidoideae given nuclear–plastid discordance found with Carduoideae (Palazzesi et al., 2022); the fifth focused on relationships within the tribe Barnadesieae (Barnadesieae); and the sixth (Overall) sampled throughout the base of the tree. The software SNaQ is computationally intensive, and we therefore limited our sampling for these analyses. For each of the six tests, we selected representative taxa from the earliest-diverging lineages plus one Goodeniaceae taxon (*Scaevola tomentosa* Gaudich.) as an outgroup (Appendix S2). We then re-ran PHYLUCE only on those select taxa, utilizing the methods detailed above, to generate gene trees and a species tree for input to PhyloNetworks.

For all analyses, we estimated the best phylogenetic network with a varying number of hybridization (h) events allowed (h examined between 0 and 10). We evaluated the model fit for each of the 0–10 h events. We then examined when the pseudolikelihood score (ploglik) for h reached a plateau and selected the best network over 10 search replicates for each value of h. In cases where the best-scoring network included a hybridization involving the outgroup, we did not consider these networks in downstream interpretation. Spurious hybridizations with the outgroup have been reported in PhyloNetworks analyses, potentially arising from deep discordance that the method resolves as gene flow in the absence of alternative explanations. We therefore interpreted our network results in two ways: (1) we present only those hybridization events that do not involve the outgroup, and (2) we present a biologically reasonable alternative network that was recovered by the approach despite the suboptimal ploglik score. The justification for presenting the alternative network is that when we constrain the hybridization to not point at the outgroup, this network would be the best one (i.e., constrained optimization; Claudia Solís-Lemus, pers. comm.). The chosen network was then run 10 times with bootstrapping and 100 replicates for node support.

Bioinformatic tutorial

To ensure reproducibility and accessibility, we compiled a web tutorial using R Markdown (https://github.com/rstudio/rmarkdown) and bookdown (Xie, 2024) containing all code

and input data to guide researchers through our methodology (Figure 2). The tutorial implements multiple coding languages (R, Julia, Bash, Python) and is structured into logical sections that include installation instructions, preprocessing and file formatting, and results visualization for all bioinformatic analyses implemented in this study, with inline comments and markdown annotations explaining each step. The tutorial encompasses topics such as: assembling sequence data, generating species phylogenetic trees, identifying phylogenetic discordance, investigating ancient hybridization, and inferring WGDs. Where applicable, each analysis is presented independently, allowing users to easily modify or adapt specific subsections to their own datasets or research questions. This tutorial is hosted on GitHub (https://github.com/erika-r-moore/Ellestad_etal_2025_APPS_Hybridizations; see Data Availability Statement), allowing users to easily submit issues, request modifications, and provide feedback, ensuring that any problems or updates can be addressed in a timely manner.

RESULTS

Sequencing statistics

Sequencing Hyb-Seq phylogenomic data from tribes Feddeeae and Polymnieae resulted in 3,284,704 total reads for *Feddea cubensis* Urb. and 4,761,696 total reads for *Polymnia canadensis* L. Newly generated transcriptome data of 18 taxa resulted in a range of 14,445,770 to 23,701,861 for total reads (Appendix S1).

Plastid sequence assembly and phylogenetic reconstruction

Using off-target sequencing data from all 267 taxa, 22 complete circular plastomes spanning Calyceraceae and seven subfamilies of Asteraceae were assembled and annotated (Appendix S3). To use one plastome reference per subfamily for mapping analyses, seven additional plastomes were downloaded from GenBank: *Scaevola taccada* (Gaertn.) Roxb. (NC040933), *Famatinanthus decussatus* (NC081938), *Gerbera piloselloides* (L.) Cass. (NC0885588), *Hyaloseris cinerea* (Griseb.) Griseb. (NC081948), *Wunderlichia mirabilis* Riedel ex Baker (NC081943), *Pentaphorus foliolosus* D.Don (NC081946), and *Schlechtendalia luzulifolia* Less. (OM892819). In Asteraceae and Calyceraceae, plastome sizes ranged from 150,801 bp (*Senecio vulgaris* L.) to 153,670 bp (*Famatinanthus decussatus*). In Goodeniaceae, however, *Scaevola taccada* exhibited a much larger plastome (179,105 bp). Among the 29 complete plastomes, structural rearrangements identified through collinearity analyses revealed five general plastome configurations (Appendix S4). Consistent with previous findings (Pascual-Díaz et al., 2021), *Scaevola taccada* exhibited a substantial amount of inversions and rearrangements compared to all the other plastomes analyzed. Most Asteraceae taxa sampled exhibited very similar genome structures, possessing a large inversion in the large single-copy region as compared to the *Gamocarpha macrocephala* S.Denham & Pozner plastome of Calyceraceae, a well-known characteristic of Asteraceae plastomes (Jansen and Palmer, 1987). In addition to that large inversion, *Gerbera piloselloides* also exhibited a much smaller inversion. Interestingly, *Schlechtendalia luzulifolia*, from the early-diverging tribe Barnadesieae, exhibited two large-scale inversions differentiating its plastome structure from its familial relatives, and only one differentiating it from its sister family Calyceraceae.

Using a complete plastome from each subfamily as a reference, reads from all taxa were then mapped for phylogenetic analyses. Due to low proportions of mapped off-target reads for many of the samples (Appendix S3), a subsetted phylogenetic analysis was also conducted for tribal representatives exhibiting the highest amount of mapped data. Although *Famatinanthus decussatus* exhibited a very low quantity of mapped reads (643 bp), it was still included as it was the only representative from Famatinantheae. Phylogenetic analyses using the 46 representative taxa revealed higher bootstrap support for most clades compared to analyses using all, even low mapping percentage, data (Appendix S5). Family and tribal relationships differed in the chloroplast tree compared to the nuclear concatenated tree (Figure 3), as well as the coalescent-based tree (Appendix S6). Notably, in the chloroplast tree, *Gamocarpha macrocephala* (Calyceraceae) and *Fulcaldea stuessyi* Roque & V.A.Funk (Asteraceae, tribe Barnadesieae) were placed with high support in sister lineages (see below for methods on nuclear phylogenetic analyses).

Nuclear phylogenetic analyses

Assigning orthology with Phyluce retained 1047 out of the targeted 1061 loci across all 267 taxa. The number of loci recovered for each taxon ranged from 16 loci in *Liabum barahonense* Urb. to 584 in *Tanacetum parthenium* (L.) Sch.Bip (Appendix S3). The concatenated and coalescent-based tree had fairly different topologies (RF_{adj} = 0.303; Figure 4, Appendix S7). Almost all tribes within the concatenated tree were monophyletic (Appendix S8), with less consistent relationships in the coalescent-based tree (Appendix S9). One notable difference between the concatenated trees in the current study and the phylogeny of Mandel et al. (2019) was the placement of *Cyclolepis genistoides* Gillies ex D.Don, which was sister to tribe Wunderlichieae but is recovered as sister to Gochnatieae in the current study, albeit with a bootstrap of 76.

