Phylogenomic challenges in polyploid-rich lineages: Insights from paralog processing and reticulation methods using the complex genus *Packera* (Asteraceae: Senecioneae)

Erika R. Moore-Pollard[1,2], Paige A. Ellestad[2], Jennifer R. Mandel[2,3]

[1]*Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, Connecticut 06269, USA*

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

[3]*ACRE: Institute for Agricultural and Conservation Research and Education, University of Memphis, Memphis, Tennessee 38152, USA*

ABSTRACT

Phylogenomic discordance is pervasive and cannot always be resolved by increasing the amount of sequencing data alone. Biological processes such as polyploidy, hybridization, and incomplete lineage sorting are major contributors to discordance and must be accounted for to avoid misleading evolutionary interpretations. To better understand how these processes influence phylogenetic reconstruction, we conducted a comprehensive phylogenomic study in the complex genus *Packera*. With over 90 species and varieties, 40% of which exhibit polyploidy, aneuploidy, or other cytological complexities, *Packera* presents significant challenges for phylogenetic reconstruction. Given these complexities, we assessed different published paralog processing methods on the resulting evolutionary relationships and phylogenetic support of this group. We

then applied three of these methods to evaluate their impact on tree topology and our understanding of *Packera*'s evolutionary history by constructing a time-calibrated phylogeny, reconstructing historical biogeography, and testing for ancient reticulation. Phylogenetic outcomes varied based on the paralog processing method used, with no method performing the best over others. Our findings highlight the large impact of orthology inference and paralog processing on phylogenomic analyses, particularly in polyploid-rich groups such as *Packera*, and we offer guidance on methodological impacts along with practical recommendations. We note that gaining a robust understanding of *Packera*'s evolutionary history requires more than computational approaches alone. While technological advancements have greatly expanded our ability to analyze genomic data, effective phylogenomic research still relies on strong taxon sampling and detailed species knowledge. Without careful attention to the biological context, such as reproductive boundaries, cytological variation, ecological interactions, and historical biogeographic processes, phylogenomic studies risk misinterpreting evolutionary history and processes. By accounting for these factors, we can begin to improve the accuracy of evolutionary reconstructions and gain deeper insights into the complex history of plant diversification.

KEYWORDS

Compositae, discordance, hybridization, molecular systematics, orthology inference

INTRODUCTION

The field of phylogenetics is drastically changing and now routinely incorporates multi-locus and genomic data. While this expansion has enabled deeper insights into evolutionary relationships across the tree of life, it has also revealed persistent and complex challenges in phylogenetic inference. One significant challenge is gene tree discordance, where phylogenetic trees estimated from individual loci disagree with each other and/or with the inferred species tree (Degnan and Rosenberg 2009; Mendes et al. 2016). These incongruences are not simply a consequence of limited data; rather, they often reflect differences driven by underlying biological processes such as incomplete lineage sorting (ILS), gene duplication and loss, and hybridization or introgression (Pease et al. 2018). Such processes are especially prevalent in plants, where whole-genome duplications (WGDs), ancient hybridization, and reticulate evolution are widespread and recurrent (Funk 1985; Vision et al. 2000; Wendel and Cronn 2003; Peirson et al. 2012; Panchy et al. 2016; Rose et al. 2021).

Polyploidy, or WGD, is a dominant evolutionary force in angiosperms, with estimates suggesting that up to 70% of extant flowering plants have polyploid ancestry (Grant 1971; Masterson 1994; Soltis and Soltis 1999; One Thousand Plant Transcriptomes Initiative 2019). These events gave rise to extensive paralogy in plant genomes, which poses significant challenges for orthology inference and phylogenetic analysis. Although these issues are particularly pronounced in plants (e.g., Asparagaceae, Bentz and Leebens-Mack 2024; Rosaceae, Morales-Briones et al. 2018, 2021a; and Amaranthaceae, Morales-Briones et al. 2021b), they are by no means exclusive to them. Similar complications have been documented in non-plants, such as

67 African clawed frogs (Evans et al. 2004), desert rodents (Gallardo et al. 2004), alpine
68 butterflies (Nice et al. 2013), and fish (Braasch and Postlethwait 2012), indicating that
69 complex histories of gene duplication and horizontal gene transfer create phylogenetic
70 uncertainty across the tree of life. As the scale and scope of phylogenomic datasets
71 expand, so too does the need to carefully model and interpret the biological processes
72 underlying gene tree discordance. Fortunately, the growing availability of phylogenomic
73 data now makes it possible to start untangling these complex histories in ways that were
74 not previously feasible; however, doing so requires tools and models that explicitly
75 account for these processes.

76 In response, several bioinformatic tools and pipelines have been developed to
77 infer orthology under increasingly realistic evolutionary models. For instance, softwares
78 such as HybPiper (Johnson et al. 2016) and Phyluce (Faircloth 2016) implement distinct
79 strategies for identifying and handling putative paralogs, with some pipelines retaining
80 only a single gene copy (e.g., HybPiper), while others exclude all loci which exhibit
81 signs of paralogy (e.g., Phyluce). Although Phyluce has shown greater success in some
82 groups (e.g., *Antennaria* Gaertn.; Thapa et al. 2020), HybPiper typically outperforms
83 Phyluce across most plant lineages because of its ability to better resolve issues
84 associated with polyploidy.

85 Additional pipelines have been developed that implement distinct strategies for
86 identifying and handling putative paralogs using graph- or tree-based methods (e.g.,
87 paragone-nf; Yang and Smith 2014; Jackson et al. 2023), whereas others consider
88 paralogy during species tree estimation itself (e.g., ASTRAL-Pro; Zhang et al. 2020;
89 Zhang and Mirarab 2022a). These different approaches can yield markedly different

90 phylogenetic outcomes, particularly in polyploid-rich systems where paralogs may be
91 the rule rather than the exception. Yet, few studies have systematically compared these
92 methods to understand their impact on phylogenetic and downstream evolutionary
93 inferences.

94 This is especially relevant for groups such as *Packera* Á. Löve & D. Löve
95 (Asteraceae: Senecioneae), a North American genus of about 91 species known for
96 polyploidy, complex patterns of hybridization, and poorly resolved taxonomy. An
97 estimated 40% of *Packera* taxa present polyploidy, aneuploidy, and/or other cytological
98 disturbances (Barkley 1988; Trock 2006), complicating phylogenetic reconstruction of
99 this group (Bain and Golden 2000). Given this, *Packera* provides an ideal system for
100 testing how alternative strategies for handling paralogs influence phylogenomic
101 inference, owing to its extensive polyploidy, ecological and geographic diversity, and
102 history of reticulation; all factors that increase the potential for gene tree discordance.

103 Here we present the most comprehensive phylogenomic investigation of *Packera*
104 to date, sampling nearly 93% of the genus and leveraging a high-resolution Hyb-Seq
105 dataset. We applied a robust framework for evaluating the impact of gene duplication by
106 comparing multiple paralog processing methods, including pipelines that remove all
107 paralogs, those that model duplication and loss, and those that retain a single
108 representative gene copy. We assessed how these approaches influence gene tree
109 concordance, species tree topology, and node support. We then tested how three of
110 these methods (a standard paralog processing method, a method that removes all
111 paralogs, and a selective paralog-aware method; see methods below) alter the
112 interpretations of downstream macroevolutionary analyses including divergence dating,

113 historical biogeography, and detection of ancient introgression. Our results demonstrate
114 that the choice of paralog processing approach strongly influences inferred species
115 relationships, gene tree concordance, node support, and evolutionary inferences,
116 highlighting the profound effect that paralog-aware decisions can have in
117 phylogenomics. Additionally, this study provides a broadly applicable framework for
118 assessing paralogs and their downstream impacts on phylogenetic inference in systems
119 shaped by gene duplication, reticulation, and rapid diversification.
120

121 **MATERIALS & METHODS**

122 *Specimen Collection*

123 A nearly complete sampling of currently recognized *Packera* species and
124 varieties were used in this study (62 *Packera* species, 22 varieties, and one
125 nothospecies; totaling 85 *Packera* taxa). *Packera* species were selected based on
126 information published in the International Plant Names Index (IPNI; available at
127 http://www.ipni.org) and then curated from discussions with botanists that have
128 previously specialized in *Packera* (D. Trock & J. Bain, pers. comm.). Outgroup taxa from
129 tribes Senecioneae (n = 19) and Anthemideae (n = 4) were selected by consulting
130 recent studies of the group (Bain and Golden 2000; Pelser et al. 2010; Mandel et al.
131 2019). Additionally, three taxa from tribe Calenduleae were included for downstream
132 analyses that require more distantly related outgroups that do not have a known WGD
133 event (Barker et al. 2016; Huang et al. 2016; Zhang et al. 2021). A complete list of the
134 111 sampled species, herbarium vouchers, and their publication status can be found in
135 Supplementary Table 1.

136 *DNA Extraction and Sequencing*

137 Leaf tissues were obtained from herbarium specimens or silica-dried field
138 collections. Dried leaf tissue (~25 mg) was ground using a Bead Mill 24 Homogenizer
139 (ThermoFisher Scientific, Atlanta, Georgia, USA), and total DNA was extracted using
140 the Omega Bio-Tek Kit (Atlanta, Georgia, USA) following manufacturer's protocols with
141 the addition of polyvinylpyrrolidone (PVP) and ascorbic acid to the first extraction buffer
142 (10 mL SQ1 buffer, 100 mg PVP, 90 mg ascorbic acid per reaction). Samples of
143 *Packera* and outgroup species that were obtained from other researchers followed
144 different DNA extraction protocols which are described in Supplementary Table 1. All
145 DNA samples used in this study were quality checked using a Nanodrop 2000 (Thermo
146 Fisher Scientific, Carlsbad, California, USA) and quantified using a Qubit High
147 Sensitivity assay (ThermoFisher Scientific, Oregon, USA). If the DNA had low
148 concentrations or possible contaminants, the samples were cleaned following the
149 E.Z.N.A. Cycle Pure Kit (Omega Biotek, Georgia, USA) and re-quantified with the Qubit
150 High Sensitivity (HS) assay. Up to 1 µg DNA per sample was sonicated with a QSonica
151 machine (ThermoCube, New York, USA) to generate fragment sizes within the 400-500
152 base pair (bp) range. The amount of DNA used was dependent on the initial total DNA
153 concentration of the targeted sample. Time allotted for sonication of each sample varied
154 depending on approximate length and degree of fragmentation of the DNA. Prior to
155 sonication, total DNA was visualized on a 1% agarose gel in 1X TBE and GelRed 3x
156 (Biotium) and the approximate length was estimated by comparing to a 100 bp
157 GeneRuler DNA ladder (ThermoFisher Scientific, Vilnius, Lithuania). If the DNA sample
158 appeared smaller than 400 bp, it was not sonicated.

159 Libraries for each sample were prepared using the NEBNext Ultra II DNA Library
160 Prep and NEBNext Multiplex Oligos for Illumina Kits (New England Biolabs, Ipswich,
161 Massachusetts, USA) following the NEBNext Ultra II Version 5 protocol with size
162 selection on DNA fragments at 300-400 bp range. The protocol was adjusted by halving
163 all amounts of reagents and DNA used. The barcoded libraries (NEBNext multiplex
164 oligos) were quality checked with an Agilent Bioanalyzer (Agilent Technologies, Santa
165 Clara, California, USA) and quantified by Qubit HS assay. Libraries were then used for
166 targeted sequence capture with the custom MyBaits probe set, Compositae-1061
167 (Mandel et al. 2014), from Arbor Biosciences (Ann Arbor, Michigan, USA) following the
168 manufacturer's protocol (v. 2.3.1). Approximately 500 ng of total DNA library in 6 µl of
169 library, or equal amount of DNA per sample when pooling (maximum of 8 samples), was
170 used in each MyBaits reaction for hybridization with the baits for 36 hours. Hybridized
171 targets were recovered using Dynabeads M-280 Streptavidin (ThermoFisher Scientific,
172 Vilnius, Lithuania) following the manufacturer's protocol. Captured targets were
173 amplified and quantified using KAPA library quantification kits (Kapa Biosystems,
174 Wilmington, Massachusetts, USA). A 1:4 ratio of library was added to each of the
175 captured target pools to obtain off-target plastid sequence data. Libraries were
176 sequenced on Illumina MiSeq (300 cycles) and NovaSeq 6000 (300 cycles) instruments
177 at Psomagen (Rockville, Maryland, USA) across several runs. Sequence stats were
178 obtained using the pxlssq function in Phyx v. 1.3.1 (Supplementary Table 2) (Brown et
179 al. 2017).

180 *Sequence Processing*

181 Raw FASTQ data were trimmed of low quality bases and adapter contamination
182 using Trimmomatic v. 0.36 (Bolger et al. 2014). The Sliding Window quality filter was
183 used in this analysis (illuminaclip 2:30:10, leading 20, trailing 20, sliding window 5:20).
184 Cleaned reads with a minimum length of 36 bp and corresponding forward and reverse
185 pairs were retained. FastQC v. 0.11.9 (Andrews 2010) was used to ensure good
186 sequence quality.

187 *Plastid Assembly*

188 Chloroplast sequences were extracted from trimmed reads using two
189 approaches. First, the GetOrganelle v. 1.7.7.1 toolkit (Jin et al. 2020) was used to
190 assemble circularized plastomes for all sequences with sufficient off-target sequencing
191 data. The recommended parameters for embryophyte plant plastomes were used,
192 except that the number of rounds and the disentangle time limit were increased (-R 75
193 and --disentangle-time-limit 7200). All complete circular plastomes were annotated
194 using GeSeq (Tillich et al. 2017). From these, a complete *Packera* representative
195 plastome was chosen and used as a reference. Alternatively, all reads were mapped to
196 the reference plastome using Bowtie2 (Langmead and Salzberg 2012), variants were
197 called using SAMtools v. 1.18 and BCFtools v. 1.18 (Danecek et al. 2021), and
198 consensus sequences were created using seqtk v. 1.4 (https://github.com/lh3/seqtk).
199 For the aligned consensus sequences, model selection and maximum likelihood
200 phylogenetic analysis were then conducted using IQ-TREE v. 2.2.2 (Minh et al. 2020)
201 and visualized using 'ggtree' (Yu et al. 2017) in R v. 4.3.1 (R Core Team 2016; RStudio
202 2020).

