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

Erika R. Moore-Pollard[1,2,4], 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*

4 *Address for correspondence*

Emails and ORCID IDs of all authors:

Erika R. Moore-Pollard, moore.erika.r@gmail.com, 0000-0002-1182-9274

Paige A. Ellestad, pllestad@memphis.edu, 0000-0003-3930-3239

Jennifer R. Mandel, jmandel@memphis.edu, 0000-0003-3539-2991

Abstract

Phylogenomic discordance is pervasive and cannot be fully addressed through increased 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 the impact of different published orthology inference methods on the resulting evolutionary relationships and phylogenetic support of this group. We then applied three paralog-processing 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. Unsurprisingly, phylogenetic outcomes varied depending on the paralog processing method used, with no method performing the best overall over others. Our findings highlight the profound 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 and 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 species

boundaries, cytological variation, and ecological interactions, 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; paralogy; molecular systematics

Introduction

The field of phylogenetics is drastically changing with the increased use of 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. A significant challenge is gene tree discordance, where 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, they are by no means exclusive to them. Similar complications have been documented in non-plants, such as African clawed frogs (Evans et al. 2004), desert rodents (Gallardo et al. 2004), alpine butterflies (Nice et al. 2013), and fish (Braasch and Postlethwait 2012), indicating that complex histories of gene duplication and horizontal gene transfer create

76 phylogenetic uncertainty across the tree of life. As the scale and scope of phylogenomic
77 datasets expand, so too does the need to carefully model and interpret the biological
78 processes underlying gene tree discordance. Fortunately, the growing availability of
79 phylogenomic data now makes it possible to start untangling these complex histories in
80 ways that were not previously feasible; however, doing so requires tools and models
81 that explicitly account for these processes.

82 In response, a number of bioinformatic tools and pipelines have been developed
83 to infer orthology under increasingly realistic evolutionary models. For instance,
84 softwares such as HybPiper (Johnson et al. 2016), Phyluce (Faircloth 2016), and
85 Captus (Ortiz et al. 2023) implement distinct strategies for identifying and handling
86 putative paralogs, with some pipelines retaining only a single gene copy (e.g., HybPiper
87 and Captus), while others aim to exclude all loci showing any sign of paralogy (e.g.,
88 Phyluce). Although Phyluce has shown greater success in some groups (e.g.,
89 *Antennaria* Gaertn.; Thapa et al. 2020), HybPiper typically outperforms across most
90 plant lineages because of the issues associated with polyploidy. Time will tell how
91 newer tools such as Captus perform compared to Phyluce or HybPiper.

92 Additional pipelines have been developed that implement distinct strategies for
93 identifying and handling putative paralogs using graph-based or tree-based methods
94 (e.g., paragone-nf; Yang and Smith 2014; Jackson et al. 2023), whereas others
95 accommodate paralogy during species tree estimation itself (e.g., ASTRAL-Pro; Zhang
96 et al. 2020; Zhang and Mirarab 2022a). These different approaches can yield markedly
97 different phylogenetic outcomes, particularly in polyploid-rich systems where paralogs
98 may be the rule rather than the exception. Yet, few studies have systematically

99 compared these methods to understand their impact on phylogenetic and downstream
100 evolutionary inferences.

101 This is especially important for groups such as *Packera* Á. Löve & D. Löve
102 (Asteraceae: Senecioneae), a North American genus of about 91 species known for
103 complex patterns of hybridization, polyploidy, and poorly resolved taxonomy. An
104 estimated 40% of *Packera* taxa present polyploidy, aneuploidy, and other cytological
105 disturbances (Barkley 1988; Trock 2006), complicating phylogenetic reconstruction of
106 this group (Bain and Golden 2000). Given this, *Packera* presents an ideal system to test
107 the effects of orthology inference differences given its extensive polyploidy, ecological
108 and geographic diversity, and history of reticulation—all factors that increase the
109 potential for gene tree discordance.

110 Our study provides a broadly applicable framework for evaluating orthology
111 inference and its downstream impacts in systems shaped by gene duplication,
112 reticulation, and rapid diversification. Here we present the most comprehensive
113 phylogenomic investigation of *Packera* to date, sampling nearly 93% of the genus and
114 leveraging a high-resolution Hyb-Seq dataset. We applied a robust framework for
115 evaluating the impact of paralog processing by comparing multiple orthology inference
116 methods, including pipelines that remove all paralogs, those that model duplication and
117 loss, and those that retain a single representative gene copy. We assessed how these
118 approaches influence gene tree concordance, species tree topology, and node support.
119 We then tested how three of these methods–a standard orthology inference method, a
120 method that removes all paralogous loci, and a selective orthology method (see
121 methods below)–alter downstream macroevolutionary analyses including divergence
122 dating, historical biogeography, and detection of ancient introgression. Our results

demonstrate that the choice of paralog processing approach strongly influences inferred species relationships, gene tree concordance, node support, and evolutionary inferences, highlighting the profound effect that orthology inference decisions can have in phylogenomics.

Materials & Methods

Specimen collection

A nearly complete sampling of currently recognized *Packera* species and varieties were used in this study (62 *Packera* species, 22 varieties, and one nothospecies; totaling 85 *Packera* taxa). *Packera* species were selected based on information published in the International Plant Names Index (IPNI; available at http://www.ipni.org) and then curated from discussions with botanists that have previously specialized in *Packera* (D. Trock & J. Bain, pers. comm.). Outgroup taxa from tribes Senecioneae (n = 19) and Anthemideae (n = 4) were selected by consulting recent studies of the group (Bain and Golden 2000; Pelser et al. 2010; Mandel et al. 2019). Additionally, three taxa from tribe Calenduleae were included for downstream analyses that require more distantly related outgroups that do not have a known WGD event (Barker et al. 2016; Huang et al. 2016; Zhang et al. 2021). A complete list of the 111 sampled species, herbarium vouchers, and their publication status can be found in Table S1.

DNA extraction and sequencing

Leaf tissue was obtained from herbarium sheets or live plants that were dried in silica beads. Dried leaf tissue (~25 mg) was ground using a Bead Mill 24 Homogenizer (ThermoFisher Scientific, Atlanta, Georgia, USA), and total DNA was extracted using

147 the Omega Bio-Tek Kit (Atlanta, Georgia, USA) following manufacturer's protocols with
148 the addition of polyvinylpyrrolidone (PVP) and ascorbic acid to the first extraction buffer
149 (10 mL SQ1 buffer, 100 mg PVP, 90 mg ascorbic acid). Samples of *Packera* and
150 outgroup species that were obtained from other researchers followed different DNA
151 extraction protocols which are outlined in Table S1. All DNA samples used in this study
152 were quality checked using a Nanodrop 2000 (Thermo Fisher Scientific, Carlsbad,
153 California, USA) and quantified using a Qubit High Sensitivity assay (ThermoFisher
154 Scientific, Oregon, USA). If the DNA had low concentrations or possible contaminants,
155 the samples were cleaned following the E.Z.N.A. Cycle Pure Kit (Omega Biotek,
156 Georgia, USA) and were re-quantified with the Qubit HS assay. Up to 1 µg DNA per
157 sample was sonicated with a QSonica machine (ThermoCube, New York, USA) to
158 generate fragment sizes within the 400-500 base pair (bp) range. The amount of DNA
159 used is dependent on the initial total DNA concentration of the targeted sample. Time
160 allotted for sonication of each sample varied depending on approximate length and
161 degree of fragmentation of the DNA. Prior to sonication, total DNA was visualized on a
162 1% agarose gel in 1X TBE and GelRed 3x (Biotium) and the approximate length was
163 estimated by comparing to a 100 bp GeneRuler DNA ladder (ThermoFisher Scientific,
164 Vilnius, Lithuania). If the sample of DNA appeared smaller than 400 bp, the samples
165 were not sonicated.