For the transcriptomes assembled from both the newly generated and the downloaded transcriptome data, BUSCO indicated there was a large range of completeness for each transcriptome assembly, from *Liabellum palmeri* (A.Gray) Rydb. only being 1.6% complete to *Platycodon grandiflorus* being 97.7% complete (Appendix S10). Mean and median

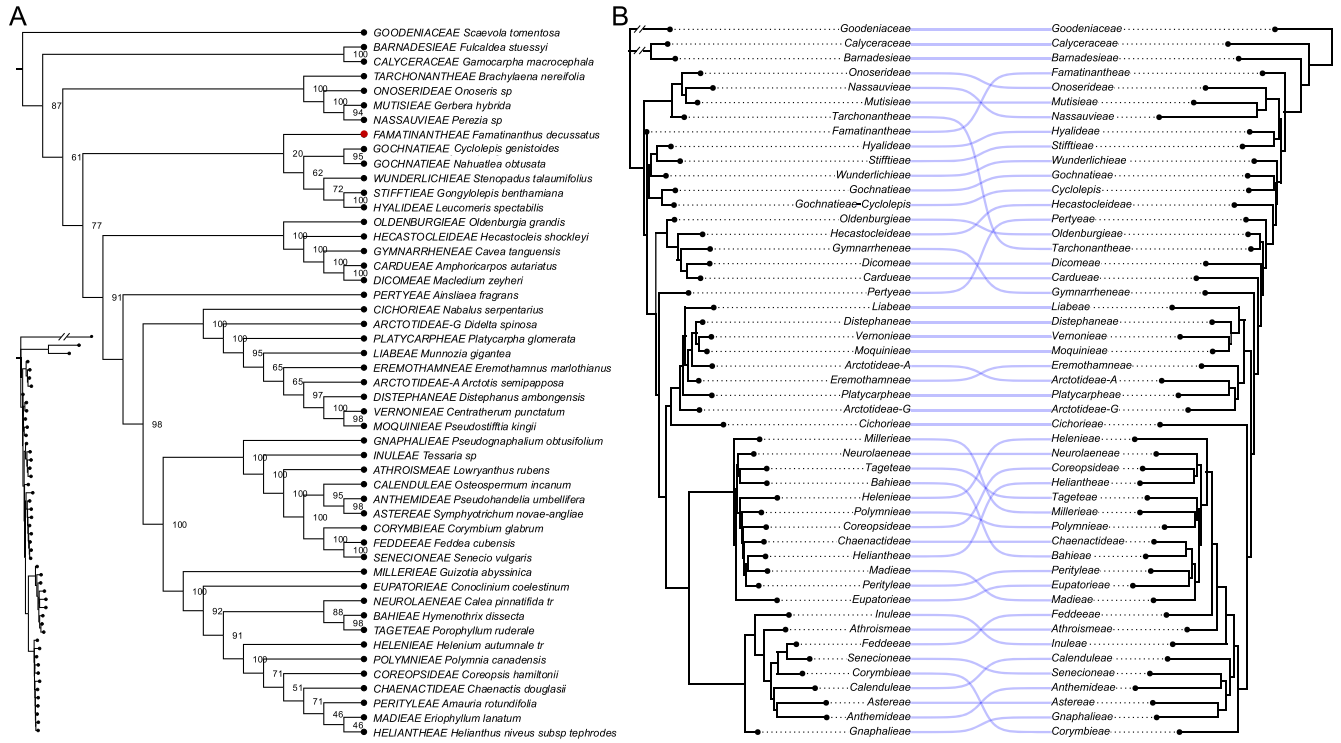

FIGURE 3 Comparison of the concatenated nuclear and plastid trees. (A) Tribe-level plastid tree generated from off-target mapped reads with bootstrap support values indicated at each node. The red circle highlights the only available tribal representative for Famatinantheae, *Famatinanthus decussatus*, which exhibited a very low quantity of mapped data (only 643 bp). (B) A tanglegram of the plastid tree (left) and concatenated nuclear tree (right) showing the differing evolutionary relationships by lines connecting tip labels. Note that a straight line does not indicate the same relationship at the node.

completeness was about 66.9% and 75.8%, respectively. The number of transcripts identified with TransDecoder also had a wide range. *Inula helenium* L. had the lowest number at 22,145 transcripts, and *Neurolaena lobata* (L.) Cass. had the most at 136,955 transcripts, with an average and median number of genes at 57,338 and 52,756 transcripts, respectively. Interestingly, *Liabellum palmeri* had the second highest number of transcripts at 120,115, although it was the most incomplete. Notably, *Famatinanthus decussatus* was not highly complete (10.7%) because it was assembled using genomic rather than transcriptomic data (Zhang et al., 2021).

Identifying phylogenetic discordance

The Quartet Sampling approach indicated several areas of discordance across both the concatenated and coalescent-based phylogenetic reconstructions of Asteraceae (Figure 4; Appendices S11, S12, S13). For the concatenated nuclear phylogeny, several backbone nodes show evidence of discordance, including when the larger Asteraceae clade diverged from its sister subfamily Barnadesiodeae (discordant node 1, Appendix S11) and the rest of the family, and the split between the clade with subfamilies Stifftioideae + Mutisioideae (discordant node 2, Appendix S11) and the remainder of the family. An additional series of strongly discordant values are present in three almost sequential nodes (discordant nodes 3–5, Appendix S11) along the backbone involving tribes Gochnatieae, Hecastocleideae, and Oldenburgieae + Tarchonantheae (Appendix S11). Interestingly, for the coalescent-based phylogeny, the backbone node indicating the divergence of Calyceraceae + tribe Barnadesieae from the rest of Asteraceae showed evidence of strong support; however, the node indicating the divergence of Calyceraceae and tribe Barnadesieae exhibited high discordance (Figure 4, Appendix 12). Similar results are found with the PhyParts analysis, where the backbone is majorly discordant, especially after the split from Barnadesieae to the remainder of the family (Appendix S14).

Inferring whole-genome duplications

K_S analyses were used to infer ancient WGD events with 69 transcriptomes of Campanulaceae, Goodeniaceae, Calyceraceae, and Asteraceae taxa, with multiple K_S peaks identified for each taxon. In summary, for Asteraceae, many taxa exhibited peaks (indicated by the modes of their lognormal a, b, c, and d optimized distributions) around 0.5, 2.25, and 3.75 (Appendices S10, S15, S16). Taxa from the closely related families Calyceraceae and Goodeniaceae also exhibited peaks around these K_S values, although taxa from