203 *Nuclear Ortholog Assembly*

204 Cleaned FASTQ data were assembled using the HybPiper v. 2.1.6 (Johnson et
205 al. 2016) pipeline to match the reads to the target loci contained in the Compositae-
206 1061 probe set (Mandel et al. 2014). In HybPiper, the command 'assemble' was used to
207 map the paired-end reads to the target loci using BWA v. 0.7.17 (Li and Durbin 2009.
208 These were then assembled into contigs with specified kmer lengths (21, 33, 55, 77, 99)
209 using SPAdes v. 3.5 (Bankevich et al. 2012). Paralog detection was carried out using
210 the 'paralog_retriever' command
211 (https://github.com/mossmatters/HybPiper/wiki/Paralogs), which flags loci that indicate
212 potential paralogs, then extracts the assemblies that were flagged. A supercontig
213 multifasta for each locus was generated and aligned in MAFFT v. 7.407 (Katoh and
214 Standley 2013) using default parameters.

215 *Phylogenetic Analyses*

216 Given the complex history of hybridization and polyploidization events in
217 *Packera*, along with the large number of paralogs flagged from HybPiper in the current
218 study (Supplementary Table 3), we sought to compare how different published
219 orthology inference methods and species tree estimation tools impacted the inferred
220 evolutionary relationships and support of this group. Our dataset is uniquely adapted to
221 tackle this issue given dense taxon sampling and high locus recovery across both
222 orthologs and paralogs, combined with decades of natural history and systematics work
223 in the group, enabling more thorough investigations into these challenges.

224 We generated multiple nuclear phylogenies utilizing different paralog processing
225 tools to infer orthology and species tree estimation: Concatenation (individual gene

226 histories not accounted for), HybPiper with ASTRAL-III (hereafter 'ASTRAL'; putative
227 single-copy orthologous loci are chosen based on coverage depth and sequence
228 similarity concerning the gene references from Compositae-1061, then inferred using
229 coalescent methods), HybPiper with wASTRAL (hereafter 'wASTRAL'; similar to
230 ASTRAL, but low support gene tree branches are contracted prior to species tree
231 estimation), HybPiper with ASTRAL-Pro (hereafter 'ASTRAL-Pro'; all single-copy and
232 multi-copy genes identified by HybPiper are considered, and the species tree that
233 optimizes a quartet similarity measure is chosen), and the 'one-to-one orthologs' (1-to-
234 1), 'maximum inclusion' (MI), 'monophyletic outgroup' (MO), and 'rooted ingroup' (RT)
235 paralog processing methods by Yang and Smith (2014) with species trees inferred
236 using ASTRAL-III (further details below; refer to Supplementary Table 3, Supplementary
237 Fig. 1). By performing comparisons among methods, we aimed to gain a better
238 understanding of the impact of paralogy on phylogenomic studies, and ultimately
239 increase phylogenomic support and concordance across species of complex lineages.
240 *Concatenation.*—Aligned supercontigs from MAFFT were concatenated with
241 FASconCAT-G v. 1.02 (Kück and Longo 2014), then implemented in PartitionFinder v.
242 2.1.1 (Lanfear et al. 2012) to determine the most appropriate nucleotide substitution
243 model (GTR, GTR+Γ, GTR+I+Γ). The overall best fitting model (Supplementary Table 4)
244 was used to build a concatenated nuclear maximum likelihood tree in RAxML v. 8.1.3
245 (Stamatakis 2014) with 1,000 bootstrap (BS) replicates.
246 ASTRAL and wASTRAL.—For the ASTRAL and wASTRAL analyses, nuclear
247 gene alignments were separated based on the most suitable nucleotide substitution
248 model for each gene as determined by PartitionFinder (Supplementary Table 4), then

249 individual gene trees were generated using RAxML. For the ASTRAL analysis, the gene
250 trees were used in ASTRAL-III v. 5.7.3 (Zhang et al. 2018) to generate the species
251 trees. For the wASTRAL analysis, or weighted ASTRAL, a species tree was generated
252 using 'astral-hybrid' in the ASTER program (https://github.com/chaoszhang/ASTER).
253 This method has shown to be more accurate than typical ASTRAL methods since
254 wASTRAL incorporates gene tree uncertainty by weighting quartets based on the
255 confidence levels of the input gene trees. In doing so, wASTRAL reduces the influence
256 of erroneous gene trees, leading to a more reliable species tree estimation (Zhang and
257 Mirarab 2022b). Local posterior probability (LPP) values were estimated for branch
258 support and indicate the probability that the branch is the true branch given the set of
259 gene trees provided (Sayyari and Mirarab 2016).
260 *ASTRAL-Pro.*—ASTRAL-Pro2 (ASTRAL for PaRalogs and Orthologs) v. 1.13.1.3
261 (Zhang et al. 2020; Zhang and Mirarab 2022a), hereafter referred to as ASTRAL-Pro,
262 provides a quartet-based species-tree inference method that allows multi-copy gene
263 trees to be used as input. Thus, all paralogous and orthologous sequences as assigned
264 by HybPiper were used to generate the species tree instead of selecting a single
265 ortholog to represent an individual gene's history (e.g., ASTRAL method). By
266 accommodating multi-copy gene trees, ASTRAL-Pro accounts for gene duplication and
267 loss events, reducing biases that arise from treating paralogs as true orthologs. This
268 approach is intended to improve species tree accuracy, particularly in datasets with
269 extensive gene family expansions, by leveraging more information from complex gene
270 histories. The ASTRAL-Pro pipeline is similar to ASTRAL, as described above, but
271 differs by having the user generate separate alignments and conduct maximum

272 likelihood analyses on the genes that are flagged as paralogs and single-copy contigs.
273 Single-copy and paralogous gene sequences were independently aligned using
274 MAFFT, and gene trees were generated under the best-fitting nucleotide substitution
275 model from PartitionFinder (Supplementary Table 4) in RAxML with 1,000 BS replicates.
276 The RAxML output files were then used as input for ASTRAL-Pro2 to generate the
277 species tree with LPP values at each node.

278 *1-to-1, MI, MO, and RT methods.*—Paralog processing was then carried out
279 using the tree-based orthology inference strategies originally proposed by Yang and
280 Smith (2014) via the paragone-nf v. 1.1.0 pipeline (Jackson et al. 2023;
281 https://github.com/chrisjackson-pellicle/paragone-nf). This pipeline investigates the loci
282 flagged as paralogous from HybPiper and determines which loci can be retained or
283 pruned given specified conditions (Yang and Smith 2014; Morales-Briones et al. 2021a).
284 All orthology inference methods proposed by Yang and Smith (2014)–the 'one-to-one
285 orthologs' (1-to-1), 'maximum inclusion' (MI), 'monophyletic outgroup' (MO), and 'rooted
286 ingroup' (RT) methods–were tested. Briefly, the 1-to-1 approach is the simplest strategy
287 in that it only retains homologs that have no duplications. The MI approach extracts the
288 subtree with the highest number of taxa without taxon duplications. The MO approach
289 filters for homolog trees where outgroup taxa (here tribe Calenduleae: *Calendula*
290 *arvensis*, *Tripteris pinnatilobatum* (Norl.) B.Nord., and *Chyrsanthemoides incana*
291 (Burm.f.) Norl.) are monophyletic and non-repeating. The RT approach searches
292 subtrees of ingroup taxa and roots them by the designated outgroups, then removes the
293 ingroup taxa as rooted trees. In turn, inferred gene duplications, or paralogs, are then

294 pruned from root to tip in the rooted tree (Yang and Smith 2014; Morales-Briones et al.
295 2021a).
296 To aid in consistency across analyses, the HybPiper run described above was
297 used as input instead of using the hybpiper-nf pipeline (Jackson et al. 2023). The three
298 Calenduleae taxa were considered the internal outgroups since a recent whole genome
299 duplication event has not been reported for this tribe, as is required for the MI and MO
300 analyses (Yang and Smith 2014). The resulting trimmed alignments produced via
301 paragone-nf were used as input for RAxML to generate gene trees, and a species tree
302 with LPP support values was generated using ASTRAL-III following the 'ASTRAL'
303 protocol described above.
304 *Measuring phylogenetic tree support*
305 The normalized quartet score was calculated for all orthology inference methods
306 that generated a species tree with ASTRAL-III using the '-q' option. Branch lengths were
307 recalculated from coalescent units to nucleotide substitutions per codon site for each
308 ASTRAL-III species tree using IQ-TREE. Parsimony informative sites (PIS) and pairwise
309 identity (PI) were calculated for each gene matrix using the
310 'parsimony_informative_sites' and 'pairwise_identity' functions, respectively, in phykit v.
311 2.0.1 (Steenwyk et al. 2021). Nuclear and plastome trees were visualized using the
312 package 'phytools' (Revell 2012) in R v. 4.0.5.
313 *Measuring Discordance*
314 Species trees from concatenation, plastid, ASTRAL, wASTRAL, and paragone-nf
315 (1-to-1, MI, MO, and RT) were analyzed for underlying discordance using Quartet
316 Sampling (Pease et al. 2018). ASTRAL-Pro was not analyzed as Quartet Sampling

317 does not allow multi-copy gene trees as input. Quartet Sampling assesses the
318 confidence, consistency, and informativeness of internal tree relationships, and the
319 reliability of each terminal branch by evaluating if a phylogeny has a lack of support due
320 to lack of information, discordance from ILS or introgression, or misplaced or erroneous
321 taxa. Three scores were calculated for each internal branch of the focal tree which
322 determines the consistency of information (Quartet Concordance [QC]), the presence of
323 secondary evolutionary histories (Quartet Differential [QD]), and the amount of
324 information (Quartet Informativeness [QI]). Quartet Fidelity (QF) scores were calculated
325 at each tip to report how frequently a taxon is included in concordant topologies. Using
326 the rooted species trees and concatenated gene alignments from FASconCAT-G
327 (described above), Quartet Sampling was run using default settings on the designated
328 analyses. The concatenation, plastid, ASTRAL, wASTRAL, 1-to-1, MI, and MO species
329 trees were rooted to the outgroup Calenduleae taxon, *Tripteris pinnatilobatum*, while the
330 RT tree was rooted to *Doronicum pardalianches* L. using the *pxrr* function in Phyx
331 (Supplementary Table 3).

332 Additionally, PhyParts v. 0.0.1 (Smith et al. 2015) was used with default settings
333 to quantify and visualize discordance in the ASTRAL, wASTRAL, and paragone-nf (1-to-
334 1, MI, MO, and RT) trees. Similar to Quartet Sampling, ASTRAL-Pro could not be
335 analyzed because of multi-copy gene trees. Additionally, the concatenation and plastid
336 trees could not be analyzed since those analyses do not use individual gene histories
337 as input. Resulting trees containing pie charts indicating whether gene trees were
338 supported, conflicted, or uninformative at each node were generated using a custom
339 python script

340 (https://github.com/mossmatters/phyloscripts/tree/master/phypartspiecharts). PhyParts
341 requires rooted gene trees and a rooted species tree as input. Similar to Quartet
342 Sampling, ASTRAL, 1-to-1, MI, and MO gene trees were rooted to *Tripteris*
343 *pinnatilobatum* and the RT tree was rooted to *Doronicum pardalianches* using the *pxrr*
344 function in Phyx. Rooting gene trees for PhyParts decreased the number of genes used
345 as input since not all gene trees contained the targeted outgroup taxon (see
346 Supplementary Table 3). For phylogenies generated with ASTRAL technologies–
347 ASTRAL, wASTRAL, ASTRAL-Pro, and paragone-nf trees (Supplementary Table 3)–
348 introgression and ILS were visualized across lineages using phytop v. 0.3.1 (Shang et
349 al. 2025) with default settings. The species trees were re-ran with the flag '-t 2' added to
350 be used as input for phytop.

351 *Tree Topology Comparison*

352 To evaluate similarity among the nuclear gene trees, pairwise adjusted
353 Robinson-Foulds (RF_{adj}) distances were calculated as outlined by (Mitchell et al. 2017)
354 between all trees using the *RF.dist()* function in package 'phangorn' (Schliep 2011) in R,
355 with unrooted trees as input and setting the "normalize" argument to TRUE (Steel and
356 Penny 1993). The Calenduleae outgroup taxa were trimmed from the trees in R using
357 the *drop.tip()* function in package 'ape' (Paradis and Schliep 2019) to make
358 concatenation, ASTRAL, ASTRAL-Pro, and paragone-nf trees directly comparable. The
359 RF_{adj} values range from zero to one, indicating whether the tree topologies are identical
360 or completely dissimilar, respectively. A principal component analysis (PCA) was
361 generated from all RF_{adj} values in R to aid in visualization.

362 All nuclear and plastid phylogenies and their associated analyses described
363 above were compared, but for simplicity only the ASTRAL, 1-to-1, and MO trees were
364 analyzed in further analyses. ASTRAL was chosen as it selects the putative single-copy
365 orthologs less rigorously than the remaining methods, creating a baseline method for
366 comparison. We then compared ASTRAL to a method that disregards all genes
367 containing paralogs (1-to-1) and a more rigorous paralog-processing method (MO). We
368 chose MO over MI and RT due to its more stringent ortholog identification, which
369 minimizes the inclusion of paralogs and reduces phylogenetic noise. Additionally, the
370 conservative filtering of MO enhances alignment quality and computational efficiency,
371 providing a more reliable foundation for downstream phylogenetic analyses (Yang and
372 Smith 2014; Morales-Briones et al. 2021a; Jackson et al. 2023).