166 Libraries for each sample were prepared using the NEBNext Ultra II DNA Library
167 Prep Kit and NEBNext Muliplex Oligos for Illumina Kits (New England Biolabs, Ipswich,
168 Massachusetts, USA) following the NEBNext Ultra II Version 5 protocol with size
169 selection on DNA fragments at 300-400 bp range. To make library preparation more
170 economic, the protocol was adjusted by halving all amounts of reagents and DNA used.

171 The barcoded libraries (NEBNext multiplex oligos) were quality checked with an Agilent
172 Bioanalyzer (Agilent Technologies, Santa Clara, California, USA) and quantified by
173 Qubit High Sensitivity assay. Libraries were then used for targeted sequence capture
174 with the custom probe set MyBaits Compositae-1061 kit (Mandel et al. 2014) from Arbor
175 Biosciences (Ann Arbor, Michigan, USA) following manufacturer's protocols (version
176 2.3.1). Approximately 500ng of total DNA library in 6 µl of library, or equal amount of
177 DNA per sample when pooling (maximum of 8 samples), was used in each MyBaits
178 reaction for hybridization with the baits for 36 hours. Hybridized targets were recovered
179 using Dynabeads M-280 Streptavidin (ThermoFisher Scientific, Vilnius, Lithuania)
180 following the manufacturer's protocol. Captured targets were amplified and quantified
181 using KAPA library quantification kits (Kapa Biosystems, Wilmington, Massachusetts,
182 USA). A 1:4 ratio of library was added to each of the captured target pools to obtain off-
183 target plastid sequence data. Sequence stats were obtained using the *pxlssq* function in
184 Phyx v. 1.3.1 (Table S2) (Brown et al. 2017).

185 **Sequence processing**

186 Raw FASTQ data were cleaned using the program Trimmomatic v. 0.36 (Bolger
187 et al. 2014) which removes Illumina adapters from low-quality bases and adapter
188 contamination. The Sliding Window quality filter was used in this analysis (illuminaclip
189 2:30:10, leading 20, trailing 20, sliding window 5:20). Cleaned reads were retained
190 when they had a minimum length of 36 bp and corresponding forward and reverse
191 pairs.

192 **Plastid assembly**

193 Chloroplast sequences were extracted from trimmed paired end reads using two
194 approaches. First, the GetOrganelle v. 1.7.7.1 toolkit (Jin et al. 2020) was used to

195 assemble circularized plastomes for all sequences with sufficient off-target sequencing
196 data. The recommended parameters for embryophyte plant plastomes were used but
197 the number of rounds and the disentangle time limit were increased (-R 75 and --
198 disentangle-time-limit 7200). All complete circular plastomes were annotated using
199 GeSeq (Tillich et al. 2017). From these, a *Packera* representative plastome was chosen
200 and used as a reference. Second, all reads were mapped to the reference plastome
201 using Bowtie2 (Langmead and Salzberg 2012), variants were called using Samtools
202 and BCFtools (Danecek et al. 2021), and consensus sequences were created using
203 seqtk (https://github.com/lh3/seqtk). For the aligned consensus sequences, model
204 selection and maximum likelihood phylogenetic analysis were then conducted using IQ-
205 TREE v. 2.2.2 (Minh et al. 2020) and visualized using 'ggtree' (Yu et al. 2017) in R v.
206 4.3.1 (R Core Team 2016; RStudio 2020).

207 **Nuclear ortholog assembly**

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

218 the run, then aligned into supercontigs using MAFFT v. 7.407 (Katoh and Standley
219 2013).

220 **Phylogenetic analyses**

221 Given the complex history of hybridization and polyploidization events in
222 *Packera*, along with the large number of paralogs flagged from HybPiper in the current
223 study (Table S3), we sought to compare how different published orthology inference
224 methods and species tree estimation tools impacted the inferred evolutionary
225 relationships and support of this group. Our dataset is uniquely adapted to tackle this
226 issue given the dense taxon sampling, high locus recovery across both orthologs and
227 paralogs, combined with decades of natural history and systematics work in the group,
228 enabling more thorough investigations into these challenges.

229 We generated multiple nuclear phylogenies utilizing different paralog processing
230 tools to infer orthology and species tree estimation: Concatenation (individual gene
231 histories not accounted for), HybPiper with ASTRAL-III (hereafter 'ASTRAL'; putative
232 single-copy orthologous loci are chosen based on coverage depth and sequence
233 similarity concerning the reference then inferred using coalescent methods), HybPiper
234 with wASTRAL (hereafter 'wASTRAL'; similar to ASTRAL but low support gene tree
235 branches are contracted prior to species tree estimation), ASTRAL-Pro (all single-copy
236 and multi-copy genes identified by HybPiper are considered, and the species tree that
237 optimizes a quartet similarity measure is chosen), and the 1-to-1, MI, MO, and RT
238 paralog processing methods by (Yang and Smith 2014) with species trees inferred
239 using ASTRAL-III (further details below; refer to Table S3). By performing intentional
240 comparisons among methods, we aim to gain a better understanding of the impact of

241 paralogy on phylogenomic studies, and ultimately increase phylogenomic support and
242 concordance across species of complex lineages.

243 Concatenation

244 Aligned supercontigs from MAFFT were concatenated with FASconCAT-G v.
245 1.02 (Kück and Longo 2014) then implemented in PartitionFinder v. 2.1.1 (Lanfear et al.
246 2012) to determine the most appropriate nucleotide substitution model. The overall best
247 fitting model (Table S4) was used to build a concatenated nuclear maximum likelihood
248 tree in RAxML v. 8.1.3 (Stamatakis 2014) with 1,000 bootstrap (BS) replicates.

249 ASTRAL and wASTRAL

250 For ASTRAL and wASTRAL analyses, nuclear gene alignments were separated
251 based on the most suitable nucleotide substitution model for each gene as determined
252 by PartitionFinder (Table S4), then individual gene trees were generated using RAxML.
253 For the ASTRAL analysis, the gene trees were implemented in ASTRAL-III v. 5.7.3
254 (Zhang et al. 2018) to generate the species trees. For the wASTRAL analysis, or
255 weighted ASTRAL, a species tree was generated using 'astral-hybrid' in the ASTER
256 program (https://github.com/chaoszhang/ASTER). This method has shown to be more
257 accurate than typical ASTRAL methods since wASTRAL incorporates gene tree
258 uncertainty by weighting quartets based on the confidence levels of the input gene
259 trees. In doing so, wASTRAL reduces the influence of erroneous gene trees, leading to
260 a more reliable species tree estimation (Zhang and Mirarab 2022b). Local posterior
261 probability (LPP) values were estimated for branch support and indicate the probability
262 that the branch is the true branch given the set of gene trees provided. LPP is
263 considered the more reliable clade support measure when running ASTRAL analyses

since LPP is computed based on a quartet score and assumes incomplete lineage sorting (ILS) (Sayyari and Mirarab 2016).