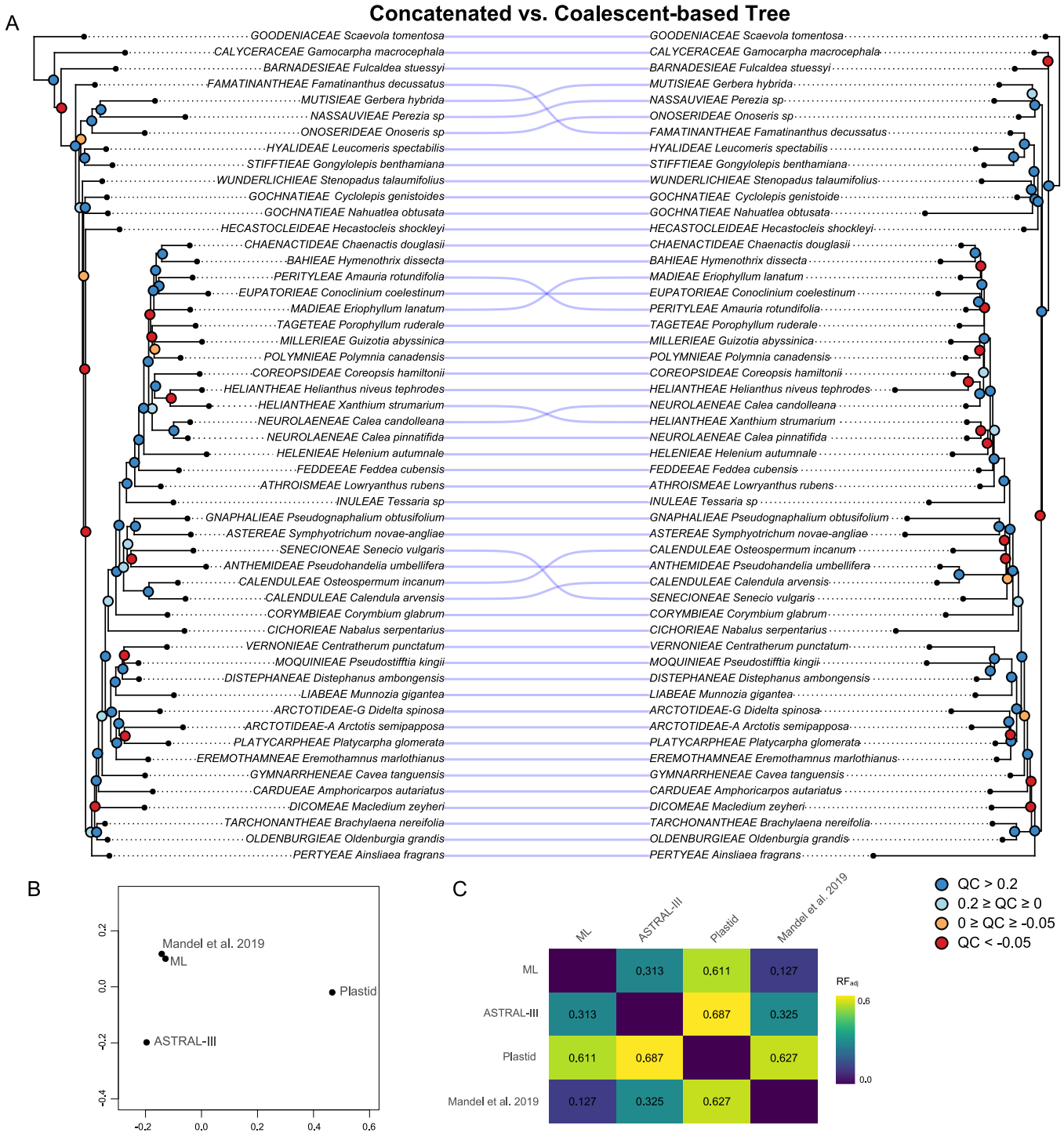

FIGURE 4 Comparing phylogenetic tree reconstructions. (A) Tanglegram of the nuclear concatenated tree (left) and the coalescent-based tree (right) showing the differing evolutionary relationships of tribes by lines connecting tip labels. Note that a straight line does not indicate the same relationship at the node. Colored circles at nodes correspond to Quartet Concordance (QC) values calculated with Quartet Sampling, and follow the color scheme at the bottom right of the tanglegram. Full Quartet Sampling results can be found in Appendices S12 and S13. (B) Classical multidimensional scaling (MDS) of the adjusted Robinson–Foulds (RF_{adj}) values calculated between the nuclear concatenated (ML), nuclear coalescent-based (ASTRAL-III), and plastid trees from this study, along with a comparison of the tree generated by Mandel et al. (2019). Trees with more similar relationships appear closer together. Raw values are found in C. (C) RF_{adj} values as a pairwise matrix that range from zero to one, with values closer to zero indicating more similar relationships and values closer to one indicating more different relationships. Boxes are color coordinated according to the legend to the right of the matrix.

Campanulaceae did not (Appendix S10, S16). Mixed paralog/ortholog K_S plots between *Acicarpha procumbens* (Calyceraceae) and taxa from Asteraceae (*Arnaldoa macbrideana* and *Gerbera delavayi*) and *Scaevola aemula* (Goodeniaceae), however, revealed similar K_S peaks (0.52, 0.53, and 0.55) between Calyceraceae and Asteraceae, but exhibited a higher K_S peak (0.64) for Goodeniaceae orthologs, indicating that Goodeniaceae likely diverged prior to a Calyceraceae/Asteraceae shared WGD event (Appendix S16C). Evidence for more recent WGD was indicated by substantial K_S peaks within the lognormal a and b distributions (Appendix S10). More recent WGD was particularly evident among taxa within the following tribes: Barnadesieae, Cardueae, Cichorieae, Gnaphalieae, Senecioneae, Anthemideae, Helenieae, Neurolaeneae, Chaenactideae, and Eupatorieae (Appendix S16). However, evidence of recent WGD events within the tribes Cardueae and Neurolaeneae remains unclear due to the lack of completeness of the transcriptomic data associated with *Carthamus lanatus* L., *Neurolaena lobata*, and *Calea pinnatifida* Banks ex Steud., which may bias inferences (Appendix S10). Multiple tribes contain taxa with dissimilar K_S distribution patterns, e.g., *Cota tinctoria* (L.) J.Gay in Anthemideae, *Chuquiraga avellanedae* Lorentz in Barnadesieae, *Carthamus lanatus* in Cardueae, *Eupatorium chinense* L. in Eupatorieae, *Leontopodium nivale* subsp. *alpinum* (Cass.) Greuter in Gnaphalieae, *Helenium autumnale* L. in Helenieae, *Ameghinoa patagonica* Speg. in Nassauvieae, and multiple taxa in Senecioneae. These K_S signals may be due to lineage-specific whole-genome or small-scale duplication events or may also be an artifact of the aforementioned low-quality transcriptomic data of some taxa.

Testing for evidence for reticulate evolution

Re-running gene tree inferences for the six network subtree groupings resulted in a varying but comparable number of gene trees: SA: 428 loci, SA-alt: 554 loci, Hecastocleis: 354 loci, Hecastocleis-African: 365 loci, Barnadesieae: 421 loci, and Overall: 531 loci. The SNaQ analyses across tribes at the base of Asteraceae show many potential reticulation events (Figure 5, Appendix S17). The ploglik scores indicate the best network models range from one to seven reticulation events dependent on the subtree, with SA having the least and Overall having the most (Appendix S18). Many events were unique to each tested subtree; however, there were also some overlapping networks at specific branches or nodes. For example, three of the tested subtrees showed tribe Mutiseae involved in a reticulation event, although the parent taxa varied depending on species sampling. This result was also seen with two subtrees of the tribe Pertyeae (Figure 5; Appendices S17, S19). Including different sets of taxa in the SA analysis (SA and SA-alt) showed one overlapping network involving Mutisieae, Onoserideae, and Nassauvieae, with SA having an additional event involving Barnadesieae taxa. Three or more reticulation events were inferred at four branch points (Figure 5; Appendices S17, S19), providing additional evidence for reticulations at those ancestral branches.

DISCUSSION

Building on the Asteraceae backbone phylogeny of Mandel et al. (2019), in addition to some recently acquired genomic and transcriptomic data, we explored how WGD events and gene flow may have shaped the evolutionary history of Asteraceae. Despite strong support for most nodes in the concatenated phylogeny of Mandel et al. (2019), Quartet Sampling revealed key areas of phylogenetic discordance within Asteraceae, particularly at major backbone nodes where conflicting topologies suggest complex evolutionary histories (Figure 4). Notably, discordance at the divergence of Barnadesioideae and within the clades containing Stifftieae, Mutisieae, Hecastocleideae, Pertyeae, and Oldenburgieae + Tarchonantheae points to potential influence of ILS or historical gene flow. Additionally, the plastid-based tribal reconstruction shows further incongruence, reinforcing the likelihood of conflicting evolutionary signals (Figure 3, Appendix S6). The prevalence of discordance, especially at the base of the Asteraceae tree, suggests that reticulate or network-like evolution may at least partially explain these patterns. However, it is important to recognize that not all discordance reflects underlying biological processes. Some gene tree conflict may stem from gene tree estimation error due to low phylogenetic information, missing data, or uneven taxon sampling across loci (Appendix S14), obscuring true evolutionary signal.

Although the investigation of potential WGD events in Asteraceae is not new, with studies dating from at least 2008 (Barker et al., 2008), the number of species analyzed is still incipient and is focused mostly on species from the Northern Hemisphere (e.g., Huang et al., 2016; Barker et al., 2016b). Most of the transcriptome data available for species of Asteraceae comes from two sources: the 1KP project (One Thousand Plant Transcriptomes Initiative, 2019) and the studies led by Hong Ma's group (Huang et al., 2016; Zhang et al., 2021). Here, we relied on this previously generated data, choosing key species and aiming to cover as much of the tribal diversity as possible, and complemented it with newly obtained sequences. Our goal was to maximize sampling of the South American, earliest-diverging clades of the family. We added 11 new transcriptomes from the focal tribes Barnadesieae, Hyalideae, Stifftieae, Mutisieae, Nassauvieae, and Gochnatieae, plus 13 new sequences from other tribes. Although most studies that investigate WGD rely on calculating and reporting K_S values and the existence of peaks that could indicate genome multiplication events, how this information is reported varies. Barker et al. (2008, 2016b) report K_S peak ranges and plots, and Huang et al. (2016) report summarized K_S ranges for each species in their phylogeny. Zhang et al. (2021), building on the dataset of Huang et al. (2016), do not report K_S values or plots. This lack of standardization on how results are reported affects the reuse