373 *Molecular Clock Dating*

374 Estimates of divergence times were calculated on the ASTRAL, 1-to-1, and MO
375 trees with RelTime (Tamura et al. 2012, 2018) using the command-line version of the
376 MEGA software, MEGA-CC v. 11.0.11 (Kumar et al. 2012, 2016), with necessary input
377 files generated with MEGA-X (Kumar et al. 2018; Mello 2018). The three Calenduleae
378 taxa were used as outgroup taxa for this analysis.

379 There are no fossils available within *Packera* or tribe Senecioneae, so instead,
380 multiple dating scenarios were tested utilizing both primary and secondary age
381 constraints (Supplementary Table 5). Briefly, a maximum (59.95 MY) and minimum (31
382 MY) age was tested for the Anthemideae/Senecioneae split given those dates are the
383 highest and lowest values within the 95% CI from a previous diversification rate analysis
384 of Asteraceae (Mandel et al. 2019). Next, a maximum age of 34.06 MY was set at the

385 *Packera* crown node from the upper age limit provided by (Mandel et al. 2019).
386 Additionally, a minimum age constraint of 12 MY was tested within the outgroup (the
387 Anthemideae clade) given the presence of a fossil pollen (Wang 2004; Hobbs and
388 Baldwin 2013; Palazzesi et al. 2022). Finally, a maximum age constraint of 14.3 MY
389 was tested for *Pericallis* D.Don (Senecioneae), a genus endemic to the Macaronesia
390 islands (Jones et al. 2014), since that is the estimated age of Porto Santo, the oldest
391 volcanic archipelago within the Madeira Islands (McDougall and Schmincke 1976;
392 Geldmacher et al. 2000). Results of all tested scenarios were compared for biological
393 relevance, though only one scenario was chosen for the discussion. This approach
394 allowed us to focus on the topology that best aligned with prior systematic
395 understanding and showed the strongest overall support across analytical frameworks.
396 *Biogeography*
397 Ancestral ranges were estimated on the dated phylogenies using the package
398 BioGeoBEARS v. 1.1 (Matzke 2014) in R. BioGeoBEARS estimates historical,
399 biogeographic ranges by implementing popular models in a likelihood framework. Some
400 of the most common models used are dispersal-extinction-cladogenesis (DEC; Ree and
401 Smith 2008), the likelihood version of DIVA (DIVALIKE; Ronquist 1997), and the
402 BayArea likelihood version of the range evolution model (BAYAREALIKE; Landis et al.
403 2013). In BioGeoBEARS, founder-event speciation can be added to any of these
404 models (+*J*) with flexibility in determining relative probability. Likelihood Ratio Test (LRT)
405 and Akaike Information Criterion (AIC) were used to determine which model best fits the
406 data.

All *Packera* taxa, along with their four most closely related outgroup genera (*Pericallis*, *Elekmania* B.Nord., *Werneria* Kunth, and *Xenophyllum* V.A.Funk), were coded as present or absent in ten geographic ranges outlined in Takhtajan (1986) (Supplementary Table 6): Arctic Province, Canadian Province, Appalachian Province, Atlantic and Gulf Coast Plain Province, North American Prairies Province, Rocky Mountain and Vancouverian Province, Madrean Region, Asia, Europe, and Central America. Distributions were determined from previously published literature (Freeman and Barkley 1995; Trock 1999, 2006; Gramling 2006; Mahoney and Kowal 2008; Elven et al. 2011; Yeatts et al. 2011). Only four of the 26 outgroup taxa were included to limit the number of geographic ranges used in the analysis. The remaining outgroup taxa were pruned from the dated tree generated with RelTime using the *pxrmt* function in Phyx. Each taxon was limited to a maximum distribution of four ranges.

Reticulation

The Species Networks applying Quartets (SNaQ) function in the Julia v. 1.8.1 (Bezanson et al. 2018) package PhyloNetworks v. 0.15.2 (Solís-Lemus and Ané 2016; Solís-Lemus et al. 2017) was used to investigate reticulate evolution among *Packera* taxa. This package uses unrooted gene trees to infer phylogenetic networks in a pseudolikelihood framework with multi-locus data, while accounting for ILS and reticulation. We tested the fit of multiple models and gradually increased the number of reticulation events (h) from 0 to 10 to find which pseudolikelihood score reached a plateau, given recommendations by (Solís-Lemus et al. 2017). This process is computationally intense, so only clades containing 14 taxa or less that had counter-support nodes in the ASTRAL, 1-to-1, and/or MO trees, as determined by a negative

430 QC score resulting from Quartet Sampling, were used as input for PhyloNetworks, with
431 the remaining taxa pruned from the tree using the *pxrmt* function in Phyx. Given this,
432 five Packera clades were investigated for ancient reticulation events (Figs. 2-4).
433 Additionally, a single species from each major clade was then used to represent
434 backbone relationships throughout the entire phylogeny, totaling six separate
435 PhyloNetwork runs that will be referenced as follows: Backbone, representing 11 taxa
436 that span the entire phylogeny; Eastern, a clade of 12, 13, or 15 species (MO, 1-to-1,
437 and ASTRAL, respectively) either distributed along eastern North America or having
438 known affiliations with those species; California, eight taxa with distributions mainly
439 within California; Mexico, seven (MO and ASTRAL) to eight (1-to-1) taxa with
440 distributions mainly within Mexico; Rocky, a clade of five taxa that are distributed in the
441 Central US, most within the Rocky Mountains; and Arctic, nine taxa that are present in
442 colder climates along the Pacific Northwest and into Canada (Supplementary Table 7).
443 Rocky could not be run for 1-to-1 since fewer than five taxa were present in that target
444 clade.
445

446 **RESULTS**

447 *Sequencing and Matrix Results*

448 Illumina sequencing resulted in approximately 733.3 million reads across 111
449 taxa. Read numbers ranged from 535,934 in *Werneria aretioides* Wedd. to 20,685,576
450 in *Senecio vulgaris* L. (full list in Supplementary Table 2). HybPiper retained 1,058 of
451 the targeted 1,061 loci across all 111 taxa with 846 of those loci having paralog
452 warnings (~80%). The number of loci recovered for each taxon ranged from 566 in

453 *Packera paupercula var. paupercula* (Michx.) Á.Löve & D.Löve to 1,054 in *Calendula*
454 *arvensis* L. (Supplementary Table 8). The number of gene trees varied across analyses,
455 from 285 in the 1-to-1 analysis to 7,388 in the MI analysis (Supplementary Table 3).
456 Parsimony informative sites (PIS) also ranged from 2.11% in the plastid (mapping) tree
457 to 99.64% in the MO tree (Supplementary Table 3). For species trees generated with
458 ASTRAL-III, normalized quartet scores were used to measure gene tree discordance,
459 which ranged from 0.516 for the RT tree to 0.534 for the MO tree (Supplementary Table
460 3), indicating that ~52-53% of the gene trees matched the species tree in those
461 analyses.

462 Off-target plastid sequencing resulted in partial (i.e., mapping; n = 111) to whole
463 (i.e., complete; n = 18) plastomes. Complete plastomes from 17 *Packera* and one
464 *Senecio* species ranged from 150,772 bp to 151,728 bp in size (Supplementary Fig. 2).
465 For each consensus sequence mapped to the reference plastid assembly, *Packera*
466 *cymbalaria*, the percentage of missing data varied drastically, with *Tephroseris*
467 *newcombei* exhibiting the most missing data (93.1%; Supplementary Table 9,
468 Supplementary Fig. 3). The mapped plastid supermatrix with 111 taxa had 2.11% PIS
469 (Supplementary Table 3).

470 *Topological Comparisons and Discordance*

471 Phylogenetic topologies varied depending on sequence type (e.g., nuclear or
472 plastid) as well as paralog processing and phylogenetic tree inference method (i.e.,
473 ASTRAL, ASTRAL-Pro, 1-to-1, etc.) (Fig. 1, Supplementary Fig. 4). When building
474 phylogenies with nuclear data, *Packera* was generally placed as the sister lineage to a
475 clade comprising *Pericallis* and *Elekmania*, with the exception of ASTRAL-Pro which

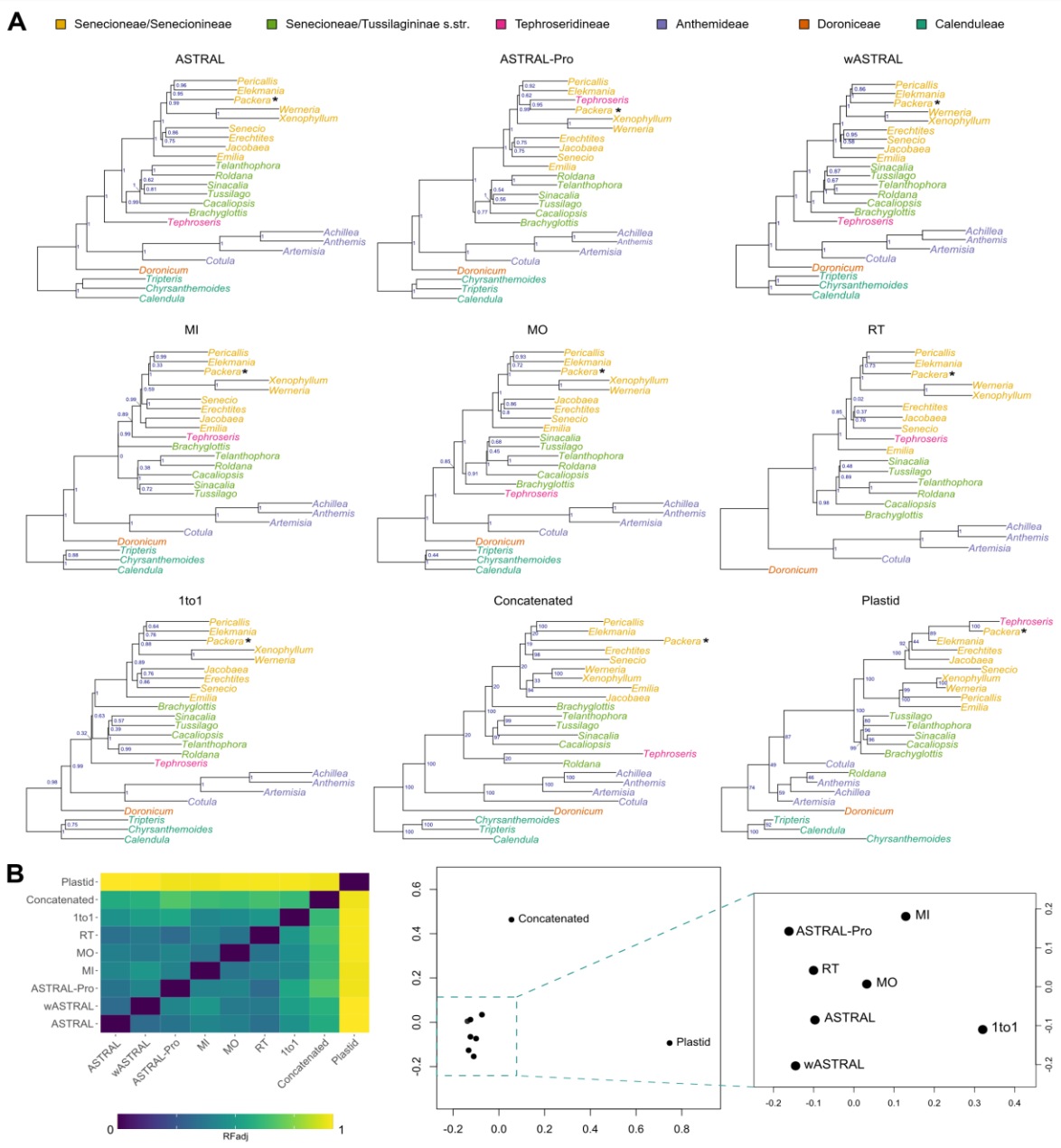


Figure 1. Genus-level phylogenies of all trees generated in this study. **A**) Genera are colored according to their tribal/subtribal classifications following the key on the top of the figure. *Packera* is marked with a black asterisk next to its placement in the trees. Navy blue text at node represents support values (LPP or BS). **B**) Bottom left is a matrix of RF_{adj} values between the nine different trees, with most similar topologies as dark blue squares (0) and least similar as yellow squares (1). The bottom right shows the RF_{adj} values plotted as a principal components analysis (PCA). Points closer together indicate higher similarity, while dots further away indicate higher dissimilarity.

476

placed *Packera* as sister to *Tephroseris* and in a larger clade with *Pericallis* and *Elekmania* (Fig. 1). Interestingly, the plastid data showed a similar relationship to ASTRAL-Pro; however, *Pericallis* was placed unexpectedly in a separate clade with other Senecionineae taxa. All nuclear phylogenies recovered monophyletic tribes and subtribes (e.g., Senecionineae and Tussilaginineae s.str.), though the relationships within the subtribes varied. The mapped plastid tree generally separated the subtribes, apart from *Roldana* (Senecionineae) in Tussilaginineae s. str. (Fig. 1, Supplementary Fig. 4). The placement of *Tephroseris* (tribe Tephroseridineae) varied across phylogenies, with its most consistent placement as sister to Senecioneae (Fig. 1). *Doronicum* was consistently placed sister to Senecioneae, providing support for its elevation to tribal status (Doroniceae; Panero 2005; Fu et al. 2016).