ASTRAL-Pro

ASTRAL-Pro2 (ASTRAL for PaRalogs and Orthologs) v. 1.13.1.3 (Zhang et al. 2020; Zhang and Mirarab 2022a), hereafter referred to as ASTRAL-Pro, provides a quartet-based species-tree inference method that allows multi-copy gene trees to be used as input. Thus, all paralogous and orthologous sequences are used to generate the species tree instead of selecting a single ortholog to represent an individual gene's history (e.g., ASTRAL method). By accommodating multi-copy gene trees, ASTRAL-Pro accounts for gene duplication and loss events, reducing biases that arise from treating paralogs as true orthologs. This approach is supposed to improve species tree accuracy, particularly in datasets with extensive gene family expansions, by leveraging more information from complex gene histories. The ASTRAL-Pro pipeline is similar to ASTRAL, as described above, but differs by having the user generate separate alignments and conduct maximum likelihood analyses on the genes that are flagged as paralogs and single-copy contigs, separately. Single-copy and paralogous gene sequences were independently aligned using MAFFT, and gene trees were generated under the best-fitting nucleotide substitution model from PartitionFinder (GTR, GTR+Γ, GTR+I+Γ; Table S4) in RAxML with 1,000 BS replicates. The RAxML output files were then used as input for ASTRAL-Pro2 to generate the species tree with LPP values at each node.

Paragone-nf

Paralog processing was then carried out using the tree-based orthology inference strategies originally proposed by (Yang and Smith 2014) via the paragone-nf

288 v. 1.1.0 pipeline (Jackson et al. 2023) (https://github.com/chrisjackson-
289 pellicle/paragone-nf). This pipeline contains a Nextflow v. 23.10.0 (Di Tommaso et al.
290 2017) script downloaded through a Singularity v. 3.7.1 container (Kurtzer et al. 2017,
291 2021) to investigate the loci flagged as paralogous from HybPiper and determines which
292 loci can be retained or pruned given specified conditions (Yang and Smith 2014;
293 Morales-Briones et al. 2021a). All orthology inference methods proposed by (Yang and
294 Smith 2014)–the 'one-to-one orthologs' (1-to-1), 'maximum inclusion' (MI),
295 'monophyletic outgroup' (MO), and 'rooted ingroup' (RT) methods–were tested. Briefly,
296 the 1-to-1 approach is the simplest strategy in that it only retains homologs that have no
297 duplications. The MI approach extracts the subtree with the highest number of taxa
298 without taxon duplications. The MO approach filters for homolog trees where outgroup
299 taxa (here tribe Calenduleae: *Calendula arvensis*, *Tripteris pinnatilobatum* (Norl.) B.
300 Nord., and *Chyrsanthemoides incana* (Burm.f.) Norl.) are monophyletic and non-
301 repeating. The RT approach searches subtrees of ingroup taxa and roots them by the
302 designated outgroups, then removes the ingroup taxa as rooted trees. In turn, inferred
303 gene duplications, or paralogs, are then pruned from root to tip in the rooted tree (Yang
304 and Smith 2014; Morales-Briones et al. 2021a).
305 To aid in consistency across analyses, the HybPiper run described above was
306 used as input instead of using the hybpiper-nf pipeline (Jackson et al. 2023), which is
307 another NextFlow script generated by the authors of the paragone-nf pipeline. The three
308 Calenduleae taxa were considered the internal outgroups since this tribe has not
309 experienced a recent whole genome duplication event, which is required for the MI and
310 MO analyses (Yang and Smith 2014). The resulting trimmed alignments produced via
311 paragone-nf were used as input for RAxML to generate gene trees, and a species tree

312 with LPP support values was generated using ASTRAL-III following the 'ASTRAL'
313 protocol described above.
314 The normalized quartet score was calculated for all orthology inference methods
315 that generated a species tree with ASTRAL-III using the '-q' option. Branch lengths were
316 recalculated from coalescent units to nucleotide substitutions per codon site for each
317 ASTRAL-III species tree using IQTree. Parsimony informative sites (PIS) and pairwise
318 identity (PI) were calculated for each gene matrix using the
319 'parsimony_informative_sites' and 'pairwise_identity' functions, respectively, in phykit v.
320 2.0.1 (Steenwyk et al. 2021). Nuclear and plastome trees were visualized using the
321 package 'phytools' (Revell 2012) in R v. 4.0.5.
322 **Measuring discordance**
323 Species trees from concatenation, plastid, ASTRAL, wASTRAL, and paragone-nf
324 (1-to-1, MI, MO, and RT) were analyzed for underlying discordance using Quartet
325 Sampling (Pease et al. 2018). ASTRAL-Pro was not analyzed as Quartet Sampling
326 does not allow multi-copy gene trees as input. Quartet Sampling assesses the
327 confidence, consistency, and informativeness of internal tree relationships, and the
328 reliability of each terminal branch by evaluating if a phylogeny has a lack of support due
329 to lack of information, discordance from ILS or introgression, or misplaced or erroneous
330 taxa. Three scores are calculated for each internal branch of the focal tree which
331 determines the consistency of information (Quartet Concordance [QC], the presence of
332 secondary evolutionary histories (Quartet Differential [QD], and the amount of
333 information (Quartet Informativeness [QI]). Quartet Fidelity (QF) scores are calculated at
334 each tip and reports how frequently a taxon is included in concordant topologies.
335 Quartet Sampling efficiently synthesizes phylogenetic tests and offers more

336 comprehensive and specific information on branch support than conventional measures
337 (i.e., BS or LPP; (Pease et al. 2018). Using the rooted species trees and concatenated
338 gene alignments from FASconCAT-G (described above), Quartet Sampling was run
339 using default settings on the designated analyses. The concatenation, plastid, ASTRAL,
340 wASTRAL, 1-to-1, MI, and MO species trees were rooted to the outgroup Calenduleae
341 taxon, *Tripteris pinnatilobatum* (Norl.) B. Nord., while the RT tree was rooted to
342 *Doronicum pardalianches* L. using the *pxrr* function in Phyx (Table S3).
343 Additionally, PhyParts v. 0.0.1 (Smith et al. 2015) was used with default settings
344 to quantify and visualize discordance in the ASTRAL, wASTRAL, and paragone-nf (1-to-
345 1, MI, MO, and RT) trees. Similar to Quartet Sampling, ASTRAL-Pro could not be
346 analyzed because of multi-copy gene trees. Additionally, the concatenation and plastid
347 trees could not be analyzed since those analyses do not use individual gene histories
348 as input. Resulting trees containing pie charts indicating whether gene trees were
349 supported, conflicted, or uninformative at each node were generated using a custom
350 python script
351 (https://github.com/mossmatters/phyloscripts/tree/master/phypartspiecharts). PhyParts
352 requires rooted gene trees and a rooted species tree as input. Similar to Quartet
353 Sampling, ASTRAL, 1-to-1, MI, and MO gene trees were rooted to *Tripteris*
354 *pinnatilobatum* and the RT tree was rooted to *Doronicum pardalianches* using the *pxrr*
355 function in Phyx. Rooting gene trees for PhyParts decreased the number of genes used
356 as input since not all gene trees contained the targeted outgroup taxon (see Table S3).
357 For phylogenies generated with ASTRAL technologies–ASTRAL, wASTRAL,
358 ASTRAL-Pro, and paragone-nf trees (Table S3)–introgression and ILS were visualized

359 across lineages using phytop v. 0.3.1 (Shang et al. 2025) using default settings. The
360 species trees were re-ran with the flag '-t 2' added to be used as input for phytop.