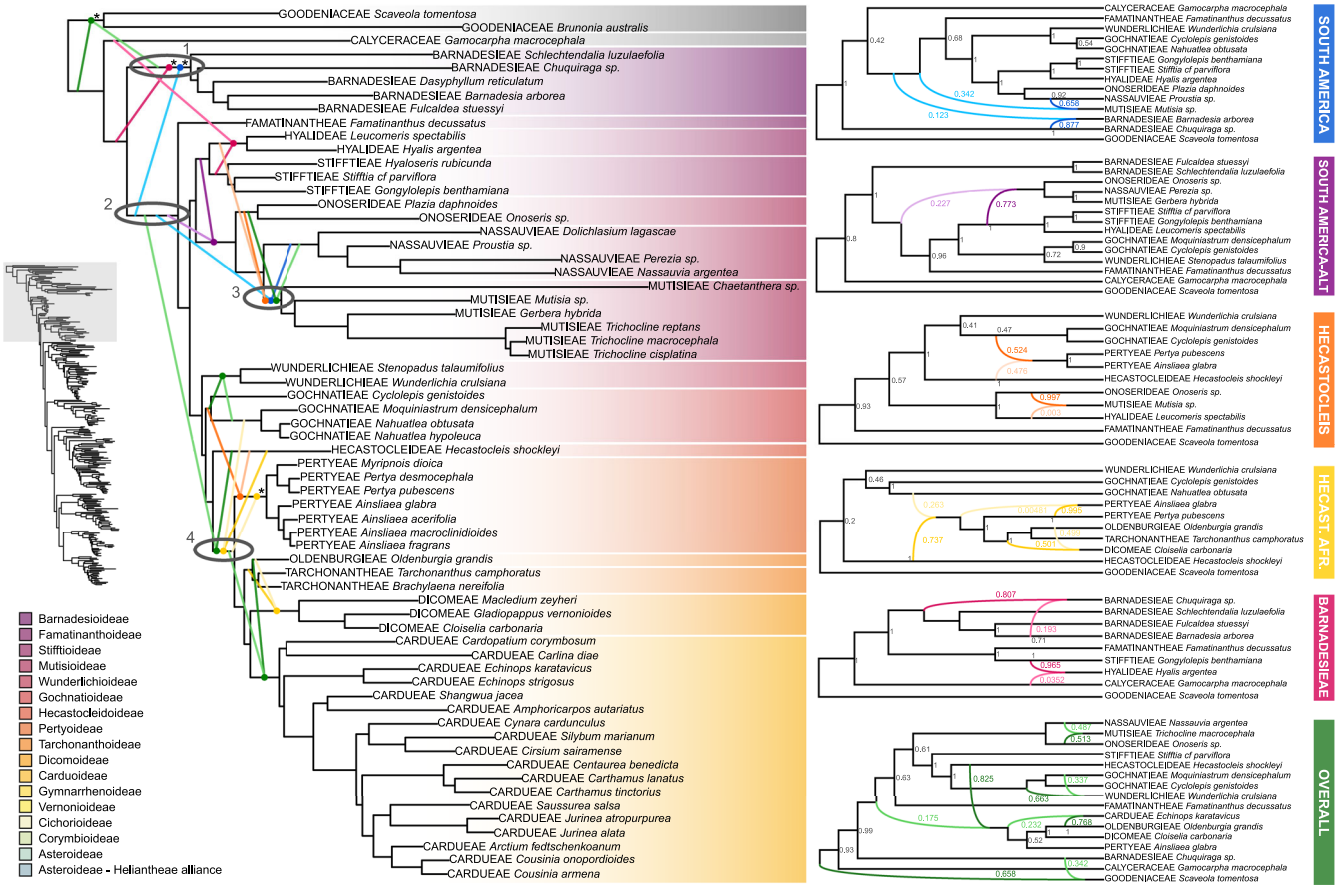

FIGURE 5 Phylogenetic relationships and inferred hybridization events at the base of the full nuclear phylogeny. Tribes are colored by subfamily as indicated in the legend on the bottom left. The left main tree shows the species relationships as bifurcating black edges, representing well-supported phylogenetic splits. Colored edges represent hybridization events inferred from six individual subtree analyses using PhyloNetworks (SNaQ), with colors corresponding to each subtree shown in the panels on the right. Hybrid edges connect two parental lineages to a single hybrid node, representing gene flow into the recipient lineage. These edges are shown as lines pointing toward the recipient taxon's crown as a dot, as inferences are only being made at the tribal level. This structure reflects the reticulate nature of the inferred network and should not be interpreted as independent bifurcations. Asterisks mark reticulations where one parent is inferred within the same tribe (intra-tribal hybridization). Numbers at nodes in the networks indicate local posterior probability (LPP) support values, and colored numbers on hybrid edges denote estimated proportions of ancestry from each parental lineage. Branches with four or more reticulations are circled and numbered 1–4. Reticulations involving the outgroup were excluded due to biological and methodological considerations (see Methods). Complete summaries of all inferred reticulations are provided in Appendix S18.

of data, especially as methods advance. For this study, we reused sequences both from Huang et al. (2016) and Zhang et al. (2021); however, due to the summarized reporting of the former and lack of reporting of the latter, we faced challenges in evaluating the congruence of our results to theirs. While our results generally resemble the K_S value trends reported by Huang et al. (2016), the clustering bins calculated for their K_S values (shown as shaded blocks in Appendix S16B) do not directly translate to our findings.

We found similar results to previous studies for the families related to Asteraceae: Campanulaceae shows signs of an ancient duplication (shared with all other taxa assessed), while Goodeniaceae, Calyceraceae, and almost all Asteraceae taxa additionally show signs of more recent WGD events (Appendix S16) (Huang et al., 2016; Barker et al., 2016b). We found further evidence of a recent WGD event in tribe Barnadesieae, the sister group to the remainder of Asteraceae, although in a more recent range than previously reported by Barker et al. (2016b). Previous studies have interpreted this evidence as indicating that Calyceraceae and Barnadesieae share a paleotetraploid ancestor (Huang et al., 2016; Barker et al., 2016b). However, recent, more widely sampled phylogenomic studies have pointed out some variability in the relative positions of Calyceraceae, Barnadesieae, and the rest of Asteraceae (Zuntini et al., 2024), with Calyceraceae nested within Asteraceae. While some claims about the monophyly of Asteraceae are probably overstated in that study, due to low branch support and backboning methods, this variability in the relationships casts doubt on whether Calyceraceae and Barnadesieae actually share the same WGD event. It is important to point out that although the divergence of these clades probably dates to the Late Cretaceous (Mandel et al., 2019), the crown ages of these groups are more recent—in the Miocene for Calyceraceae (Brignone et al., 2023) and later in the Cretaceous for Barnadesieae

(although with most of the extant diversity originating in the Oligocene/Miocene [see Lörch et al., 2021]). With the K_S data indicating more recent potential WGDs in these clades, it could indicate that Calyceraceae and Barnadesieae each had independent, rather than shared, WGDs. Very few species of these groups have been included in transcriptomic studies, likely due to their restricted South American distribution. A dedicated focus on these two groups has the potential to bring resolution to this issue.

We similarly found evidence of WGDs in other South American tribes, namely Famatinantheae, Hyalideae, Onoserideae, Mutisieae, and Gochnatieae. The occurrence of a WGD in Mutisieae was previously posited in several studies (Barker et al., 2008; Huang et al., 2016; Zhang et al., 2021), and our wider sampling here supports this. The position of this duplication is still debatable, however, as Huang et al. (2016) indicated that it occurred somewhere on the stem of the clade, while Zhang et al. (2021) suggested that its occurrence was deeper within the clade. Zhang et al. (2021) sampled representatives of most of these South American tribes but did not detect WGDs specific to these clades, which suggests that the peaks we observe may be a relic from the WGD that is shared by most Asteraceae. Nevertheless, the Mutisieae and its allied tribes need additional dedicated studies to resolve the placement of this duplication. Furthermore, we recovered conflicting evidence of WGDs in tribes usually considered to have one, such as Senecioneae and Gnaphalieae (Huang et al., 2016; Zhang et al., 2021, respectively); within these tribes, some species that we examined exhibited a more recent K_S peak, while others did not (Appendix S16). Incongruences in WGD tribal inferences among studies could be due to the different taxa sampled in each study. With WGDs nested within large clades (e.g., Senecioneae has ca. 3500 species and Gnaphalieae ca. 2100; Mandel et al., 2019), it is probable that a number of species are not members of the clade that share the WGD.