Phylogenetic topologies remained highly incongruent independent of methods used (Figs. 2-4, Supplementary Table 10). The ASTRAL and wASTRAL tree topologies were most similar (RF_{adj} = 0. 324) (Fig. 1, Supplementary Table 10), while the plastid tree was consistently the most dissimilar to the remaining trees (RF_{adj} ranging from 0.895 to 0.914), with the concatenated tree being the next most dissimilar (RF_{adj} ranging from 0.562 to 0. 648; Supplementary Table 10). Additionally, the normalized quartet scores from ASTRAL indicate that only about half of the gene trees match the species trees in the ASTRAL and paragone-nf analyses (Supplementary Table 3). Most of the topological conflict was among *Packera* taxa, with minor differences among the remaining Senecioneae taxa and outgroups (Supplementary Fig. 4). Support values varied depending on which sequence type and method was used, though generally remained low. The wASTRAL and MI trees showed the highest node support with 58

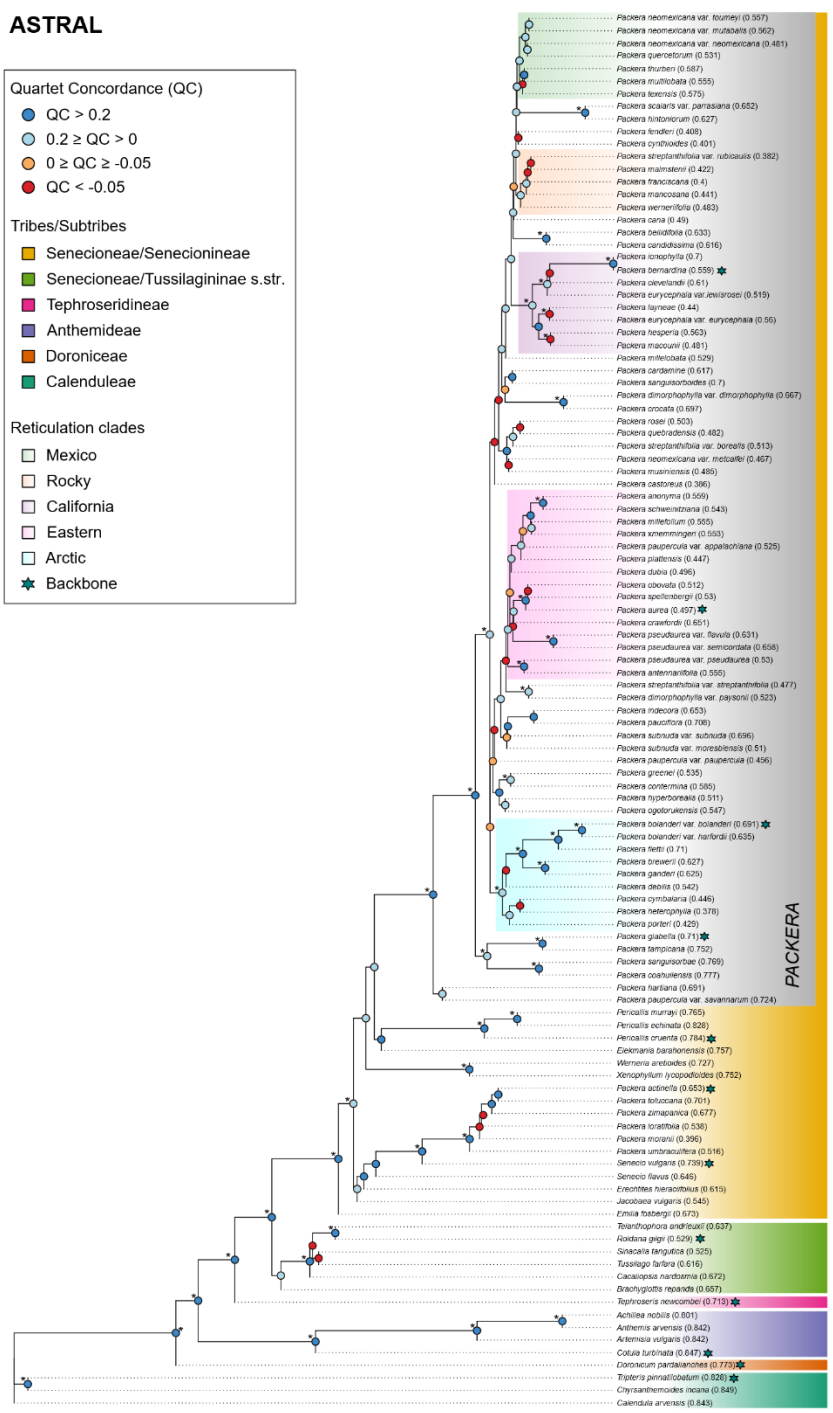


Figure 2. Full phylogeny of the ASTRAL analysis. Tribal/subtribal status is indicated on the right of the phylogeny and follows the color code as indicated in the key. Numbers next to tip labels correspond to the Quartet Fidelity (QF) score from Quartet Sampling, which ranges from 0 to 1, with values closer to 0 indicating erroneous taxa. Circles at nodes show the Quartet Concordance (QC) factors from Quartet Sampling, where red and orange indicate counter-support nodes. Asterisks next to QC circles indicate high support (≥ 95 LPP) at that node. Highlighted clades and teal stars next to some tip labels, indicate clades where reticulation events were tested, following the color key on the left.

501

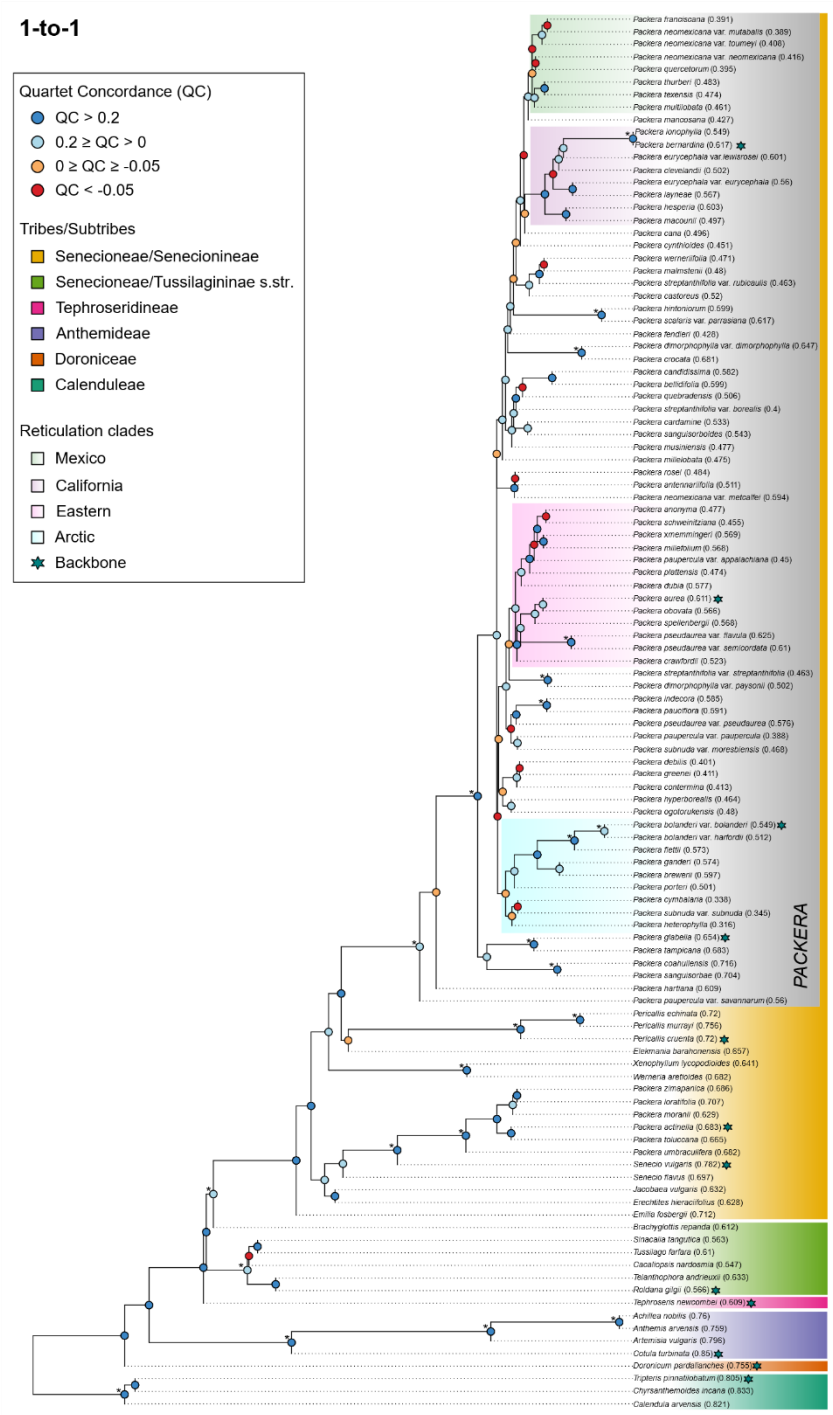


Figure 3. Full phylogeny of the 'one-to-one ortholog' (1-to-1) analysis. Tribal/subtribal status is indicated on the right of the phylogeny and follows the color code as indicated in the key. Numbers next to tip labels correspond to the Quartet Fidelity (QF) score from Quartet Sampling, which ranges from 0 to 1, with values closer to 0 indicating erroneous taxa. Circles at nodes show the Quartet Concordance (QC) factors from Quartet Sampling, where red and orange indicate counter-support nodes. Asterisks next to QC circles indicate high support (≥ 95 LPP) at that node. Highlighted clades and teal stars next to some tip labels, indicate clades where reticulation events were tested, following the color key on the left.

502

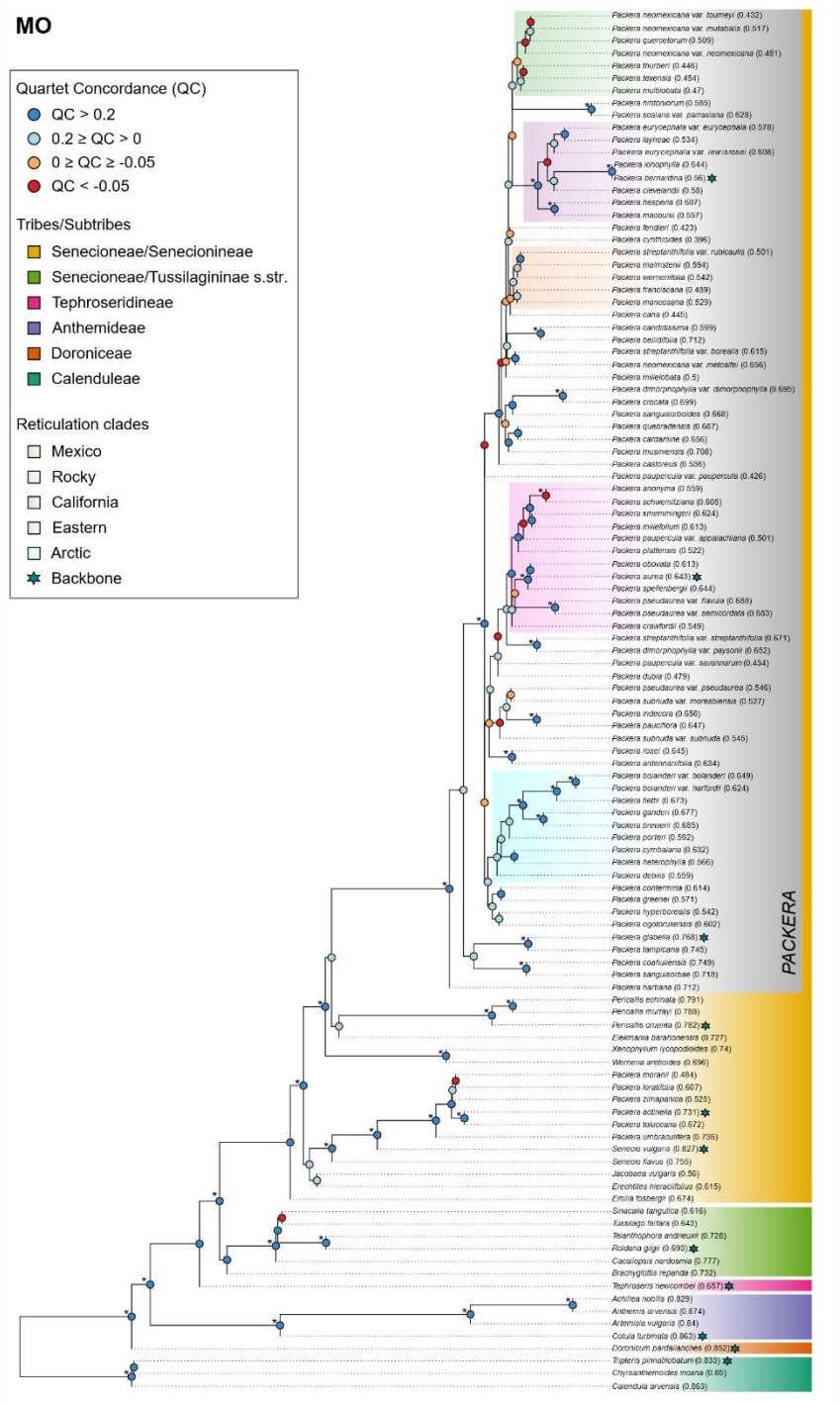


Figure 4. Full phylogeny of the 'monophyletic outgroup' (MO) analysis. Tribal/subtribal status is indicated on the right of the phylogeny and follows the color code as indicated in the key. Numbers next to tip labels correspond to the Quartet Fidelity (QF) score from Quartet Sampling, which ranges from 0 to 1, with values closer to 0 indicating erroneous taxa. Circles at nodes show the Quartet Concordance (QC) factors from Quartet Sampling, where red and orange indicate counter-support nodes. Asterisks next to QC circles indicate high support (≥ 95 LPP) at that node. Highlighted clades and teal stars next to some tip labels, indicate clades where reticulation events were tested, following the color key on the left.