361 **Tree topology comparison**

362 To evaluate similarity among the all nuclear gene trees, pairwise adjusted
363 Robinson-Foulds (RF_{adj}) distances were calculated as outlined by (Mitchell et al. 2017)
364 between all trees using the *RF.dist()* function in package 'phangorn' (Schliep 2011) in R,
365 with unrooted trees as input. The Calenduleae outgroup taxa were trimmed from the
366 trees in R using the *drop.tip()* function in package ape (Paradis and Schliep 2019) to
367 make concatenation, ASTRAL, ASTRAL-Pro, and paragone-nf trees directly
368 comparable. The RF distance was adjusted by setting the "normalize" argument to
369 TRUE, which divides RF by the maximal possible distance as described in (Steel and
370 Penny 1993). Given that the input trees are unrooted and binary, this value is then 2n-6,
371 where n is the number of nodes on the tree (resulting equation: $RF_{adj} = RF/(2n-6)$). The
372 RF_{adj} values range from zero to one, indicating whether the tree topologies are identical
373 or completely dissimilar, respectively. A principal component analysis (PCA) was
374 generated from all RF_{adj} values in R to aid in visualization.

375 All nuclear and plastid phylogenies and their associated analyses described
376 above will be provided and compared, but for simplicity only the ASTRAL, 1-to-1, and
377 MO trees will be analyzed in further analyses. ASTRAL was chosen as it selects the
378 putative single-copy orthologs less rigorously than the remaining methods, creating a
379 baseline method to compare to. We then wanted to compare ASTRAL to a method that
380 disregards genes containing paralogs (1-to-1) and a more rigorous paralog-processing
381 method (MO). We chose MO over MI and RT due to its more stringent ortholog
382 identification, which minimizes the inclusion of paralogs and reduces phylogenetic

383 noise. While RT and MI retain larger gene sets by tolerating ambiguous orthologs, they
384 increase the risk of incorporating non-orthologous sequences, leading to potential
385 topological artifacts and reduced tree accuracy. Additionally, the conservative filtering of
386 MO enhances alignment quality and computational efficiency, providing a more reliable
387 foundation for downstream phylogenetic analyses (Yang and Smith 2014; Morales-
388 Briones et al. 2021a; Jackson et al. 2023). Unfortunately, the paralog processing
389 method that includes all genes, ASTRAL-Pro, could not be included since further
390 analyses require only single-copy orthologs and the underlying gene trees of ASTRAL-
391 Pro are single- and multi-copy.

392 **Molecular clock dating**

393 Estimates of divergence times were calculated on the ASTRAL, 1-to-1, and MO
394 trees using RelTime (Tamura et al. 2012, 2018), which utilizes the relative rate
395 framework (RRF) and has been shown to be efficient for larger datasets and produces
396 similar results to Bayesian approaches than other non-Bayesian dating methods such
397 as treePL (Smith and O'Meara 2012; Tao et al. 2020a, 2020b; Barba-Montoya 2021;
398 Costa et al. 2022). RelTime was implemented using the command-line version of the
399 MEGA software, MEGA-CC v. 11.0.11 (Kumar et al. 2012, 2016), with necessary input
400 files generated with MEGA-X (Kumar et al. 2018; Mello 2018). The three Calenduleae
401 taxa were used as outgroup taxa for this analysis. Since the branch lengths of the three
402 phylogenies were in coalescent units from ASTRAL-III, branch lengths of each species
403 tree were recalculated to the number of nucleotide substitutions per codon site prior to
404 running RelTime using the GTR+I+Γ model in IQTree.

405 There are no fossils available within *Packera* or tribe Senecioneae, so instead,
406 multiple dating scenarios were tested utilizing both primary and secondary age

407 constraints (Table S5). Briefly, a maximum (59.95 MY) and minimum (31 MY) age was
408 tested for the Anthemideae/Senecioneae split given those dates are the highest and
409 lowest values within the 95% CI from a previous diversification rate analysis of
410 Asteraceae (Mandel et al. 2019). Additionally, a maximum age of 34.06 MY was set at
411 the *Packera* crown node from the upper age limit provided by (Mandel et al. 2019).
412 Next, a minimum age constraint of 12 MY was tested within the outgroup-Anthemideae
413 clade given the presence of a fossil pollen (Wang 2004; Hobbs and Baldwin 2013;
414 Palazzesi et al. 2022). Finally, a maximum age constraint of 14.3 MY was tested for
415 *Pericallis* D.Don (Senecioneae), a genus endemic to the Macaronesia islands (Jones et
416 al. 2014), since that is the estimated age of Porto Santo, the oldest volcanic archipelago
417 within the Madeira Islands (McDougall and Schmincke 1976; Geldmacher et al. 2000).
418 Results of all tested scenarios were compared for biological relevance, though only one
419 scenario was chosen for the discussion. This approach allowed us to focus on the
420 topology that best aligned with prior systematic understanding and showed the
421 strongest overall support across analytical frameworks.

422 **Biogeography**

423 Ancestral ranges were estimated on the dated phylogenies using the package
424 BioGeoBEARS v. 1.1 (Matzke 2014) in R. BioGeoBEARS estimates historical,
425 biogeographic ranges by implementing popular models in a likelihood framework. Some
426 of the most common models used are dispersal-extinction-cladogenesis (DEC; (Ree
427 and Smith 2008), the likelihood version of DIVA (DIVALIKE; (Ronquist 1997), and the
428 BayArea likelihood version of the range evolution model (BAYAREALIKE; (Landis et al.
429 2013). In BioGeoBEARS, founder-event speciation can be added to any of these
430 models (+*J*) with flexibility in determining relative probability. Likelihood Ratio Test (LRT)

431 and Akaike Information Criterion (AIC) was used to determine which model best fit the
432 data.