The SNaQ analyses across the tested Asteraceae tribes reveal widespread signals of reticulation (Figure 5), strengthening the hypothesis that hybridization played a role in the family's early diversification. The best-fitting network models suggest varying numbers of reticulation events (ranging from one to seven), with the SA grade showing the fewest events and the Overall dataset the most (Appendix S18), indicating that the extent of detected reticulation is influenced by taxon sampling. While many reticulation events were unique to individual subtrees, some patterns recurred across multiple analyses (Appendix S19), suggesting consistent historical signals. Notably, four branches of the tree provide multiple forms of evidence for reticulations (Figure 5). Two of these branches occur before and two occur after the divergence of the monotypic tribes Famatinantheae and Hecastocleideae. The detection of a reticulation event preceding the divergence of *Famatinanthus decussatus* (tribe Famatinantheae) suggests that hybridizations may have contributed to its origin and unique morphological traits (Panero et al., 2014; Erbar and Leins, 2025). In contrast, the evidence for a reticulation event following the divergence of *Hecastocleis shockleyi* A.Gray (tribe Hecastocleideae) suggests an alternative evolutionary scenario. Post-divergence hybridization may have introduced advantageous genetic variation, enhancing its resilience to extreme conditions. Beyond these monotypic lineages, we also detected multiple lines of evidence for hybridizations among the branches leading to tribes Mutisieae and Barnadesieae (Figure 5, Appendix S19). Mutisieae was consistently involved in reticulation events across three subtrees, either as the contributor to or the product of reticulation, with multiple analyses showing Nassauvieae and Onoserideae as the parent taxa. Barnadesieae was involved in multiple reticulation events across the tested subtrees, although the inferred parental taxa varied with species sampling, highlighting the complex nature of these evolutionary interactions and/or uncertainty associated with different sampling schemes. Similarly, multiple subtrees suggested reticulation events within Pertyeae, further supporting the idea that network-like evolution is a common feature in Asteraceae, especially when there is incongruence between nuclear and plastid phylogenies. There was an overlap between the SA and SA-alt analyses, with both identifying a network involving Mutisieae, Onoserideae, and Nassauvieae. Additionally, the identification of three or more reticulation events in four specific branches suggests that these ancestral nodes represent key points of evolutionary complexity (Figure 5) and may explain why there is wide morphological overlap between these tribes in Mutisioideae, a subfamily that has been historically difficult to circumscribe (Katinas et al., 2008; Funk et al., 2009). Evidence for reticulation based on subsetted phylogenetic networks, such as these, provides an informed platform for testing hybridization hypotheses based solely on morphology and will help to provide a more comprehensive and nuanced understanding of evolutionary relationships and trait evolution within Asteraceae.

CONCLUSIONS

Our findings provide compelling evidence for and emphasize the significant role that WGDs, ancient hybridization, and introgression have played in shaping the early diversification of Asteraceae, likely acting as catalysts for morphological innovation, ecological adaptation, and lineage diversification. We also note that extinction has likely played a role in discordance, particularly at the base of the tree, which unfortunately is something that cannot be sampled. The continued expansion of genomic resources for the family and broader taxonomic sampling will enable deeper investigations into gene flow, phylogenetic relationships, and genome evolution, helping to unravel the complexities of its evolutionary history. High-resolution genomic data, coupled with novel phylogenetic network tools, will be essential to refine our understanding of the prevalence and consequences of ancient reticulation events, introgression, and lineage diversification within the family. We hope that by including a tutorial for this study via R Markdown, future research will more easily incorporate

reticulation-based analyses to further expand our knowledge on hybridization events within the family. While the complexity of evolution has long been recognized, the increasing availability of genomic datasets and computational approaches now allows us to dissect and more accurately document the intricate evolutionary histories that contribute to the "tree," or perhaps more accurately, the network of life.

AUTHOR CONTRIBUTIONS

J.R.M. and L.E.W. conceived of the study that evolved over time with C.M.S. J.M.B. and C.M.S. collected and identified specimens for this study. R.T. and C.M.S. generated transcriptome and genome data. E.R.M.P., P.A.E., J.R.M., and C.M.S. all performed bioinformatic analyses. E.R.M.P., P.A.E., J.R.M., L.E.W., and C.M.S. wrote the manuscript and interpreted the results. All authors provided edits and approved the final version of the manuscript.

ACKNOWLEDGMENTS

The authors thank Eric Spangler and Kristian Skjervold, administrators of the University of Memphis High-Performance Computing (HPC) cluster, and Mike Barker and Zheng Li for valuable discussions on an earlier version of this manuscript. We also thank James Beck, Alfonso Susanna, Zheng Li, Paola De Lima Ferreira, and Javier Francisco-Ortega for providing plant samples used in this study.

DATA AVAILABILITY STATEMENT

Raw sequence data are available through the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) and GenBank portals. Previously published transcriptome data are available in BioProjects PRJNA994483, PRJNA636629, and PRJEB4922. Newly sequenced transcriptomes are deposited at NCBI SRA under BioProject PRJNA1246973. The two newly sequenced Hyb-Seq datasets are uploaded to NCBI SRA under BioProject PRJNA1246973, with the remaining taxa found in BioProject PRJNA540287. Complete plastomes are uploaded to NCBI under BioProject PRJNA1246973. Associated tree files, alignments, and visualizations of plastome annotation and structure are available on Zenodo (https://doi.org/10.5281/zenodo.16111839; Ellestad et al., 2025).

ORCID

Paige A. Ellestad https://orcid.org/0000-0003-3930-3239
Erika R. Moore-Pollard https://orcid.org/0000-0002-1182-9274
Carolina M. Siniscalchi https://orcid.org/0000-0003-3349-5081
Ramhari Thapa https://orcid.org/0000-0002-4730-1076
Linda E. Watson https://orcid.org/0000-0001-6098-779X
J. Mauricio Bonifacino https://orcid.org/0000-0003-0377-3853
Jennifer R. Mandel https://orcid.org/0000-0003-3539-2991

REFERENCES

Arnold, M. L. 1992. Natural hybridization as an evolutionary process. *Annual Review of Ecology and Systematics* 23(1): 237–261.

Arnold, M. L. 2007. Evolution through genetic exchange, 272. Oxford University Press, London, United Kingdom.

Baker, W. J., P. Bailey, V. Barber, A. Barker, S. Bellot, D. Bishop, L. R. Botigué, et al. 2022. A comprehensive phylogenomic platform for exploring the angiosperm tree of life. *Systematic Biology* 71(2): 301–319.

Barker, M. S., N. C. Kane, M. Matvienko, A. Kozik, R. W. Michelmore, S. J. Knapp, and L. H. Rieseberg. 2008. Multiple paleopolyploidizations during the evolution of the Compositae reveal parallel patterns of duplicate gene retention after millions of years. *Molecular Biology and Evolution* 25(11): 2445–2455.

Barker, M. S., N. Arrigo, A. E. Baniaga, Z. Li, and D. A. Levin. 2016a. On the relative abundance of autopolyploids and allopolyploids. *The New Phytologist* 210(2): 391–398.

Barker, M. S., Z. Li, T. I. Kidder, C. R. Reardon, Z. Lai, L. O. Oliveira, M. Scascitelli, and L. H. Rieseberg. 2016b. Most Compositae (Asteraceae) are descendants of a paleohexaploid and all share a paleotetraploid ancestor with the Calyceraceae. *American Journal of Botany* 103(7): 1203–1211.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15): 2114–2120.

Brignone, N. F., N. Mazet, R. Pozner, and S. S. Denham. 2023. Calyceraceae: Unexpected diversification pattern in the Southern Andes. *Perspectives in Plant Ecology, Evolution and Systematics* 60: e125744.