503 out of 110 nodes (52.7%) having ≥ 0.95 LPP (Supplementary Table 3,
504 Supplementary Fig. 4). In contrast, the plastid and concatenated trees showed much
505 lower resolution with the fewest number of nodes having high support (36 of 110 nodes,
506 32.7%, have ≥ 95 BS; Supplementary Table 3, Supplementary Fig. 4).
507 Quartet Sampling indicated that underlying gene tree discordance remained high,
508 independent of the method used. ASTRAL and concatenated trees had the most fully
509 supported nodes (QS = 1/–/1) when compared to the remaining phylogenies
510 (Supplementary Fig. 5), though the number was still notably low (~10%). Even so, the
511 concatenated tree also had the highest proportion of counter-support nodes (QC < 0;
512 26.4%; see Supplementary Fig. 5 for all Quartet Sampling results). PhyParts indicated
513 that the ASTRAL and MO trees had the most concordant gene trees. The wASTRAL
514 analysis had the highest percentages of gene trees supporting the top alternative
515 bipartition. The MI tree harbored the highest discordance and uninformative data
516 (Supplementary Fig. 6).
517 This underlying discordance could be caused by ILS, which was common in all
518 tested ASTRAL trees (Supplementary Fig. 7). Introgressive hybridization (IH) was not
519 detected without the presence of ILS in any of the nodes, though IH was most common
520 in the nodes that were shown to be discordant (Supplemental Figs. 5, 7). Interestingly,
521 crown *Packera* consistently had high IH values, potentially indicating introgression at
522 that node. Backbone nodes were generally considered well resolved with high
523 confidence (ILS index < 50% and IH index < 10%; Shang et al. 2025), especially in the
524 outgroups, whereas *Packera* taxa appeared less resolved.

525 *Molecular Clock Dating*

526 The resulting divergence rate estimates differed among ASTRAL, 1-to-1, and
527 MO, with estimated ages of the MO and 1-to-1 trees older and sometimes almost
528 double the ASTRAL ages. For ASTRAL, the mean crown age for *Packera* ranged from
529 11.4 MY in Scenario 5 and 6 to 13.5 MY in Scenario 1-4. Alternatively, the mean crown
530 age of *Packera* were almost doubled in the 1-to-1 and MO analyses, with ranges from
531 ~21 MY in Scenario 5 to 35.5–35.7 MY in Scenario 4 (Supplementary Table 11,
532 Supplementary Fig. 8). Scenario 2 was chosen as the best-fitting dating model because
533 it incorporated all available primary calibration points and yielded divergence time
534 estimates that align well with known patterns of evolutionary history (Pelser et al. 2010;
535 Mandel et al. 2019), while minimizing reliance on secondary calibrations
536 (Supplementary Table 5) (Müller and Reisz 2005; Benton and Donoghue 2007; Parham
537 et al. 2012). The dated trees from all six scenarios can be found in Supplementary
538 Figure 8.

539 *Biogeography*

540 Evaluation of the six different model choices demonstrated that BAYAREALIKE
541 was the best fit given our data based on its lowest log likelihood (Ln L) and highest AIC
542 score, as well as the highest weighted AICc score for all three of our analyses
543 (Supplementary Table 12). Generally, models including jump speciation (+J) provided a
544 better fit than those without in the DEC and DIVALIKE models, and were the same or
545 very similar in the BAYAREALIKE models (Supplementary Table 12), indicating a
546 potential for jump dispersal or founder-event speciation in our data. Trees of all model
547 runs are in Supplementary Figure 9.

The historical biogeography analysis revealed distinct geographic patterns across the different orthology methods (Fig. 5, Supplementary Fig. 9). ASTRAL identified the ancestral range as concentrated in the North American Prairies Province (N), Rocky Mountain and Vancouverian Province (R), and Madrean (M) regions. In contrast, MO showed the ancestral regions being only the Rocky Mountain and Vancouverian Province (R), and Madrean (M) regions. 1-to-1 was intermediate with crown *Packera* being ancestral to the North American Prairies Province (N), Rocky Mountain and Vancouverian Province (R), and Madrean (M) regions, similar to ASTRAL; however, the remaining backbone was Rocky Mountain and Vancouverian Province (R), and Madrean (M) regions, similar to MO. Notably, all analyses showed one major expansion into the eastern United States–Appalachian Province (P) and Atlantic and Gulf Coast Plain Province (T)–represented by the Eastern clade (Fig. 5, Supplementary Fig. 9), with a few instances of individual introductions to the eastern United States. Many of these taxa originated initially in the North American Prairies Province (N) and/or Rocky Mountain and Vancouverian Province (R) ranges. ASTRAL and MO showed two early expansions into the Arctic Province (I) and Canadian Province (C) at the base of the phylogeny, while 1-to-1 showed at least three expansions north. 1-to-1 exhibited the most geographic inconsistency, followed by MO. Even so, geography proved to be a reliable indicator for larger clades across all analyses, although widespread species were scattered across the phylogeny and did not always align consistently within geographic clades (Figs. 2-4, Supplementary Fig. 9).

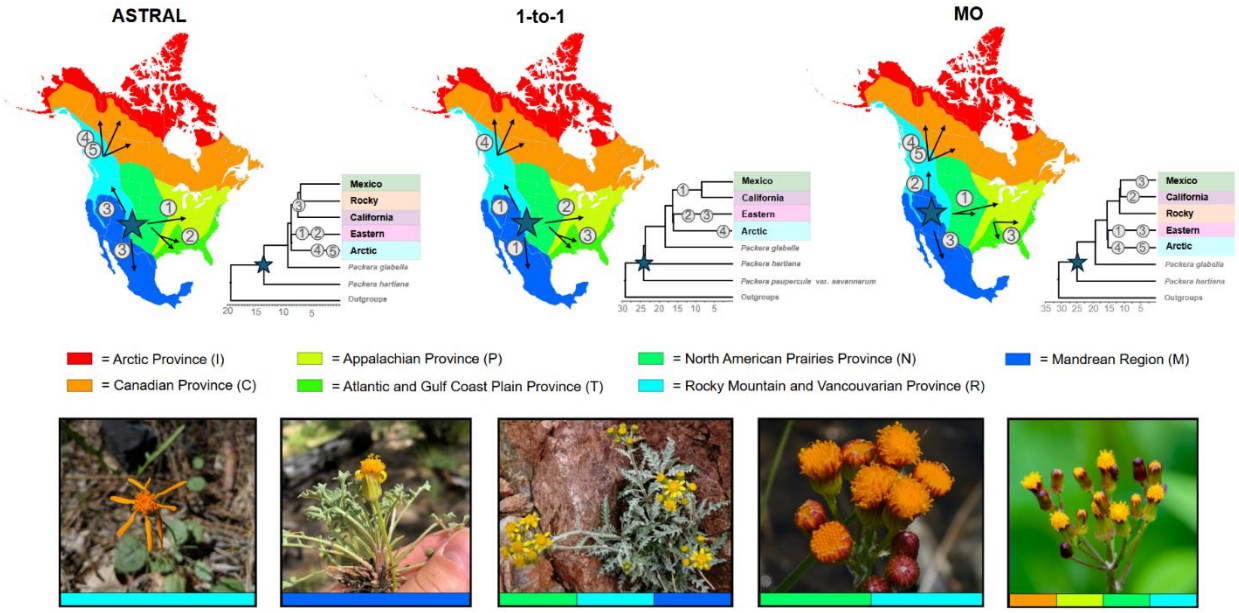


Figure 5. Predicted biogeographic region summaries generated from BioGeoBEARS of the ASTRAL, 1-to-1, and MO analyses. The maps of North America are colored according to the ten geographic regions outlined by Takhtajan (1986): Arctic Province (I), Canadian Province (C), Appalachian Province (P), Atlantic and Gulf Coast Plain Province (T), North American Prairies Province (N), Rocky Mountain and Vancouvarian Province (R), Madrean Region (M), Asia (A), Europe (E), and Central America (S). Stars indicate the predicted ancestral range of crown *Packera*. Arrows and numbers correspond to their predicted range expansions in chronological order, with 5 being the most recent major expansion. Numbers also correspond to the dated tree to the right of the phylogeny, which indicates the nodes the expansions occurred in relation to the major geographic clades designated in Figures 2-4. All BioGeoBears results can be found in Supplementary Figure 9. Images of *Packera* species highlighting some of the morphological diversity seen in the group. Colored bars on the bottom of the image correspond to their current distributions, following the color scheme of the map and key. From left to right: Flame ragwort, *P. greenei* (A.Gray) W.A.Weber & Á.Löve; Tehachapi ragwort, *P. ionophylla* (Greene) W.A.Weber & Á.Löve; Fendler's ragwort, *Packera fendleri* (A.Gray) W.A.Weber & Á.Löve; Weak groundsel, *P. debilis* (Nutt.) W.A.Weber & Á.Löve; and Rayless mountain ragwort, *P. indecora* (Greene) Á.Löve & D.Löve. Image credits: Erika R. Moore-Pollard and Robert Lagier.

Reticulation events

Five *Packera* clades were tested for ancient reticulation events given a large number of counter-support (QC < 0) nodes from Quartet Sampling on the ASTRAL, 1-to-1, and MO trees (Figs. 2-4). Additionally, a single species from each major clade was then used to represent backbone relationships throughout the entire phylogeny, totaling six separate PhyloNetwork runs named: Arctic, Backbone, California, Eastern, Mexico, and Rocky (Supplementary Table 13). Log-likelihood values indicated that the most

578 likely number of reticulation events (h) varied depending on the run (Supplementary Fig.
579 10). Additionally, Arctic, Mexico, and Backbone had varying numbers of optimal
580 reticulation events between the paralog-processing methods (Supplementary Table 7,
581 Supplementary Fig. 10). For ASTRAL, Arctic and Mexico had the highest number of
582 reticulation events (h = 3), followed by California and Backbone (h = 2), then Eastern
583 and Rocky (h = 1). For 1-to-1, Backhone had the most (h = 6), then Mexico (h = 3),
584 Arctic and California (h = 2), and then Eastern (h = 1). For MO, Arctic, California, and
585 Backbone had the highest number of reticulations (h = 2) with Eastern, Mexico, and
586 Rocky having the lowest (h = 1) (Fig. 6, Supplementary Fig. 10, Supplementary Table
587 7). The taxa also differed between analyses given different phylogenetic relationships
588 between clades (Supplementary Table 13), impacting some of the networks identified.
589 Some networks showed different events depending on the phylogenetic method
590 and subset of data (Fig. 6); however, some consistent patterns emerged. For example,
591 the Eastern clade continuously showed the same reticulation event regardless of the
592 paralog processing method, and parentage percentages were very similar. In contrast,
593 Rocky did not show the same reticulation events, but the networks generally involved
594 *Packera mancosana* and *P. streptanthifolia*. The California clade showed differing
595 results across methods: both ASTRAL and 1-to-1 analyses identified *P. bernardina* as
596 the product of reticulation with the same parentage, whereas MO did not yield the same
597 results, and no consistent trends emerged. The Mexico clade was generally
598 inconsistent, but featured reticulation events involving *P. neomexicana* and its varieties.
599 Additionally, *P. texensis* and *P. quercetorum* were involved in reticulation events in two
600 of the three paralog processing networks. Among all the clades, the Arctic clade

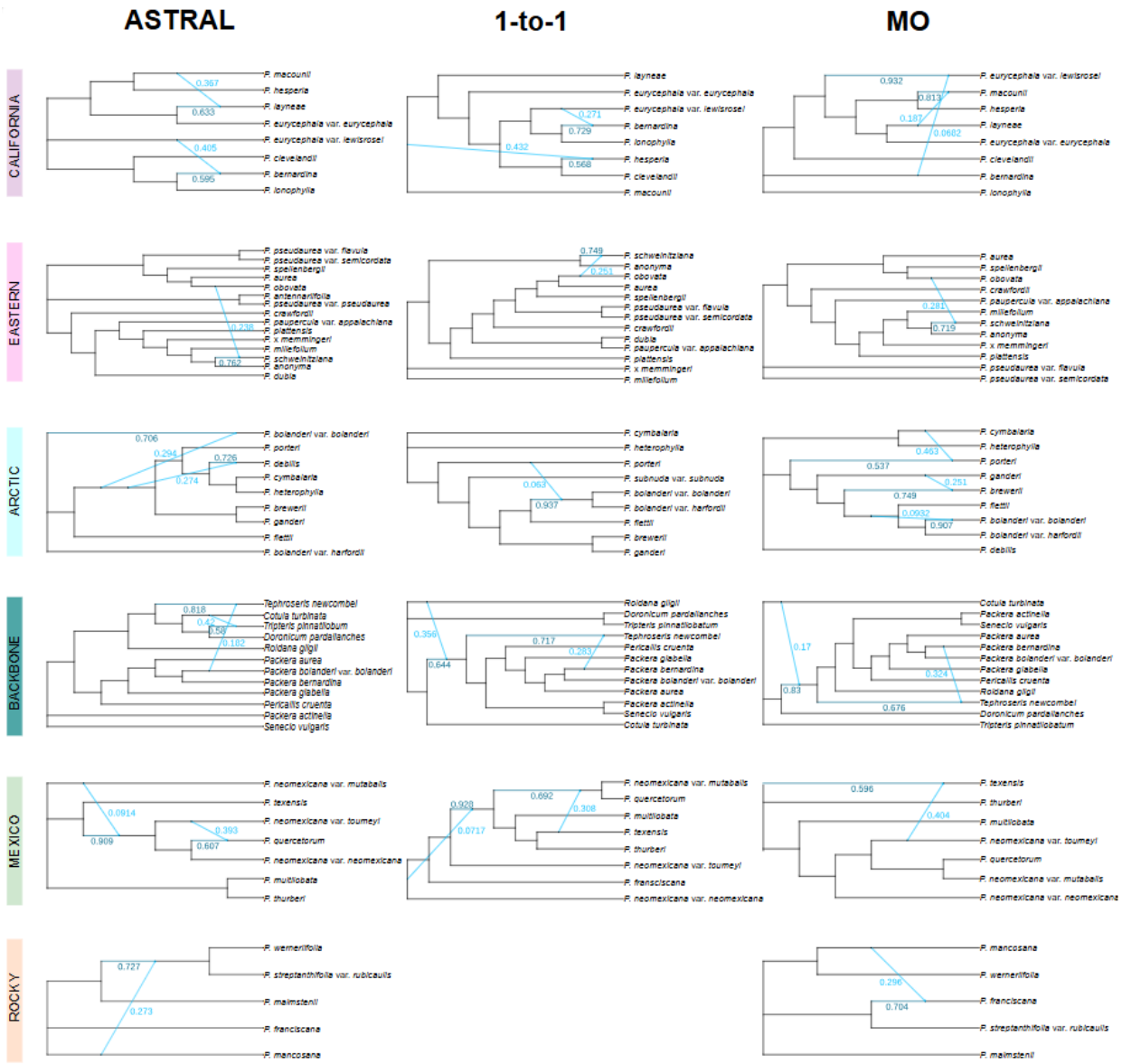


Figure 6. PhyloNetworks figures of the three tested analyses: ASTRAL, 1-to-1, and MO. The various areas of the phylogeny that were tested are shown on the left and can be read as a row. Dark green edges indicate the major hybrid edge, while the light blue shows the minor hybrid edge. The associated numbers next to each branch correspond to the percentage of parentage given by that branch. Number of hybridizations correspond with the best log likelihood value. The 1-to-1 analysis does not have a Rocky analysis since the clade in that tree only had four taxa.