433 All *Packera* taxa, along with their four most closely related outgroup genera
434 (*Pericallis*, *Elekmania* B.Nord., *Werneria* Kunth, and *Xenophyllum* V.A.Funk), were
435 coded as present or absent in ten geographic ranges outlined in Takhtajan (1986)
436 (Table S6): Arctic Province, Canadian Province, Appalachian Province, Atlantic and
437 Gulf Coast Plain Province, North American Prairies Province, Rocky Mountain and
438 Vancouverian Province, Madrean Region, Asia, Europe, and Central America.
439 Distributions were determined from previously published literature (Freeman and
440 Barkley 1995; Trock 1999, 2006; Gramling 2006; Mahoney and Kowal 2008; Elven et al.
441 2011; Yeatts et al. 2011). Only four of the 26 outgroup taxa were included to limit the
442 number of geographic ranges used in the analysis. The remaining outgroup taxa were
443 pruned from the dated tree generated with RelTime using the *pxrmt* function in Phyx.
444 Each taxon was limited to a max distribution of four ranges.

445 **Reticulation**

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

455 support nodes in the ASTRAL, 1-to-1, and/or MO trees, as determined by a negative
456 Quartet Concordance (QC) score resulting from Quartet Sampling, were used as input
457 for PhyloNetworks, with the remaining taxa pruned from the tree using the *pxrmt*
458 function in Phyx. Given this, five *Packera* clades were investigated for ancient
459 reticulation events (Figs. 1-3). Additionally, a single species from each major clade was
460 then used to represent backbone relationships throughout the entire phylogeny, totaling
461 six separate PhyloNetwork runs that will be referenced as: Backbone: representing 11
462 taxa that span the entire phylogeny; Eastern: a clade of 12, 13, or 15 species (MO, 1-to-
463 1, and ASTRAL, respectively) either distributed along eastern North America or having
464 known affiliations with those species; California: eight taxa with distributions mainly
465 within California; Mexico: seven (MO and ASTRAL) to eight (1-to-1) taxa with
466 distributions mainly within Mexico; Rocky: a clade of five taxa that are distributed in the
467 Central US, most within the Rocky Mountains; and Arctic: nine taxa that are present in
468 colder climates along the Pacific Northwest and into Canada Table S7). Rocky could not
469 be run for 1-to-1 since less than five taxa were present in that target clade (Fig. 2).
470 Hybridization networks were visualized using Dendroscope v. 3.7.6 (Huson and
471 Scornavacca 2012).
472

473 **Results**

474 **Matrix results**

475 Illumina sequencing resulted in approximately 733.3 million reads across 111
476 taxa. The minimum and maximum number of reads ranged from 535,934 in *Werneria*
477 *aretioides* Wedd. and 20,685,576 in *Senecio vulgaris* L. (full list in Table S2). HybPiper
478 retained 1,058 of the targeted 1,061 loci across all 111 taxa with 846 of those loci

479 having paralog warnings (~80%). The number of loci recovered for each taxon ranged
480 from 566 in *Packera paupercula var. paupercula* (Michaux) Á. Löve & D. Löve to 1,054
481 in *Calendula arvensis* L. (Table S8). The number of gene trees varied across analyses,
482 with the least being 285 in the 1-to-1 analysis to 7,388 in the MI analysis (Table S3).
483 Parsimony informative sites (PIS) also ranged from 2.11% in the plastid (mapping) tree
484 to 99.64% in the MO tree (Table S3). For species trees generated with ASTRAL-III,
485 normalized quartet scores were used to measure gene tree discordance. In doing so,
486 the normalized quartet score values ranged from 0.516 with the RT tree to 0.534 with
487 the MO tree (Table S3), indicating that ~52-53% of the gene trees match the species
488 tree in those analyses.

489 Off-target plastid sequencing resulted in partial (i.e., mapping; n = 111) to whole
490 (i.e., complete; n = 18) plastomes. Complete plastomes from 7 *Packera* and one
491 *Senecio* species ranged from 150,772 bp to 151,728 bp (Fig. S1). The percentage of
492 missing data for each mapped consensus sequence varied drastically, with *Tephroseris*
493 *newcombei* exhibiting the most missing data (93.1%; Table S9, Fig. S2). The complete
494 plastome supermatrix with 18 taxa had 0.123% PIS while the mapping supermatrix with
495 111 taxa had 2.11% PIS (Table S3).

496 Raw read and assembly information for the 111 taxa can be found in Tables S2
497 & S8. General matrix information and summary statistics for all analyses can be found
498 in Table S3.

499 **Topological comparisons and discordance**

500 Phylogenetic topologies varied depending on sequence type (e.g., nuclear or
501 plastid) or orthology inference and phylogenetic tree inference method (i.e., ASTRAL,
502 ASTRAL-Pro, 1-to-1, etc.) (Figs. 4 & S3). With the nuclear data, *Packera* is generally

503 placed as the sister lineage to a clade comprising *Pericallis* and *Elekmania*, with the
504 exception of ASTRAL-Pro which placed *Packera* as sister to *Tephroseris* and in a larger
505 clade with *Pericallis* and *Elekmania* (Fig. 4). Interestingly, the plastid data showed a
506 similar relationship to ASTRAL-Pro; however, *Pericallis* is placed unexpectedly in a
507 separate clade with other Senecionineae taxa. All nuclear phylogenies recovered
508 monophyletic tribes and subtribes (e.g., Senecionineae and Tussilaginineae s.str.),
509 though the relationships within the subtribes varied. The mapped plastid tree generally
510 separated the subtribes, with the exception of *Roldana* (Senecionineae) in
511 Tussilaginineae s.str. (Figs. 4 & S3). *Tephroseris* (tribe Tephroseridineae) moved
512 constantly with its most consistent placement being sister to Senecioneae (Fig. 4).
513 *Doronicum* is consistently placed sister to Senecioneae, providing support for its
514 elevation to tribal status (Doroniceae; (Panero 2005; Fu et al. 2016).
515 Phylogenetic topologies remained highly incongruent independent of methods
516 used (Figs. 1-3, Table S10). The ASTRAL and wASTRAL tree topologies were most
517 similar (RF_{adj} = 0.324) (Fig. 4, Table S10), while the plastid tree was consistently the
518 most dissimilar to the remaining trees (RF_{adj} ranging from 0.895 to 0.914), with the
519 concatenated tree being the next most dissimilar (RF_{adj} ranging from 0.562 to 0.648;
520 Table S10). Additionally, the normalized quartet scores from ASTRAL indicate that only
521 half of the gene trees match the species trees in the ASTRAL and paragone-nf analyses
522 (Table S3). Most of the topological conflicts were among *Packera* taxa, with minor
523 differences among the remaining Senecioneae taxa and outgroups (Fig. S3). Support
524 values varied depending on which sequence type and method was used, though
525 generally remained low. The wASTRAL and MI trees showed the highest node support
526 with 58 out of 110 nodes (52.7%) having ≥0.95 LPP (Table S3 & Fig. S3). In contrast,

527 the plastid and concatenated trees showed much lower resolution with the fewest
528 number of nodes having high support (36 of 110 nodes, 32.7%, have ≥95 BS; Table S3
529 & Fig. S3).