Brown, J. W., J. F. Walker, and S. A. Smith. 2017. Phyx: Phylogenetic tools for UNIX. *Bioinformatics* 33(12): 1886–1888.

Calvo, J., I. Álvarez, C. Aedo, and P. B. Pelser. 2013. A phylogenetic analysis and new delimitation of *Senecio* sect. *Crociseris* (Compositae: Senecioneae), with evidence of intergeneric hybridization. *Taxon* 62(1): 127–140.

Chen, H., and A. Zwaenepoel. 2023. Inference of ancient polyploidy from genomic data. *Methods in Molecular Biology* 2545: 3–18.

Chen, H., A. Zwaenepoel, and Y. Van de Peer. 2024. wgd v2: A suite of tools to uncover and date ancient polyploidy and whole-genome duplication. *Bioinformatics* 40(5): btae272.

Criado-Ruiz, D., J. Vallès, R. J. Bayer, L. Palazzesi, J. Pellicer, I. P. Lorenzo, O. Maurin, et al. 2024. A phylogenomic approach to disentangling the evolution of the large and diverse daisy tribe Anthemideae (Asteraceae). *Journal of Systematics and Evolution* 64: 282–306.

Danecek, P., J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, et al. 2021. Twelve years of SAMtools and BCFtools. *GigaScience* 10(2): giab008.

Darling, A. C. E., B. Mau, F. R. Blattner, and N. T. Perna. 2004. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Research* 14(7): 1394–1403.

Edger, P. P., H. M. Heidel-Fischer, M. Bekaert, J. Rota, G. Glöckner, A. E. Platts, D. G. Heckel, et al. 2015. The butterfly plant arms-race escalated by gene and genome duplications. *Proceedings of the National Academy of Sciences, USA* 112(27): 8362–8366.

Edwards, S. V., S. Potter, C. J. Schmitt, J. G. Bragg, and C. Moritz. 2016. Reticulation, divergence, and the phylogeography-phylogenetics continuum. *Proceedings of the National Academy of Sciences, USA* 113(29): 8025–8032.

Ellestad, P., E. Moore-Pollard, C. Siniscalchi, R. Thapa, L. Watson, M. Bonifacino, and J. Mandel. 2025. Ancient hybridization and phylogenetic discordance: Exploring evolutionary complexity in Asteraceae. *Zenodo* https://doi.org/10.5281/zenodo.16111839 [posted 18 July 2025; accessed 10 April 2026].

Ellstrand, N. C. 2014. Is gene flow the most important evolutionary force in plants? *American Journal of Botany* 101(5): 737–753.

Erbar, C., and P. Leins. 2025. Styles of Asteraceae: A synopsis of their morphological and functional diversity. *Capitulum* 3(2): 1–38. https://doi.org/10.53875/capitulum.03.2.01.

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

Feng, C., J. Wang, L. Wu, H. Kong, L. Yang, C. Feng, K. Wang, et al. 2020. The genome of a cave plant, *Primulina huaijiensis*, provides insights into adaptation to limestone karst habitats. *The New Phytologist* 227(4): 1249–1263.

Freire, S. E. 2009. Pertyeae (Pertyoideae), 315–326. In V. A. Funk, A. Susanna, T. F. Stuessy, and R. J. Bayer [eds.], Systematics, evolution, and biogeography of Compositae. International Association for Plant Taxonomy, Vienna, Austria.

Fu, L., B. Niu, Z. Zhu, S. Wu, and W. Li. 2012. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics (Oxford, England)* 28(23): 3150–3152.

Funk, V. A., A. Susanna, T. F. Stuessy, and R. J. Bayer [eds.]. 2009. Systematics, evolution, and biogeography of Compositae. International Association for Plant Taxonomy (IAPT), Vienna, Austria.

Givnish, T. J. 2010. Ecology of plant speciation. *Taxon* 59(5): 1326–1366.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29(7): 644–652.

Haas, B. J., A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B. Couger, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols* 8(8): 1494–1512.

Hernández, F., R. B. Vercellino, M. Todesco, N. Bercovich, D. Alvarez, J. Brunet, A. Presotto, and L. H. Rieseberg. 2024. Admixture with cultivated sunflower likely facilitated establishment and spread of wild sunflower (*Helianthus annuus*) in Argentina. *Molecular Ecology* 33(22): e17560.

Hosner, P. A., B. C. Faircloth, T. C. Glenn, E. L. Braun, and R. T. Kimball. 2016. Avoiding missing data biases in phylogenomic inference: An empirical study in the landfowl (Aves: Galliformes). *Molecular Biology and Evolution* 33(4): 1110–1125.

Huang, C. H., C. Zhang, M. Liu, Y. Hu, T. Gao, J. Qi, and H. Ma. 2016. Multiple polyploidization events across Asteraceae with two nested events in the early history revealed by nuclear phylogenomics. *Molecular Biology and Evolution* 33(11): 2820–2835.

Jansen, R. K., and J. D. Palmer. 1987. A chloroplast DNA inversion marks an ancient evolutionary split in the sunflower family (Asteraceae). *Proceedings of the National Academy of Sciences, USA* 84(16): 5818–5822.

Jin, J.-J., W.-B. Yu, J.-B. Yang, Y. Song, C. W. dePamphilis, T.-S. Yi, and D.-Z. Li. 2020. GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biology* 21(1): e241.

Katinas, L., J. Pruski, G. Sancho, and M. C. Tellería. 2008. The subfamily Mutisioideae (Asteraceae). *The Botanical Review* 74(4): 469–716.

Katoh, K., and D. M. Standley. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* 30(4): 772–780.

Lambert, S. M., T. W. Reeder, and J. J. Wiens. 2015. When do species-tree and concatenated estimates disagree? An empirical analysis with higher-level scincid lizard phylogeny. *Molecular Phylogenetics and Evolution* 82(Pt A): 146–155.

Landis, J. B., D. E. Soltis, Z. Li, H. E. Marx, M. S. Barker, D. C. Tank, and P. S. Soltis. 2018. Impact of whole-genome duplication events on diversification rates in angiosperms. *American Journal of Botany* 105(3): 348–363.

Lanfear, R., and M. W. Hahn. 2024. The meaning and measure of concordance factors in phylogenomics. *Molecular Biology and Evolution* 41(11): msae214.

Lanfear, R., B. Calcott, S. Y. W. Ho, and S. Guindon. 2012. PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution* 29: 1695–1701.

Langmead, B., and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9(4): 357–359.

Li, Z., M. T. W. McKibben, G. S. Finch, P. D. Blischak, B. L. Sutherland, and M. S. Barker. 2021. Patterns and processes of diploidization in land plants. *Annual Review of Plant Biology* 72(1): 387–410.

Lörch, M., J. Mutke, M. Weigend, and F. Luebert. 2021. Historical biogeography and climatic differentiation of the *Fulcaldea-Archidasyphyllum-Arnaldoa* clade of Barnadesioideae (Asteraceae) suggest a Miocene, aridity-mediated Andean disjunction associated with climatic niche shifts. *Global and Planetary Change* 201: 103495.

Mandel, J. R., R. B. Dikow, V. A. Funk, R. R. Masalia, S. E. Staton, A. Kozik, R. W. Michelmore, et al. 2014. A target enrichment method for gathering phylogenetic information from hundreds of loci: An example from the Compositae. *Applications in Plant Sciences* 2(2): e1300085.

Mandel, J. R., R. B. Dikow, and V. A. Funk. 2015. Using phylogenomics to resolve mega-families: An example from Compositae. *Journal of Systematics and Evolution* 53(5): 391–402.

Mandel, J. R., M. S. Barker, R. J. Bayer, R. B. Dikow, T. G. Gao, K. E. Jones, S. Keeley, et al. 2017. The Compositae tree of life in the age of phylogenomics. *Journal of Systematics and Evolution* 55(4): 405–410.