exhibited the most variability in the number of reticulation events, with the *P. bolanderi* varieties consistently involved in some form of reticulation. However, the specifics of these events varied depending on sampling and paralog processing methods. In the Backbone subtree analysis, *Tephroseris* was frequently identified as a product of reticulation, with 1-to-1 and MO showing the same parentage. In contrast, ASTRAL did

606 not show the same reticulation events but did identify a reticulation event involving
607 *Tephroseris* with a different parentage. Notably, the parentage of *Tephroseris*
608 always included a *Packera* taxon, though the specific taxon varied.
609

610 **DISCUSSION**

611 Gene duplication was pervasive in *Packera*, with more than 80% of nuclear loci
612 showing evidence of paralogy (Supplementary Table 3). Such widespread paralogy
613 poses major challenges for species tree inference, especially in Hyb-Seq datasets,
614 where probes may unintentionally enrich for multiple homeologs (Salmon et al. 2012;
615 Akama et al. 2014; Crameri et al. 2022), complicating the recovery of true single-copy
616 orthologs for phylogenetic reconstructions. Although some paralogy is expected in
617 target enrichment studies (Siniscalchi et al. 2021), the extent observed in *Packera* is
618 unusually high and reflects lineage-specific genome dynamics, rather than a general
619 feature of all species. This has important methodological implications, as HybPiper's
620 orthology inference relies largely on single-copy filtering (Johnson et al. 2016), and
621 using its outputs alone can introduce noise and bias into gene tree reconstruction,
622 ultimately distorting species tree inference. Consistent with our study, using only
623 putative single-copy orthologs designated by HybPiper was insufficient for accurate
624 phylogenetic inference in *Packera*, a group characterized by extensive polyploidy and
625 hybridization (Fig. 2) (Trock 2006). These challenges are not unique to *Packera*, as
626 many plant and non-plant lineages also experience difficulties reconstructing
627 phylogenies due to hybridization and gene/genome duplications. Together, these

628 examples highlight the complexity of orthology inference in phylogenomics and the need
629 for rigorous paralog processing across both plant and non-plant systems.
630 The most recent phylogeny of *Packera* prior to this study was based on ITS data
631 and revealed extremely low sequence variation, with only 0–5% divergence and several
632 *Packera* taxa sharing identical sequences (Bain and Golden 2000). This led to a largely
633 unresolved tree, likely due to few phylogenetically informative sites, sparse species
634 sampling, and potential reticulation events (Funk 1985; Linder and Rieseberg 2004). By
635 using Hyb-Seq data and broader taxon sampling, we substantially improved
636 phylogenetic resolution across *Packera*, including within major clades. However, our
637 analyses also revealed that gene tree concordance and species tree resolution varied
638 substantially across methods (Supplementary Fig. 4), exemplifying the influence of
639 paralog processing strategies on downstream phylogenetic outcomes. Methods that
640 explicitly model gene duplication and loss, such as those implemented in paragone-nf
641 (Yang and Smith 2014; Jackson et al. 2023), generally yielded more concordant gene
642 trees and better-resolved species trees (Figs. 3, 4). However, substantial conflict
643 persisted across the trees, indicating that even sophisticated modeling approaches
644 cannot fully resolve deep relationships in groups with complex histories like *Packera*.
645 These discrepancies may in part reflect differences in the number and identity of genes
646 retained after filtering (Supplementary Table 3). For instance, the 1-to-1 method, which
647 excludes all loci with paralogs and resembles approaches used in tools like Phyluce,
648 eliminates a significant portion of the dataset and may oversimplify the data, especially
649 in duplication-rich systems such as *Packera*. While Phyluce and 1-to-1 filtering both
650 remove paralogs, the underlying methods differ (Faircloth 2016; Jackson et al. 2023),

651 causing Phyluce to outperform HybPiper in certain cases (Mandel et al. 2019; Thapa et
652 al. 2020). Similarly, the RT method naturally loses information since it removes
653 outgroups to determine orthology (Yang and Smith 2014; Morales-Briones et al. 2021b,
654 2021a) and potentially limits phylogenetic signals. These findings highlight that while
655 some approaches have potential to improve concordance, none fully resolved the
656 underlying discordance of *Packera*, emphasizing that the choice of method, underlying
657 data, and biological context must guide ortholog selection in phylogenomic analyses.
658 Results from a previous phylogenetic study of *Packera* showed that its
659 evolutionary relationships tend to loosely follow a geographic pattern (Bain and Golden
660 2000), which is consistent with our findings (Figs. 2-4, Supplementary Fig. 9). Our
661 nuclear phylogenies showed that major clades tended to contain taxa within similar
662 geographic ranges, independent of which paralog processing method was utilized.
663 Additionally, most taxa within these major clades were conserved across methods,
664 though their placement along the backbone, and some taxa, did vary (Figs. 2-4). For
665 example, the taxa in the Eastern clade was fairly consistent between the orthology
666 inference methods, although some taxa did move in and out of the clade
667 (Supplementary Fig. 4). Interestingly, several of these taxa are widespread and engage
668 in interspecific hybridization with other *Packera* species where their ranges overlap. For
669 example, *P. pseudaurea* and its varieties are typically found in central North America,
670 though their phylogenetic placement is generally in the Eastern clade. These results are
671 unsurprising as they have an extensive history of hybridization with other *Packera* taxa
672 in the Eastern clade, which is believed to have created much taxonomic uncertainty
673 (Barkley 1962; Trock 1999), potentially explaining their inconsistent phylogenetic

674 placements (Supplementary Fig. 4). However, conflicting signals may also reflect
675 additional evolutionary processes beyond hybridization, such as cryptic speciation
676 within widespread taxa or the influence of unsampled and/or extinct lineages.
677 Our results also reveal that evolutionary interpretations such as divergence
678 dating and biogeographic reconstruction are highly sensitive to the paralog processing
679 approach used to infer the underlying phylogeny. Divergence time estimates varied
680 across methods, with some species trees suggesting that crown *Packera*'s
681 diversification occurred more recently (ASTRAL: 13.5 MY), while others placed key
682 divergences deeper in time (1-to-1: 24.2 MY and MO: 25 MY; Supplementary Fig. 8,
683 Supplementary Table 11). These differences in divergence times are not trivial, as they
684 affect our understanding of how *Packera* taxa respond to climatic shifts, mountain uplift,
685 and other historical events, and have broader implications for comparative studies
686 across Asteraceae and other radiating plant lineages. For example, tribal- and family-
687 wide studies predicted the crown age of *Packera* to be fairly young, between 7.8 MY
688 and 20.2 MY (Pelser et al. 2010; Mandel et al. 2019; Zhang et al. 2021), which is more
689 in-line with the ASTRAL analysis. While more *Packera*-specific predictions are older,
690 with (Barkley 1988) predicting the origin of *Packera* in the later stages of the mid-
691 Tertiary (~22-37 MY), closely aligning with the paragone-nf analyses (Supplementary
692 Fig. 8). Similarly, biogeographic patterns shifted depending on the inferred topology,
693 with each analysis showing different migration patterns over time (Fig. 5,
694 Supplementary Fig. 9). Even so, the broad pattern is relatively unchanged. For
695 example, all topologies generally agreed that *Packera* originated in western North
696 America with one major expansion east and then later expansions north from the

western clades. From there, individual taxa expanded or became specialized to certain regions, with most of the diversity present in western North America. Because the topologies are consistent across methods, we can be more confident that they likely reflect the true evolutionary history of *Packera* instead of methodological artifacts.

Despite improvements from paralogy-aware filtering, the species trees for *Packera* were not fully resolved, indicating that underlying biological processes such as ILS, hybridization, and introgression also play significant roles in shaping *Packera*'s evolutionary history. This is supported by strong cytonuclear discordance, where plastid phylogenetic relationships vary significantly from the nuclear trees (Fig. 1, Supplementary Fig. 4), consistent with chloroplast capture or ancient hybridization events (Doyle 1992; Maddison 1997; Galtier and Daubin 2008; Soltis and Soltis 2009; Smith et al. 2015; Stull et al. 2020). Directly testing for gene flow revealed multiple introgression events, particularly along short internal branches susceptible to ILS (Supplementary Fig. 7). However, the specific reticulation patterns recovered varied depending on the paralog processing method used. These discrepancies highlight a critical point: gene selection and paralog processing directly shapes reticulation inference. Each gene has its own evolutionary trajectory, shaped by recombination and gene duplication and loss (Betrán 2015), and filtering strategies that exclude paralogs or collapse duplications can erase signals of past hybridization. As a consequence of different orthology and paralog processing methods, some reticulation events, such as those involving Eastern *Packera* lineages, were consistently recovered across datasets, while others appeared only in specific analyses.

719 Network analyses, such as PhyloNetworks, are inherently sensitive to sampling
720 bias (Solís-Lemus and Ané 2016). Even though we sampled 85 out of a targeted 91
721 *Packera* taxa (~93% coverage of the genus), missing taxa may have impacted our
722 results. Including the remaining and unsampled *Packera* taxa in future studies could
723 help increase our understanding of these relationships. Additionally, limiting the
724 analyses to specified clades can ignore other potential reticulation events occurring
725 across groups. We recognize this limitation; however, expanding sampling to more taxa
726 or loci is currently restricted by the intensive and often unfeasible computational
727 requirements of reticulation network analyses (see Hejase and Liu 2016). With recent
728 bioinformatic advances and increased interest in non-bifurcating trees, we anticipate
729 that less computationally intense and more efficient phylogenetic and network tools will
730 soon be developed, allowing researchers to test for reticulations across large
731 phylogenies.
732 Ghost introgressions, which are ancient introgression events that leave genetic
733 signatures of extinct species in present-day species (Ottenburghs 2020), also pose
734 challenges to inferences made by hybridization networks. Accounting for ghost
735 introgressions in macroevolutionary studies may help improve the accuracy of trait
736 evolution and biogeographic reconstructions, refine divergence time estimates, and
737 reveal hidden contributions of hybridization to diversification (Ottenburghs 2020; Tricou
738 et al. 2022). Network methods such as PhyloNetworks can detect ghost introgressions,
739 but these methods do not specify that the hybrid edge is a ghost introgression, so it is
740 up to the researcher to recognize its presence, which can be difficult especially without

741 background knowledge of the group (Solís-Lemus and Ané 2016; Solís-Lemus et al.
742 2017).

743 Our findings emphasize that accurately reconstructing reticulate evolutionary
744 histories requires embracing the independent histories of genes and carefully
745 considering how orthology is defined. In *Packera*, gene tree heterogeneity is expected
746 and evolutionarily informative rather than mere noise. Methods that oversimplify gene
747 relationships, such as pipelines restricted to single-copy orthologs, risk masking the
748 very patterns needed to understand complex evolutionary processes. By comparing
749 alternative strategies for handling paralogs and integrating gene flow analyses, we show
750 that approaches which explicitly model gene duplication and loss provide modest
751 improvements in resolution and concordance but are insufficient on their own to fully
752 resolve relationships. Direct tests for reticulation and ILS revealed the extent to which
753 *Packera*'s evolutionary history has been shaped by hybridization and gene flow,
754 underscoring the broader challenge of capturing complexity in phylogenomic datasets.
755 Within polyploid systems, the tradeoff between excluding paralogs to reduce conflict
756 and retaining them to preserve evolutionary signals becomes especially pronounced.

757 These results also highlight how paralog handling influences not only
758 phylogenetic reconstruction but also the biological inferences drawn from the inferred
759 relationships. Macroevolutionary analyses such as divergence dating, trait evolution,
760 and biogeography are highly sensitive to orthology inference strategies and are only as
761 reliable as the phylogenetic framework on which they are built (e.g., Magalhaes et al.
762 2021). Improving that framework will require both methodological innovation and deeper
763 biological insight. For example, targeted probe sets designed to enrich truly single-copy

764 loci (e.g., Compositae-ParaLoss-1272; Moore-Pollard et al. 2024) offer one path
765 forward. Yet entirely excluding paralogs may discard informative loci, while retaining
766 them without appropriate modeling risks misleading inference. Integrating
767 morphological, ecological, and geographic evidence is therefore essential for assessing
768 whether patterns inferred from large-scale datasets reflect genuine evolutionary history
769 or methodological artifacts.