530 Quartet Sampling indicated that underlying gene tree discordance remains high,
531 independent of the method used. ASTRAL and concatenated trees had the most fully
532 supported nodes (QS = 1/–/1) when compared to the remaining phylogenies (Fig. S4),
533 though the number was still low (~10%). Even so, the concatenated tree also had the
534 highest proportion of counter-support nodes (QC = <0; 26.4%; see Fig. S4 for all
535 Quartet Sampling results). PhyParts indicated that the ASTRAL and MO trees have the
536 most concordant gene trees. The wASTRAL analysis had the highest percentages of
537 gene trees supporting the top alternative bipartition. The MI tree harbored the highest
538 discordance and uninformative data (Fig. S5).

539 This underlying discordance could be caused by ILS, as it was common in all
540 tested ASTRAL trees (Fig. S6). Introgressive hybridization (IH) was not detected without
541 the presence of ILS in any of the nodes, though IH was most common in the nodes that
542 were shown to be discordant (Figs. S4 & S6). Interestingly, crown *Packera* consistently
543 had high IH values, potentially indicating introgression at that node. Backbone nodes
544 were generally considered well resolved with high confidence (ILS index < 50% and IH
545 index <10%; (Shang et al. 2025), especially in the outgroups, whereas *Packera* taxa
546 appeared less resolved.

547 **Molecular clock dating**

548 The resulting divergence rate estimates differed among ASTRAL, 1-to-1, and
549 MO, with estimated ages of the MO and 1-to-1 trees older and sometimes almost
550 double the ASTRAL ages. For ASTRAL, the age for crown *Packera* ranged from 11.4

551 MY in Scenario 5 and 6 to 13.5 MY in Scenario 1-4. Alternatively, the crown age of
552 *Packera* were almost doubled in the 1-to-1 and MO analyses, with ranges from ~21 MY
553 in Scenario 5 to 35.5 MY-35.7 MY in Scenario 4 (Table S11, Fig. S7). Scenario 2 was
554 chosen as the best-fitting dating model because it incorporated all available primary
555 calibration points and yielded divergence time estimates that align well with known
556 patterns of evolutionary history (Pelser et al. 2010; Mandel et al. 2019), while minimizing
557 reliance on secondary calibrations (Table S5) (Müller and Reisz 2005; Benton and
558 Donoghue 2007; Parham et al. 2012). The dated trees from all six scenarios can be
559 found in Fig. S7.

560 **Biogeography**

561 Evaluation of the six different model choices demonstrated that BAYAREALIKE
562 was the best fit given our data based on its lowest log likelihood (Ln L) and highest AIC
563 score, as well as the highest weighted AICc score for all three of our analyses (Table
564 S12). The BAYAREALIKE model is a simplified likelihood interpretation of the BayArea
565 program described by (Landis et al. 2013), which tests the importance of cladogenesis
566 on a particular dataset, but it does not depend on geographic distance between taxa
567 and allows for a larger number of areas. Generally, models including jump speciation
568 (+*J*) provided a better fit than those without in the DEC and DIVALIKE models, and were
569 the same or very similar in the BAYAREALIKE models (Table S12), indicating a
570 potential for jump dispersal or founder-event speciation in our data. Trees of all model
571 runs are in Fig. S8.

572 The historical biogeography analysis revealed distinct geographic patterns
573 across the different orthology methods (Figs. 5 & S8). ASTRAL identified the ancestral
574 range as concentrated in the North American Prairies Province (N), Rocky Mountain

575 and Vancouverian Province (R), and Mandrean (M) regions. In contrast, MO showed
576 the ancestral regions being only the Rocky Mountain and Vancouverian Province (R),
577 and Mandrean (M) regions. 1-to-1 was intermediate with crown *Packera* being ancestral
578 to the North American Prairies Province (N), Rocky Mountain and Vancouverian
579 Province (R), and Mandrean (M) regions, similar to ASTRAL; however, the remaining
580 backbone was Rocky Mountain and Vancouverian Province (R), and Mandrean (M)
581 regions, similar to MO. Notably, all analyses showed one major expansion into the
582 eastern United States–Appalachian Province (P) and Atlantic and Gulf Coast Plain
583 Province (T)–represented by the Eastern clade (Figs. 5 & S8), with a few instances of
584 individual introductions to the eastern United States. Many of these taxa originated
585 initially in the North American Prairies Province (N) and/or R ranges. ASTRAL and MO
586 showed two early expansions into the Arctic Province (I) and Canadian Province (C) at
587 the base of the phylogeny, while 1-to-1 showed at least three expansions north. 1-to-1
588 exhibited the most geographic inconsistency, followed by MO. Even so, geography
589 proved to be a reliable indicator for larger clades across all analyses, although
590 widespread species were scattered across the phylogeny and did not always align
591 consistently within geographic clades (Figs. 1-3 & S8).

592 **Reticulation events**

593 Five *Packera* clades were tested for ancient reticulation events given a large
594 number of counter-support (QC < 0) nodes from Quartet Sampling on the ASTRAL, 1-
595 to-1, and MO trees (Figs. 1-3). Additionally, a single species from each major clade was
596 then used to represent backbone relationships throughout the entire phylogeny, totaling
597 six separate PhyloNetwork runs named: Arctic, Backbone, California, Eastern, Mexico,
598 and Rocky (Table S13). Log likelihood values indicated that the most likely number of

599 reticulation events (h) varied depending on the run (Fig. S9). Additionally, Arctic,
600 Mexico, and Backbone had varying numbers of optimal reticulation events between the
601 paralog-processing methods (Table S7, Fig. S9). For ASTRAL, Arctic and Mexico had
602 the highest number of reticulation events (h = 3), followed by California and Backbone
603 (h = 2), then Eastern and Rocky (h = 1). For 1-to-1, Backhone had the most (h = 6),
604 then Mexico (h = 3), Arctic and California (h = 2), and then Eastern (h = 1). For MO,
605 Arctic, California, and Backbone had the highest number of reticulations (h = 2) with
606 Eastern, Mexico, and Rocky having the lowest (h = 1) (Figs. 6 & S9, Table S7). The
607 taxa also differed between analyses given different phylogenetic relationships between
608 clades (Table S13), impacting some of the networks identified.

609 Some networks showed different events depending on the phylogenetic method
610 and subset of data (Fig. 6); however, some consistent patterns emerged. For example,
611 the Eastern clade continuously showed the same reticulation event regardless of the
612 paralog processing method, and parentage percentages were very similar. In contrast,
613 Rocky did not show the same reticulation events, but the networks generally involved
614 *Packera mancosana* and *P. streptanthifolia*. The California clade showed differing
615 results across methods: both ASTRAL and 1-to-1 analyses identified *P. bernardina* as
616 the product of reticulation with the same parentage, whereas MO did not yield the same
617 results, and no consistent trends emerged. The Mexico clade was generally
618 inconsistent, but did feature reticulation events involving *P. neomexicana* and its
619 varieties. Additionally, *P. texensis* and *P. quercetorum* were involved in reticulation
620 events in two of the three paralog processing networks. Among all the clades, the Arctic
621 clade exhibited the most variability in the number of reticulation events, with the *P.*
622 *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 not show the same reticulation events but did identify a reticulation event involving *Tephroseris* with a different parentage. Notably, the parentage of *Tephroseris* always included a *Packera* taxon, though the specific taxon varied.