Mandel, J. R., R. B. Dikow, C. M. Siniscalchi, R. Thapa, L. E. Watson, and V. A. Funk. 2019. A fully resolved backbone phylogeny reveals numerous dispersals and explosive diversifications throughout the history of Asteraceae. *Proceedings of the National Academy of Sciences, USA* 116(28): 14083–14088.

Manni, M., M. R. Berkeley, M. Seppey, and E. M. Zdobnov. 2021. BUSCO: Assessing genomic data quality and beyond. *Current Protocols* 1(12): e323.

Mendes, F. K., and M. W. Hahn. 2016. Gene tree discordance causes apparent substitution rate variation. *Systematic Biology* 65(4): 711–721.

Minh, B. Q., H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A. Von Haeseler, R. Lanfear, and E. Teeling. 2020. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* 37(5): 1530–1534.

Mitchell, N., P. O. Lewis, E. M. Lemmon, A. R. Lemmon, and K. E. Holsinger. 2017. Anchored phylogenomics improves the resolution of evolutionary relationships in the rapid radiation of *Protea* L. *American Journal of Botany* 104(1): 102–115.

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

One Thousand Plant Transcriptomes Initiative. 2019. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574(7780): 679–685.

Palazzesi, L., J. Pellicer, V. D. Barreda, B. Loeuille, J. R. Mandel, L. Pokorny, C. M. Siniscalchi, et al. 2022. Asteraceae as a model system for evolutionary studies: From fossils to genomes. *Botanical Journal of the Linnean Society* 200(2): 143–164.

Panero, J. L., and V. A. Funk. 2002. Toward a phylogenetic subfamilial classification for the Compositae (Asteraceae). *Proceedings of the Biological Society of Washington* 115(4): 760–773.

Panero, J. L., and V. A. Funk. 2008. The value of sampling anomalous taxa in phylogenetic studies: Major clades of the Asteraceae revealed. *Molecular Phylogenetics and Evolution* 47(2): 757–782.

Panero, J. L., S. E. Freire, L. Ariza Espinar, B. S. Crozier, G. E. Barboza, and J. J. Cantero. 2014. Resolution of deep nodes yields an improved backbone phylogeny and a new basal lineage to study early evolution of Asteraceae. *Molecular Phylogenetics and Evolution* 80: 43–53.

Pascual-Díaz, J. P., S. Garcia, and D. Vitales. 2021. Plastome diversity and phylogenomic relationships in Asteraceae. *Plants* 10(12): e2699.

Pease, J. B., J. W. Brown, J. F. Walker, C. E. Hinchliff, and S. A. Smith. 2018. Quartet sampling distinguishes lack of support from conflicting support in the green plant tree of life. *American Journal of Botany* 105(3): 385–403.

Posit Team. 2025. RStudio: Integrated Development for R. Posit Software, PBC, Boston, Massachusetts, USA.

Prjibelski, A., D. Antipov, D. Meleshko, A. Lapidus, and A. Korobeynikov. 2020. Using SPAdes de novo assembler. *Current Protocols in Bioinformatics* 70(1): e102.

R Core Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Revell, L. J. 2012. phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* 3(2): 217–223.

Rieseberg, L. H. 1995. The role of hybridization in evolution: Old wine in new skins. *American Journal of Botany* 82(7): 944–953.

Rieseberg, L. H. 2006. Hybrid speciation in wild sunflowers. *Annals of the Missouri Botanical Garden* 93(1): 34–48.

Roch, S., and T. Warnow. 2015. On the robustness to gene tree estimation error (or lack thereof) of coalescent-based species tree methods. *Systematic Biology* 64(4): 663–676.

Salichos, L., and A. Rokas. 2013. Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature* 497: 327–331.

Schliep, K. P. 2011. phangorn: Phylogenetic analysis in R. *Bioinformatics (Oxford, England)* 27(4): 592–593.

Schranz, M. E., S. Mohammadin, and P. P. Edger. 2012. Ancient whole genome duplications, novelty and diversification: The WGD radiation lag-time model. *Current Opinion in Plant Biology* 15(2): 147–153.

Semple, J. C., and K. Watanabe. 2009. A review of chromosome numbers in Asteraceae with hypotheses on chromosomal base number, 61–72. In V. A. Funk, A. Susanna, T. F. Stuessy, and R. J. Bayer [eds.], Systematics, evolution, and biogeography of Compositae. International Association of Plant Taxonomists, Vienna, Austria.

Sensalari, C., S. Maere, and R. Lohaus. 2022. ksrates: Positioning whole-genome duplications relative to speciation events in K_S distributions. *Bioinformatics (Oxford, England)* 38(2): 530–532.

Smith, S. A., M. J. Moore, J. W. Brown, and Y. Yang. 2015. Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. *BMC Evolutionary Biology* 15(1): e150.

Solís-Lemus, C., and C. Ané. 2016. Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. *PLoS Genetics* 12(3): e1005896.

Solís-Lemus, C., P. Bastide, and C. Ané. 2017. PhyloNetworks: A package for phylogenetic networks. *Molecular Biology and Evolution* 34(12): 3292–3298.

Soltis, P. S., and D. E. Soltis. 2016. Ancient WGD events as drivers of key innovations in angiosperms. *Current Opinion in Plant Biology* 30: 159–165.

Stamatakis, A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9): 1312–1313.

Stull, G. W., K. K. Pham, P. S. Soltis, and D. E. Soltis. 2023. Deep reticulation: The long legacy of hybridization in vascular plant evolution. *The Plant Journal* 114(4): 743–766.

Susanna, A., B. G. Baldwin, R. J. Bayer, J. M. Bonifacino, N. Garcia-Jacas, S. C. Keeley, J. R. Mandel, et al. 2020. The classification of the Compositae: A tribute to Vicki Ann Funk (1947–2019). *Taxon* 69(4): 807–814.

Tank, D. C., J. M. Eastman, M. W. Pennell, P. S. Soltis, D. E. Soltis, C. E. Hinchliff, J. W. Brown, et al. 2015. Nested radiations and the pulse of angiosperm diversification: Increased diversification rates often follow whole genome duplications. *The New Phytologist* 207(2): 454–467.

Tellería, M. C., and L. Katinas. 2005. The unusual occurrence of tricolpate pollen within Mutisieae (Asteraceae). *Grana* 44(2): 91–97.

Than, C., D. Ruths, and L. Nakhleh. 2008. PhyloNet: A software package for analyzing and reconstructing reticulate evolutionary relationships. *BMC Bioinformatics* 9(1): 322.

Tillich, M., P. Lehwark, T. Pellizzer, E. S. Ulbricht-Jones, A. Fischer, R. Bock, and S. Greiner. 2017. GeSeq: Versatile and accurate annotation of organelle genomes. *Nucleic Acids Research* 45(W1): W6–W11.

Walter, G. M., R. J. Abbott, A. C. Brennan, J. R. Bridle, M. Chapman, J. Clark, D. Filatov, et al. 2020. *Senecio* as a model system for integrating studies of genotype, phenotype and fitness. *The New Phytologist* 226: 326–344.

Watson, L. E., C. M. Siniscalchi, and J. R. Mandel. 2020. Phylogenomics of the hyperdiverse daisy tribes: Anthemideae, Astereae, Calenduleae, Gnaphalieae, and Senecioneae. *Journal of Systematics and Evolution* 58(6): 841–852.

Weitemier, K., S. C. K. Straub, R. C. Cronn, M. Fishbein, R. Schmickl, A. McDonnell, and A. Liston. 2014. Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenomics. *Applications in Plant Sciences* 2(9): e1400042.

Wen, D., Y. Yu, J. Zhu, and L. Nakhleh. 2018. Inferring phylogenetic networks using PhyloNet. *Systematic Biology* 67(4): 735–740.

White, O. W., J. A. Reyes-Betancort, M. A. Chapman, and M. A. Carine. 2020. Geographical isolation, habitat shifts and hybridisation in the diversification of the Macaronesian endemic genus *Argyranthemum* (Asteraceae). *The New Phytologist* 228(6): 1953–1971.

Wick, R. R., M. B. Schultz, J. Zobel, and K. E. Holt. 2015. Bandage: Interactive visualization of de novo genome assemblies. *Bioinformatics (Oxford, England)* 31(20): 3350–3352.