770 Rather than treating paralogy as a technical obstacle to be circumvented in
771 phylogenetic studies, our findings underscore its potential to illuminate fundamental
772 evolutionary processes, including gene duplication, polyploidy, and reticulation, that
773 have shaped the diversification of lineages across the tree of life. While these
774 challenges are common to many plant and non-plant groups, our densely sampled
775 *Packera* dataset provides a uniquely powerful system for testing and improving
776 phylogenomic methods. As the field advances toward increasingly nuanced and
777 biologically realistic models of genome evolution, systems like *Packera* will be essential
778 for developing tools that can fully capture the intricacy of complex evolutionary histories.

ACKNOWLEDGMENTS

The authors thank Drs. Carol Siniscalchi and Matthew Pollard for their bioinformatic help, Drs. Randall Bayer, L. Ellie Buckland, Duane McKenna, Anna Pardo, and Linda Watson for reading and editing prior to submission, and Dr. John Bain, Dr. Ed Schilling, and Brandon Fuller for donating genetic material for this study. We would also like to thank the following herbaria for allowing us to visit and/or destructively sample leaf tissue: ALA, ARIZ, ASU, CAS, MEM, NY, OSC, TEX, UNC, UCR, WTU. Permits to collect in the San Bernardino National Forest and Shasta-Trinity National Forest in California were obtained prior to collection of specimens through the USDA Forest Service. We also thank the University of Memphis High-Performance Cluster (HPC) administrators, Eric Spangler and Kristian Skjervold, for their assistance with the HPC and willingness to help. This work was supported by the Center for Biodiversity Research at the University of Memphis.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Raw sequence data are available in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under BioProjects PRJNA907383, PRJNA978568, PRJNA540287, and PRJNA516161.

800 **REFERENCES**

801 Akama S., Shimizu-Inatsugi R., Shimizu K.K., Sese J. 2014. Genome-wide
802 quantification of homeolog expression ratio revealed nonstochastic gene regulation in
803 synthetic allopolyploid *Arabidopsis*. *Nucleic Acids Res.* 42:e46.

804 Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data.
805 Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

806 Bain J.F., Golden J.L. 2000. A phylogeny of *Packera* (Senecioneae; Asteraceae) based
807 on internal transcribed spacer region sequence data and a broad sampling of
808 outgroups. *Mol. Phylogenet. Evol*. 16:331–338.

809 Bankevich A., Nurk S., Antipov D., Gurevich A.A., Dvorkin M., Kulikov A.S., Lesin V.M.,
810 Nikolenko S.I., Pham S., Prjibelski A.D., Pyshkin A.V., Sirotkin A.V., Vyahhi N., Tesler
811 G., Alekseyev M.A., Pevzner P.A. 2012. SPAdes: A new genome assembly algorithm
812 and its applications to single-cell sequencing. *J. Comput. Biol*. 19:455–477.

813 Barker M.S., Li Z., Kidder T.I., Reardon C.R., Lai Z., Oliveira L.O., Scascitelli M.,
814 Rieseberg L.H. 2016. Most Compositae (Asteraceae) are descendants of a
815 paleohexaploid and all share a paleotetraploid ancestor with the Calyceraceae. *Am. J.*
816 *Bot*. 103:1203–1211.

817 Barkley T.M. 1962. A Revision of *Senecio aureus*. *Trans. Kans. Acad. Sci*. 65:318–364.

818 Barkley T.M. 1988. Variation among the Aureoid Senecios of North America: A
819 geohistorical interpretation. *Bot. Rev*. 54:82–106.

820 Benton M.J., Donoghue P.C.J. 2007. Paleontological evidence to date the tree of life.
821 *Mol. Biol. Evol*. 24:26–53.

822 Bentz P.C., Leebens-Mack J. 2024. Developing Asparagaceae1726: An Asparagaceae-
823 specific probe set targeting 1726 loci for Hyb-Seq and phylogenomics in the family.
824 *Appl. Plant Sci*. 12:e11597.

825 Betrán E. 2015. The “life histories” of genes. *J. Mol. Evol*. 80:186–188.

826 Bezanson J., Chen J., Chung B., Karpinski S., Shah V.B., Vitek J., Zoubritzky L. 2018.
827 Julia: dynamism and performance reconciled by design. *Proc. ACM Program. Lang*.
828 2:1–23.

829 Bolger A.M., Lohse M., Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina
830 sequence data. *Bioinformatics*. 30:2114–2120.

831 Braasch I., Postlethwait J.H. 2012. Polyploidy in Fish and the Teleost Genome
832 Duplication. In *Polyploidy and Genome Evolution*. Springer Berlin Heidelberg. p. 341–
833 383.

834 Brown J.W., Walker J.F., Smith S.A. 2017. Phyx: Phylogenetic tools for unix.
835 *Bioinformatics*. 33:1886–1888.

836 Crameri S., Fior S., Zoller S., Widmer A. 2022. A target capture approach for
837 phylogenomic analyses at multiple evolutionary timescales in rosewoods (*Dalbergia*
838 spp.) and the legume family (Fabaceae). *Mol. Ecol. Resour*. 22:3087–3105.

839 Danecek P., Bonfield J.K., Liddle J., Marshall J., Ohan V., Pollard M.O., Whitwham A.,
840 Keane T., McCarthy S.A., Davies R.M., Li H. 2021. Twelve years of SAMtools and
841 BCFtools. *Gigascience*. 10:giab008.

842 Degnan J.H., Rosenberg N.A. 2009. Gene tree discordance, phylogenetic inference and
843 the multispecies coalescent. *Trends Ecol. Evol*. 24:332–340.

844 Doyle J.J. 1992. Gene trees and species trees: Molecular systematics as one-character
845 taxonomy. *Syst. Bot*. 17:144.

846 Elven R., Murray D.F., Volodya Yu. Razzhivin B.A.Y. 2011. Annotated Checklist of the
847 Panarctic Flora (PAF) Vascular plants version 1.0. Available from
848 http://panarcticflora.org/results?biogeographic=&bioclimatic=&region=&name=packera+
849 heterophylla#paf-863101 .

850 Evans B.J., Kelley D.B., Tinsley R.C., Melnick D.J., Cannatella D.C. 2004. A
851 mitochondrial DNA phylogeny of African clawed frogs: Phylogeography and implications
852 for polyploid evolution. *Mol. Phylogenet. Evol*. 33:197–213.

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

855 Freeman C.C., Barkley T.M. 1995. A synopsis of the genus *Packera* (Asteraceae:
856 Senecioneae) in Mexico. *SIDA Contrib. Bot*. 16:699–709.

857 Fu Z.X., Jiao B.H., Nie B., Zhang G.J., Gao T.G., Chen Z.D., Lu A.M., Kong H.Z., Wang
858 X.Q., Wang Y.Z., Zhou S.L., Zhang S.Z., Wang X.M., Liu Z.J., Wang Q.F., Li J.H., Li
859 D.Z., Yi T.S., Hong M.A., Soltis D.E., Soltis P.S., Li J.H., Fu C.X., Liu Q.X. 2016. A
860 comprehensive generic-level phylogeny of the sunflower family: Implications for the
861 systematics of Chinese Asteraceae. *J. Syst. Evol*. 54:416–437.

862 Funk V.A. 1985. Phylogenetic Patterns and Hybridization. *Ann. Mo. Bot. Gard*. 72:681.

863 Gallardo M.H., Kausel G., Jiménez A., Bacquet C., González C., Figueroa J., Köhler N.,
864 Ojeda, R. 2004. Whole-genome duplications in South American desert rodents
865 (Octodontidae). *Biol. J. Linn. Soc*. 82:443–451.

866 Galtier N., Daubin V. 2008. Dealing with incongruence in phylogenomic analyses.
867 *Philos. Trans. R. Soc. Lond. B Biol. Sci*. 363:4023–4029.

868 Geldmacher J., Van Den Bogaard P., Hoernle K., Schmincke H.U. 2000. The 40Ar/39Ar
869 age dating of the Madeira Archipelago and hotspot track (eastern North Atlantic).
870 *Geochem. Geophys. Geosyst*. 1.

871 Gramling A. 2006. A conservation assessment of *Packera millefolium*, a Southern
872 Appalachian Endemic [Dissertation]. The University of North Carolina at Chapel Hill.

https://www.proquest.com/dissertations-theses/conservation-assessment-i-packera-millefolium/docview/305279534/se-2?accountid=30699.

Grant V. 1971. Plant Speciation. New York: Columbia University Press.

Hejase H.A., Liu K.J. 2016. A scalability study of phylogenetic network inference methods using empirical datasets and simulations involving a single reticulation. *BMC Bioinf*. 17:422.

Hobbs C.R., Baldwin B.G. 2013. Asian origin and upslope migration of Hawaiian *Artemisia* (Compositae-Anthemideae). *J. Biogeogr*. 40:442–454.

Huang C.H., Zhang C., Liu M., Hu Y., Gao T., Qi J., Ma H. 2016. Multiple Polyploidization Events across Asteraceae with Two Nested Events in the Early History Revealed by Nuclear Phylogenomics. *Mol. Biol. Evol*. 33:2820–2835.

Jackson C., McLay T., Schmidt-Lebuhn A.N. 2023. hybpiper-nf and paragone-nf: Containerization and additional options for target capture assembly and paralog resolution. *Appl. Plant Sci*. 11:e11532.

Jin J.-J., Yu W.-B., Yang J.-B., Song Y., dePamphilis C.W., Yi T.-S., Li D.-Z. 2020. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol*. 21:241.

Johnson M.G., Gardner E.M., Liu Y., Medina R., Goffinet B., Shaw A.J., Zerega N.J.C., Wickett N.J. 2016. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Appl. Plant Sci*. 4:1600016.

Jones K.E., Alfredo Reyes-Betancort J., Hiscock S.J., Carine M.A. 2014. Allopatric diversification, multiple habitat shifts, and hybridization in the evolution of *Pericallis* (Asteraceae), a Macaronesian endemic genus. *Am. J. Bot*. 101:637–651.

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

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

Kumar S., Stecher G., Li M., Knyaz C., Tamura K. 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol*. 35:1547–1549.

Kumar S., Stecher G., Peterson D., Tamura K. 2012. MEGA-CC: Computing core of molecular evolutionary genetics analysis program for automated and iterative data analysis. *Bioinformatics*. 28:2685–2686.

Kumar S., Stecher G., Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol*. 33:1870–1874.

909 Landis M.J., Matzke N.J., Moore B.R., Huelsenbeck J.P. 2013. Bayesian analysis of
910 biogeography when the number of areas is large. *Syst. Biol*. 62:789–804.

911 Lanfear R., Calcott B., Ho S.Y.W., Guindon S. 2012. PartitionFinder: Combined
912 selection of partitioning schemes and substitution models for phylogenetic analyses.
913 *Mol. Biol. Evol*. 29:1695–1701.

914 Langmead B., Salzberg S.L. 2012. Fast gapped-read alignment with Bowtie 2. *Nat.*
915 *Methods*. 9:357–359.

916 Li H., Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler
917 transform. *Bioinformatics*. 25:1754–1760.

918 Linder C.R., Rieseberg L.H. 2004. Reconstructing patterns of reticulate evolution in
919 plants. *Am. J. Bot*. 91:1700–1708.

920 Maddison W.P. 1997. Gene trees in species trees. *Syst. Biol*. 46:523–536.

921 Magalhaes I.L.F., Santos A.J., Ramírez M.J. 2021. Incorporating Topological and Age
922 Uncertainty into Event-Based Biogeography of Sand Spiders Supports Paleo-Islands in
923 Galapagos and Ancient Connections among Neotropical Dry Forests. *Diversity*. 13:418.

924 Mahoney A.M., Kowal R.R. 2008. Three new varieties of *Packera paupercula*
925 (Asteraceae, Senecioneae) in midwestern and southeastern North America. *Novon: J.*
926 *Bot. Nomencl*. 18:220–228.

927 Mandel J.R., Dikow R.B., Funk V.A., Masalia R.R., Staton S.E., Kozik A., Michelmore
928 R.W., Rieseberg L.H., Burke J.M. 2014. A Target Enrichment Method for Gathering
929 Phylogenetic Information from Hundreds of Loci: An Example from the Compositae.
930 *Appl. Plant Sci*. 2:1300085.

931 Mandel J.R., Dikow R.B., Siniscalchi C.M., Thapa R., Watson L.E., Funk V.A. 2019. A
932 fully resolved backbone phylogeny reveals numerous dispersals and explosive
933 diversifications throughout the history of Asteraceae. *Proc. Natl. Acad. Sci. U.S.A.*
934 116:14083–14088.

935 Masterson J. 1994. Stomatal size in fossil plants: evidence for polyploidy in majority of
936 angiosperms. *Science*. 264:421–424.

937 Matzke N.J. 2014. Model selection in historical biogeography reveals that founder-event
938 speciation is a crucial process in island clades. *Syst. Biol*. 63:951–970.

939 McDougall I., Schmincke H.U. 1976. Geochronology of Gran Canaria, Canary Islands:
940 Age of shield building volcanism and other magmatic phases. *Bulletin Volcanologique*.
941 40:57–77.

942 Mello B. 2018. Estimating timetrees with MEGA and the timetree resource. *Mol. Biol.*
943 *Evol*. 35:2334–2342.

944 Mendes F.K., Hahn Y., Hahn M.W. 2016. Gene tree discordance can generate patterns

of diminishing convergence over time. *Mol. Biol. Evol*. 33:3299–3307.

Minh B.Q., Schmidt H.A., Chernomor O., Schrempf D., Woodhams M.D., Von Haeseler A., Lanfear R., Teeling E. 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol*. 37:1530–1534.

Mitchell N., Lewis P.O., Lemmon E.M., Lemmon A.R., Holsinger K.E. 2017. Anchored phylogenomics improves the resolution of evolutionary relationships in the rapid radiation of *Protea* L. *Am. J. Bot*. 104:102–115.