Discussion

Gene duplication was pervasive in *Packera*, with more than 80% of nuclear loci showing evidence of paralogy (Table S3). Such widespread paralogy poses major challenges for species tree inference, especially in Hyb-Seq datasets, where probes may unintentionally enrich for multiple homeologs (Salmon et al. 2012; Akama et al. 2014; Crameri et al. 2022), complicating the recovery of true single-copy orthologs for phylogenetic reconstructions. Because HybPiper's orthology inference relies largely on single-copy filtering (Johnson et al. 2016), using its outputs alone can introduce noise and bias into gene tree reconstruction, ultimately distorting species tree inference. Consistent with our study, using only putative single-copy orthologs designated by HybPiper was insufficient for accurate phylogenetic inference in *Packera*, a group characterized by extensive polyploidy and hybridization (Trock 2006). These challenges are not unique to *Packera*, as many plant lineages, such as Asparagaceae (Bentz and Leebens-Mack 2024), Rosaceae (Morales-Briones et al. 2018, 2021a), and Amaranthaceae (Morales-Briones et al. 2021b), also experience difficulties reconstructing phylogenies due to hybridization and gene/genome duplications. Similar issues occur in non-plant systems, including the bacterial order Enterobacteriales

647 (Adeolu et al. 2016), annelids (Struck 2013), and modern amphibians (Siu-Ting et al.
648 2019). Together, these examples highlight the complexity of orthology inference in
649 phylogenomics and the need for rigorous paralog processing across both plant and non-
650 plant systems.

651 The most recent phylogeny of *Packera* prior to this study was based on ITS data
652 and revealed extremely low sequence variation, with only 0–5% divergence and several
653 *Packera* taxa sharing identical sequences (Bain and Golden 2000). This led to a largely
654 unresolved tree, likely due to few phylogenetically informative sites, sparse species
655 sampling, and potential reticulation events (Funk 1985; Linder and Rieseberg 2004). By
656 using Hyb-Seq data and broader taxon sampling, we substantially improved
657 phylogenetic resolution across *Packera*, including within major clades. However, our
658 analyses also revealed that gene tree concordance and species tree resolution varied
659 substantially across paralog processing methods (Fig. S3), exemplifying the influence of
660 orthology inference strategies on downstream phylogenetic outcomes. Methods that
661 explicitly model gene duplication and loss, such as those implemented in paragone-nf
662 (Yang and Smith 2014; Jackson et al. 2023), generally yielded more concordant gene
663 trees and better-resolved species trees (Figs. 2-3). However, substantial conflict
664 persisted across the trees, indicating that even sophisticated modeling approaches
665 cannot fully resolve deep relationships in groups with complex histories like *Packera*.
666 These discrepancies may in part reflect differences in the number and identity of genes
667 retained after filtering (Table S3). For instance, the 1-to-1 method, which excludes all
668 loci with paralogs and resembles approaches used in tools like Phyluce, eliminates a
669 significant portion of the dataset and may oversimplify the data, especially in
670 duplication-rich systems such as *Packera*. While Phyluce and 1-to-1 filtering both

remove paralogs, the underlying methods differ (Faircloth 2016; Jackson et al. 2023), causing Phyluce to outperform HybPiper in certain cases (Mandel et al. 2019; Thapa et al. 2020). Similarly, the RT method naturally loses information since it removes outgroups to determine orthology (Yang and Smith 2014; Morales-Briones et al. 2021b, 2021a) and potentially limits phylogenetic signals. These findings highlight that while some approaches have potential to improve concordance, none fully resolve the underlying discordance of *Packera*, emphasizing that the choice of method, underlying data, and biological context must guide ortholog selection in phylogenomic analyses.

Results from a previous phylogenetic study of *Packera* showed that its relationships tend to follow a loose geographic pattern (Bain and Golden 2000), which is consistent with our findings (Figs. 1-3 & S8). Our nuclear phylogenies showed that major clades tended to contain taxa within similar geographic ranges, independent of which orthology inference method was utilized. Additionally, most taxa within these major clades were conserved across methods, though their placement along the backbone, and some taxa, did vary (Figs. 1-3). For example, the taxa in the Eastern clade was fairly consistent between the orthology inference methods, though some taxa did move in and out of the clade (Fig. S3). Interestingly, some of these taxa are widespread species that interspecifically hybridize with other *Packera* taxa they come into contact with. For example, *P. pseudaurea* and its varieties are typically found in central North America, though their phylogenetic placement is generally in the Eastern clade. These results are unsurprising as they have an extensive history of hybridization with other *Packera* taxa in the Eastern clade, which is believed to have created much taxonomic uncertainty (Barkley 1962; Trock 1999), potentially explaining their inconsistent phylogenetic placements (Fig. S3).

695 Our results also reveal that evolutionary interpretations such as divergence
696 dating and biogeographic reconstruction are highly sensitive to the paralog processing
697 approach used to infer the underlying phylogeny. Divergence time estimates varied
698 across methods, with some species trees suggesting that crown *Packera*'s
699 diversification occurred more recently (ASTRAL: 13.5 MY), while others placed key
700 divergences deeper in time (1-to-1: 24.2 MY and MO: 25 MY; Fig. S7, Table S11).
701 These differences are not trivial as they affect our understanding of how *Packera*
702 responded to climatic shifts, mountain uplift, and other historical events, with broader
703 implications for comparative studies across Asteraceae and other radiating plant
704 lineages. For example, tribal- and family-wide studies predicted the crown age of
705 *Packera* to be fairly young, between 7.8 MY and 20.2 MY (Pelser et al. 2010; Mandel et
706 al. 2019; Zhang et al. 2021), which is more in-line with the ASTRAL analysis. While
707 more *Packera*-specific predictions are older, with (Barkley 1988) predicting the origin of
708 *Packera* in the later stages of the mid-Tertiary (~22-37 MY), closely aligning with our
709 paragone-nf analyses (Fig. S7). Similarly, biogeographic patterns shifted depending on
710 the inferred topology, with each analysis showing different migration patterns over time
711 (Figs. 5 & S8). Even so, the broad pattern is relatively unchanged. For example, all
712 topologies generally agreed that *Packera* originated in western North America with one
713 major expansion east and then later expansions north from the western clades. From
714 there, individual taxa expanded or became specialized to certain regions, with most of
715 the diversity present in western North America. The consistency across topologies lends
716 confidence that these patterns reflect the underlying evolutionary history within *Packera*,
717 rather than artifacts of methodological bias.