Wickham, H. 2016. ggplot2: Elegant graphics for data analysis. Springer-Verlag, New York, New York, USA.

Wodehouse, R. P. 1929. Pollen grains in the identification and classification of plants. IV. The Mutisieae. *American Journal of Botany* 16: 297–313.

Wood, T. E., N. Takebayashi, M. S. Barker, I. Mayrose, P. B. Greenspoon, and L. H. Rieseberg. 2009. The frequency of polyploid speciation in vascular plants. *Proceedings of the National Academy of Sciences, USA* 106(33): 13875–13879.

Xie, Y. 2024. bookdown: Authoring books and technical documents with R Markdown. CRC Press, Boca Raton, Florida, USA.

Zhang, C., M. Rabiee, E. Sayyari, and S. Mirarab. 2018. ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* 19(Suppl 6): 15–30.

Zhang, C., C.-H. Huang, M. Liu, Y. Hu, J. L. Panero, F. Luebert, T. Gao, and H. Ma. 2021. Phylotranscriptomic insight into Asteraceae diversity, polyploidy, and morphological innovation. *Journal of Integrative Plant Biology* 63(7): 1273–1293.

Zuntini, A. R., T. Carruthers, O. Maurin, P. C. Bailey, K. Leempoel, G. E. Brewer, N. Epitawalage, et al. 2024. Phylogenomics and the rise of the angiosperms. *Nature* 629(8013): 843–850.

SUPPORTING INFORMATION

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

Appendix S1: Voucher information for taxa (Hyb-Seq and transcriptomes) included in this study, including ingroup and outgroup designations, taxonomic placement (family, tribe, genus, and species), species name with authority, collection locality, collector(s), collection or accession number, herbarium of deposit, collection date, herbarium sheet barcode or identifier, and data source. Raw read count for newly generated transcriptomes is also included.

Appendix S2: The base of the phylogenetic tree of Asteraceae, indicating which taxa were included in each PhyloNetwork run. A colored block on the right indicates the taxon in that row was included in the subtree analysis designated at the top of each column.

Appendix S3: Attributes of chloroplast extraction for all sequences, including whether it was used as a reference plastome, the reference plastome used, and the percentage of missing data when mapped to the respective plastome.

Appendix S4: Collinearity plot generated through Mauve showing plastome rearrangements.

Appendix S5: Phylogenetic reconstruction of chloroplast off-target sequences extracted from target-enriched sequencing data of all taxa. Percentages of missing data indicated by colored circles on tips.

Appendix S6: Comparison of chloroplast phylogenetic tree with concatenated nuclear tree and coalescent-based nuclear tree. Tip labels are shown as species in one tree and tribes in the next.

Appendix S7: Tanglegram of coalescent (ASTRAL-III; left) and concatenated (RAxML; right) trees of all 267 taxa. Lines

between phylogenies connect matching tip labels of trees. Local posterior probability (left) and bootstrap (right) support values are designated at each node of the ASTRAL-III and RAxML tree, respectively.

Appendix S8: RAxML concatenated tree of all 267 taxa rooted to Goodeniaceae taxa. Numbers at each node indicate bootstrap (BS) support values. Branch lengths represent the number of nucleotide substitutions per site. Tip labels correspond to the currently recognized tribe and species name.

Appendix S9: ASTRAL-III tree of all 267 taxa rooted to Goodeniaceae taxa. Numbers at each node indicate local posterior probability (LPP) support values. Branch lengths are in coalescent units. Tip labels correspond to the currently recognized tribe and species name.

Appendix S10: BUSCO scores, number of genes, and K_S peak values of all 69 transcriptomes used in this study.

Appendix S11: Quartet Sampling results of RAxML concatenated tree. The Quartet Concordance (QC), Quartet Differential (QD), and Quartet Informativeness (QI) scores are found at each node as QC/QD/QI. Colored circles at the branch tips correspond to the QF values of each taxon and are colored according to the legend on the top left. Individual QF values can be found in Appendix S14. Discordant backbone nodes are circled and numbered 1–5 in purple font.

Appendix S12: Quartet Sampling results of ASTRAL-III coalescent tree. The Quartet Concordance (QC), Quartet Differential (QD), and Quartet Informativeness (QI) scores are found at each node as QC/QD/QI. Colored circles at the branch tips correspond to the QF values of each taxon and are colored according to the legend on the top left. Individual QF values can be found in Appendix S14. Discordant backbone nodes are circled and numbered 1–5 in purple font.

Appendix S13: Quartet Sampling node score results for the RAxML concatenated and ASTRAL-III coalescent trees. Results are associated with the trees found in Appendices S11 and S12, respectively.

Appendix S14: PhyParts results of the concatenated RAxML phylogeny. Numbers above the branch indicate the number of concordant gene trees, while the numbers below the branch indicate the number of conflicting gene trees. Blue: concordant gene trees, yellow: top alternative bipartition, orange: all other alternative bipartitions, light gray: missing data, dark gray: uninformative data.

Appendix S15: K_S plots of all 69 transcriptomes generated with wgd v2. The *x*-axis indicates the node averaged K_S values, while the *y*-axis indicates the node averaged number of duplicated genes. Estimated whole genome duplication peaks are measured with the lognormal optimized curves a, b, c, and sometimes d and are listed in Appendix S10.

Appendix S16: (A) Histogram of modes of lognormal (a, b, c, and d) optimized *Ks* distributions as inferred from wgd v2 for all Asteraceae taxa. (B) Modes of lognormal (a, b, c, and d) optimized *Ks* distribution for all taxa assessed, plotted by tribe and/or family. Pink shaded regions indicate the clustering bins calculated for the K_S values reported by Huang et al. (2016). (C) Mixed paralog/ortholog *Ks* plot for *Acicarpha procumbens* (Calyceraceae) compared to *Scaevola taccada* (Goodeniaceae), *Arnoldoa macbrideana* (Asteraceae), and *Gerbera delavayi* (Asteraceae) modeled with an exponential-lognormal mixture in wgd v2. Anchor *Ks* components represent clustered peaks of collinear duplicate pairs. Estimated WGD events are indicated by peaks and highlighted by their lognormal optimized modes. Vertical dashed lines represent the modes of the orthologous *Ks* distributions between *A. procumbens* and related taxa to provide relative phylogenetic positioning of WGD. Ortholog *Ks* estimates are rescaled to the focal species' substitution rate.

Appendix S17: All PhyloNetworks subtree networks plotted on the phylogeny. The direct output of the network is represented by a solid line, while potential networks at the tribal level are indicated by a dotted line. Both lines are included because sampling was sparse and inferences are only being made at the tribal level.

Appendix S18: Plot of the log-likelihood values from 0 to 10 for each PhyloNetworks subtree analysis. The *x*-axis indicates the network number (from 0 to 10). The *y*-axis indicates the loglikelihood value of that individual subtree analysis. The point colored red in the plot indicates the optimal number of reticulation events chosen for that subtree.

Appendix S19: The six subtree PhyloNetwork runs and their individual placement on the phylogeny. The number next to the analysis name indicates the optimal number of reticulation events. Numbered black circles represent the same reticulation event in the network (top) and phylogeny (bottom). Black branches indicate bifurcating branches, while colored edges indicate reticulate branches. Hybrid edges connect two parental lineages to a single hybrid node, representing gene flow into the recipient lineage. These edges are shown as lines pointing toward the recipient taxon's crown as a dot. Numbers at nodes in the networks indicate local posterior probability (LPP) support values, while colored numbers on hybrid edges denote estimated proportions of ancestry from each parental lineage. Networks were generated with Dendroscope v.3.7.6 (Huson et al., 2007).

How to cite this article: Ellestad, P. A., E. R. Moore-Pollard, C. M. Siniscalchi, R. Thapa, L. E. Watson, J. M. Bonifacino, and J. R. Mandel. 2026. Ancient hybridization and phylogenetic discordance: Exploring evolutionary complexity in Asteraceae. *Applications in Plant Sciences* 14(3): e70058. https://doi.org/10.1002/aps3.70058