Moore-Pollard E.R., Jones D.S., Mandel J.R. 2024. Compositae-ParaLoss-1272: A complementary sunflower-specific probe set reduces paralogs in phylogenomic analyses of complex systems. *Appl. Plant Sci*. 12:e11568.

Morales-Briones D.F., Gehrke B., Huang C.-H., Liston A., Ma H., Marx H.E., Tank D.C., Yang Y. 2021a. Analysis of paralogs in target enrichment data pinpoints multiple ancient polyploidy events in *Alchemilla* s.L. (Rosaceae). *Syst. Biol*. 71:190–207.

Morales-Briones D.F., Kadereit G., Tefarikis D.T., Moore M.J., Smith S.A., Brockington S.F., Timoneda A., Yim W.C., Cushman J.C., Yang Y. 2021b. Disentangling sources of gene tree discordance in phylogenomic data sets: Testing ancient hybridizations in Amaranthaceae s.l. *Syst. Biol*. 70:219–235.

Morales-Briones D.F., Liston A., Tank D.C. 2018. Phylogenomic analyses reveal a deep history of hybridization and polyploidy in the Neotropical genus *Lachemilla* (Rosaceae). *New Phytol*. 218:1668–1684.

Müller J., Reisz R.R. 2005. Four well-constrained calibration points from the vertebrate fossil record for molecular clock estimates. *BioEssays*. 27:1069–1075.

Nice C.C., Gompert Z., Fordyce J.A., Forister M.L., Lucas L.K., Buerkle C.A. 2013. Hybrid speciation and independent evolution in lineages of alpine butterflies. *Evolution*. 67:1055–1068.

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

Ottenburghs J. 2020. Ghost introgression: Spooky gene flow in the distant past. *BioEssays*. 42:e2000012.

Palazzesi L., Pellicer J., Barreda V.D., Loeuille B., Mandel J.R., Pokorny L., Siniscalchi C.M., Tellería M.C., Leitch I.J., Hidalgo O. 2022. Asteraceae as a model system for evolutionary studies: from fossils to genomes. *Bot. J. Linn. Soc*. 200:143–164.

Panchy N., Lehti-Shiu M., Shiu S.H. 2016. Evolution of gene duplication in plants. *Plant Physiol*. 171:2294–2316.

Panero J.L. 2005. New combinations and infrafamilial taxa in the Asteraceae. *Phytologia*. 87:1–14.

981 Paradis E., Schliep K. 2019. Ape 5.0: An environment for modern phylogenetics and
982 evolutionary analyses in R. *Bioinformatics*. 35:526–528.

983 Parham J.F., Donoghue P.C.J., Bell C.J., Calway T.D., Head J.J., Holroyd P.A., Inoue
984 J.G., Irmis R.B., Joyce W.G., Ksepka D.T., Patané J.S.L., Smith N.D., Tarver J.E., Van
985 Tuinen M., Yang Z., Angielczyk K.D., Greenwood J.M., Hipsley C.A., Jacobs L.,
986 Makovicky P.J., Müller J., Smith K.T., Theodor J.M., Warnock R.C.M., Benton M.J.
987 2012. Best practices for justifying fossil calibrations. *Syst. Biol*. 61:346–359.

988 Pease J.B., Brown J.W., Walker J.F., Hinchliff C.E., Smith S.A. 2018. Quartet Sampling
989 distinguishes lack of support from conflicting support in the green plant tree of life. *Am.*
990 *J. Bot*. 105:385–403.

991 Peirson J.A., Reznicek A.A., Semple J.C. 2012. Polyploidy, infraspecific cytotype
992 variation, and speciation in Goldenrods: The cytogeography of *Solidago* subsect.
993 *Humiles* (Asteraceae) in North America. *Taxon*. 61:197–210.

994 Pelser P.B., Kennedy A.H., Tepe E.J., Shidler J.B., Nordenstam B., Kadereit J.W.,
995 Watson L.E. 2010. Patterns and causes of incongruence between plastid and nuclear
996 Senecioneae (Asteraceae) phylogenies. *Am. J. Bot*. 97:856–873.

997 R Core Team. 2016. R: A Language and Environment for Statistical Computing. Vienna:
998 R Foundation for Statistical Computing.

999 Ree R.H., Smith S.A. 2008. Maximum likelihood inference of geographic range
1000 evolution by dispersal, local extinction, and cladogenesis. *Syst. Biol*. 57:4–14.

1001 Revell L.J. 2012. phytools: An R package for phylogenetic comparative biology (and
1002 other things). *Methods Ecol. Evol*. 3:217–223.

1003 Ronquist F. 1997. Dispersal-vicariance analysis: A new approach to the quantification of
1004 historical biogeography. *Syst. Biol*. 46:195–203.

1005 Rose J.P., Toledo C.A.P., Moriarty Lemmon E., Lemmon A.R., Sytsma K.J. 2021. Out
1006 of Sight, Out of Mind: Widespread Nuclear and Plastid-Nuclear Discordance in the
1007 Flowering Plant Genus *Polemonium* (Polemoniaceae) Suggests Widespread Historical
1008 Gene Flow Despite Limited Nuclear Signal. *Syst. Biol*. 70:162–180.

1009 RStudio. 2020. RStudio: Integrated Development for R. Boston, MA: RStudio, PBC.

1010 Salmon A., Udall J.A., Jeddeloh J.A., Wendel J. 2012. Targeted capture of
1011 homoeologous coding and noncoding sequence in polyploid cotton. *G3*. 2:921–930.

1012 Sayyari E., Mirarab S. 2016. Fast Coalescent-Based Computation of Local Branch
1013 Support from Quartet Frequencies. *Mol. Biol. Evol*. 33:1654–1668.

1014 Schliep K.P. 2011. phangorn: phylogenetic analysis in R. *Bioinformatics*. 27:592–593.

1015 Shang H.-Y., Jia K.-H., Li N.-W., Zhou M.-J., Yang H., Tian X.-L., Ma Y.-P., Zhang R.-G.
1016 2025. Phytop: a tool for visualizing and recognizing signals of incomplete lineage sorting

1017 and hybridization using species trees output from ASTRAL. *Hortic. Res*. 12:uhae330.

1018 Siniscalchi C.M., Hidalgo O., Palazzesi L., Pellicer J., Pokorny L., Maurin O., Leitch I.J.,
1019 Forest, F., Baker W.J. Mandel J.R. 2021. Lineage-specific vs. universal: A comparison
1020 of the Compositae1061 and Angiosperms353 enrichment panels in the sunflower family.
1021 *Appl. Plant Sci*. 9.

1022 Smith S.A., Moore M.J., Brown J.W., Yang Y. 2015. Analysis of phylogenomic datasets
1023 reveals conflict, concordance, and gene duplications with examples from animals and
1024 plants. *BMC Evol. Biol*. 15:150.

1025 Solís-Lemus C., Ané C. 2016. Inferring Phylogenetic Networks with Maximum
1026 Pseudolikelihood under Incomplete Lineage Sorting. *PLoS Genet*. 12:1–21.

1027 Solís-Lemus C., Bastide P., Ané C. 2017. PhyloNetworks: A package for phylogenetic
1028 networks. *Mol. Biol. Evol*. 34:3292–3298.

1029 Soltis D.E., Soltis P.S. 1999. Polyploidy: Recurrent formation and genome evolution.
1030 *Trends Ecol. Evol*. 14:348–352.

1031 Soltis P.S., Soltis D.E. 2009. The role of hybridization in plant speciation. *Annu. Rev.*
1032 *Plant Biol*. 60:561–588.

1033 Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-
1034 analysis of large phylogenies. *Bioinformatics*. 30:1312–1313.

1035 Steel M.A., Penny D. 1993. Distributions of tree comparison metrics—some new results.
1036 *Syst. Biol*. 42:126–141.

1037 Steenwyk J.L., Buida T.J., Labella A.L., Li Y., Shen X.-X., Rokas A. 2021. PhyKIT: a
1038 broadly applicable UNIX shell toolkit for processing and analyzing phylogenomic data.
1039 *Bioinformatics*. 37:2325–2331.

1040 Stull G.W., Soltis P.S., Soltis D.E., Gitzendanner M.A., Smith S.A. 2020. Nuclear
1041 phylogenomic analyses of asterids conflict with plastome trees and support novel
1042 relationships among major lineages. *Am. J. Bot*. 107:790–805.

1043 Takhtajan A. 1986. Floristic Regions of the World. Berkley and Los Angeles, California:
1044 University of California Press.

1045 Tamura K., Battistuzzi F.U., Billing-Ross P., Murillo O., Filipski A., Kumar S. 2012.
1046 Estimating divergence times in large molecular phylogenies. *Proc. Natl. Acad. Sci.*
1047 *U.S.A.* 109:19333–19338.

1048 Tamura K., Tao Q., Kumar S. 2018. Theoretical foundation of the RelTime method for
1049 estimating divergence times from variable evolutionary rates. *Mol. Biol. Evol*. 35:1770–
1050 1782.

1051 Thapa R., Bayer R.J., Mandel J.R. 2020. Phylogenomics Resolves the Relationships
1052 within *Antennaria* (Asteraceae, Gnaphalieae) and Yields New Insights into its

1053 Morphological Character Evolution and Biogeography. *Syst. Bot*. 45:387–402.

1054 Tillich M., Lehwark P., Pellizzer T., Ulbricht-Jones E.S., Fischer A., Bock R., Greiner S.
1055 2017. GeSeq–versatile and accurate annotation of organelle genomes. *Nucleic Acids*
1056 *Res*. 45:W6–W11.

1057 Tricou T., Tannier E., de Vienne D.M. 2022. Ghost lineages highly influence the
1058 interpretation of introgression tests. *Syst. Biol*. 71:1147–1158.

1059 Trock D.K. 1999. A revisionary synthesis of the genus *Packera* (Asteraceae:
1060 Senecioneae) [Dissertation]. Kansas State University.
1061 https://www.proquest.com/dissertations-theses/revisionary-synthesis-genus-packera-
1062 asteraceae/docview/304521955/se-2?accountid=30699.

1063 Trock D.K. 2006. *Packera* Á. Löve & D. Löve. Flora of North America Editorial
1064 Committee, eds. 1993+. New York and Oxford: p. 570–602.

1065 Vision T.J., Brown D.G., Tanksley S.D. 2000. The origins of genomic duplications in
1066 *Arabidopsis*. *Science*. 290:2114–2117.

1067 Wang W.M. 2004. On the origin and development of *Artemisia* (Asteraceae) in the
1068 geological past. *Bot. J. Linn. Soc*. 145:331–336.

1069 Wendel J.F., Cronn R.C. 2003. Polyploidy and the evolutionary history of cotton. *Adv.*
1070 *Agron*. 23:139–186.

1071 Yang Y., Smith S.A. 2014. Orthology inference in nonmodel organisms using
1072 transcriptomes and low-coverage genomes: Improving accuracy and matrix occupancy
1073 for phylogenomics. *Mol. Biol. Evol*. 31:3081–3092.

1074 Yeatts L., Schneider A., Schneider B. 2011. *Packera mancosana* (Asteraceae:
1075 Senecioneae), a new species and shale barren endemic of southwestern Colorado.
1076 *Phytoneuron*. 26:1–8.

1077 Yu G., Smith D.K., Zhu H., Guan Y., Lam T.T.-Y. 2017. GGTREE: An R package for
1078 visualization and annotation of phylogenetic trees with their covariates and other
1079 associated data. *Methods Ecol. Evol*. 8:28–36.

1080 Zhang C., Huang C.-H., Liu M., Hu Y., Panero J.L., Luebert F., Gao T., Ma H. 2021.
1081 Phylotranscriptomic insight into Asteraceae diversity, polyploidy, and morphological
1082 innovation. *J. Integr. Plant Biol*. 63:1273–1293.

1083 Zhang C., Mirarab S. 2022a. ASTRAL-Pro 2: ultrafast species tree reconstruction from
1084 multi-copy gene family trees. *Bioinformatics*. 38:4949–4950.

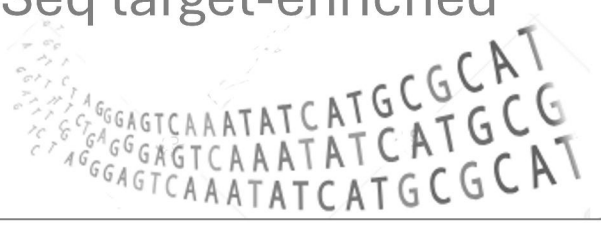
1085 Zhang C., Mirarab S. 2022b. Weighting by gene tree uncertainty improves accuracy of
1086 quartet-based species trees. *Mol. Biol. Evol*. 39:msac215.

1087 Zhang C., Rabiee M., Sayyari E., Mirarab S. 2018. ASTRAL-III: Polynomial time species
1088 tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*. 19:15–30.

1089 Zhang C., Scornavacca C., Molloy E.K., Mirarab S. 2020. ASTRAL-Pro: Quartet-based
1090 species-tree inference despite paralogy. *Mol. Biol. Evol*. 37:3292–3307.

Step 1

Trimmomatic (Hyb-Seq target-enriched paired end reads)



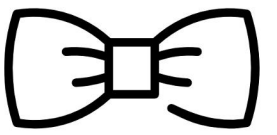
NUCLEAR

PLASTID

Step 2

HybPiper (mapped to Comp-1061 genes)

Bowtie2 (mapped to reference genome)



Step 3

ASTRAL
Orthologs are chosen based on coverage depth or sequence similarity

wASTRAL
Similar to ASTRAL, but low support branches are contracted

ASTRAL-Pro
All single-copy and multi-copy genes are considered

1-to-1
Only retains homologs without duplications

MI
Retains subtrees with the highest number of taxa without duplications

MO
Only retains homologs with monophyletic outgroups

RT
Prune by extracting ingroup clades, then cut paralogs from root to tip

ASTRAL

RAxML

Concatenated
Genes are not given individual evolutionary histories

Concatenated
Genes are not given individual evolutionary histories