718 Despite improvements from paralogy-aware filtering, the species trees for
719 *Packera* were not fully resolved, indicating that underlying biological processes such as
720 ILS, hybridization, and introgression also play significant roles in shaping *Packera*'s
721 evolutionary history. This is supported by strong cytonuclear discordance, where plastid
722 phylogenetic relationships vary significantly from the nuclear trees (Figs. 4 & S3),
723 consistent with chloroplast capture or ancient hybridization events (Doyle 1992;
724 Maddison 1997; Galtier and Daubin 2008; Soltis and Soltis 2009; Smith et al. 2015; Stull
725 et al. 2020). Directly testing for gene flow reveals multiple introgression events,
726 particularly along short internal branches susceptible to ILS (Fig. S6). However, the
727 specific reticulation patterns recovered varied depending on the paralog processing
728 method used. These discrepancies highlight a critical point: gene selection—and
729 whether or how paralogs are retained—directly shapes reticulation inference. Each
730 gene has its own evolutionary trajectory, shaped by recombination and gene duplication
731 and loss (Betrán 2015), and filtering strategies that exclude paralogs or collapse
732 duplications can erase signals of past hybridization. Consequently, some reticulation
733 events, such as those involving Eastern *Packera* lineages, were consistently recovered
734 across datasets, while others appeared only in specific analyses.
735 Our findings emphasize that accurately inferring reticulate evolutionary histories
736 requires embracing the independent histories of genes and carefully considering how
737 orthology is defined. These differences should not be dismissed as noise. Rather, they
738 may reflect the underlying biological complexity of the group. In *Packera*, where
739 hybridization and genome duplication are prevalent, gene tree heterogeneity is
740 expected and informative. Our findings emphasize that accurately inferring reticulate
741 evolutionary histories requires embracing the independent histories of genes and

742 carefully considering how orthology is defined. Methods that oversimplify gene
743 relationships risk masking the very patterns needed to understand complex evolutionary
744 processes.

745 We note that network analyses such as PhyloNetworks, are highly sensitive to
746 sampling bias, so even though we sampled 85 out of a targeted 91 *Packera* taxa (~93%
747 coverage of the genus), missing taxa could impact the results. Including the remaining
748 and unsampled *Packera* taxa in future studies could help increase our understanding of
749 these relationships. Additionally, limiting the analyses to specified clades can ignore
750 other potential reticulation events occurring across groups. We recognize this limit;
751 however, computational requirements of current reticulation network analyses limit the
752 ability to sample more taxa or loci (see (Hejase and Liu 2016). With recent bioinformatic
753 advances and increased interest in non-bifurcating trees, we anticipate that less
754 computationally intense and efficient phylogenetic and network tools will soon be
755 developed, allowing researchers to test for reticulations across large phylogenies.

756 Another necessary consideration with hybridization networks is the idea of ghost
757 introgressions, which are ancient introgression events that leave genetic signatures of
758 extinct species in present-day species (Ottenburghs 2020). Accounting for ghost
759 introgressions has important implications for macroevolutionary studies, such as
760 improving the accuracy of trait evolution and biogeographic reconstructions, refining
761 divergence time estimates, and revealing hidden contributions of hybridization to
762 diversification (Ottenburghs 2020; Tricou et al. 2022). Network methods such as
763 PhyloNetworks can detect ghost introgressions, but these methods do not specify that
764 the hybrid edge is a ghost introgression, so it is up to the researcher to recognize its

765 presence, which can be difficult especially without background knowledge of the group
766 (Solís-Lemus and Ané 2016; Solís-Lemus et al. 2017).
767 Overall, our study highlights the necessity of employing multiple paralog
768 processing pipelines and integrating gene flow analyses to accurately infer species
769 relationships in groups with complex evolutionary histories. In such systems, pipelines
770 that rely solely on single-copy orthologs, like ASTRAL, can oversimplify complex
771 histories and mislead species tree inference. By employing methods that explicitly
772 model gene duplication and loss, we yielded modest improvements in resolution and
773 concordance, but alone were not sufficient to fully resolve relationships. By directly
774 testing for reticulation and ILS, we uncovered the extent to which *Packera's* evolutionary
775 history has been shaped by hybridization and gene flow. These findings underscore the
776 broader challenge of capturing evolutionary complexity in phylogenomic datasets,
777 particularly when orthology inference is itself a major source of uncertainty. Within this
778 context, the tradeoff between excluding paralogs to reduce noise and retaining them to
779 preserve evolutionary signal becomes especially pronounced in polyploid systems.
780 The results of this study exemplifies how critical tree topology is for
781 macroevolutionary analyses and how paralog processing shapes not only the
782 phylogenetic relationships but the biological inferences we make from them. These
783 findings accentuate a central issue in phylogenomic and systematic studies, which is
784 that macroevolutionary analyses are highly sensitive to orthology inference strategies
785 and are only as reliable as the phylogenetic framework they are built on, and that
786 framework is deeply shaped by the methods used to handle paralogy. Looking ahead,
787 improving phylogeny construction will require both technical innovation and deeper
788 biological insight. Targeted probe sets designed to enrich truly single-copy genes, such

789 as Compositae-ParaLoss-1272 (Moore-Pollard et al. 2024), offers a promising strategy
790 for reducing paralog-related conflict. Yet, excluding paralogs entirely may discard
791 informative loci, while retaining them without appropriate modeling can mislead species
792 tree inference. As -omics technologies and phylogenomic methods become increasingly
793 sophisticated, the accurate interpretation of downstream analyses, such as divergence
794 dating, biogeography, trait evolution, and gene flow, still hinges on robust taxonomic
795 and systematic knowledge. In groups like *Packera*, where orthology inference can yield
796 conflicting results, understanding organismal biology is essential for evaluating the
797 plausibility of evolutionary scenarios. Morphology, ecology, and geography provide the
798 critical context needed to assess whether macroevolutionary patterns reflect true
799 evolutionary history or methodological artifacts.

800 Rather than treating paralogy as a problem to be avoided, our results highlight its
801 potential to illuminate processes like gene duplication, polyploidy, and reticulation. While
802 these challenges extend across many plant and non-plant groups, our densely sampled
803 *Packera* dataset offers a uniquely powerful tool for testing and improving phylogenomic
804 methods. As the field pushes toward more nuanced models of genome evolution,
805 systems like *Packera* will be essential for developing tools that can fully capture the
806 intricacy of complex evolutionary histories.

807 **Acknowledgements**

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

819

820 **Data availability**

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

824

825 **Author contributions**

826 ERMP designed the study, generated and analyzed data, and wrote the manuscript.
827 JRM provided guidance on the project. PE analyzed some of the data. JRM and PE
828 provided edits to the manuscript. All authors approved of the final version.

829

830 **Literature Cited**

831 Adeolu M., Alnajar S., Naushad S., S Gupta R. 2016. Genome-based phylogeny and
832 taxonomy of the “Enterobacteriales”: proposal for Enterobacterales ord. nov.
833 divided into the families Enterobacteriaceae, Erwiniaceae fam. nov.,
834 Pectobacteriaceae fam. nov., Yersiniaceae fam. nov., Hafniaceae fam. nov.,
835 Morganellaceae fam. nov., and Budviciaceae fam. nov. Int. J. Syst. Evol.
836 Microbiol. 66:5575–5599.

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

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

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

848 Barba-Montoya J. 2021. GBE Assessing Rapid Relaxed-Clock Methods for
849 Phylogenomic Dating. 13:1–14.

850 Barker M.S., Li Z., Kidder T.I., Reardon C.R., Lai Z., Oliveira L.O., Scascitelli M.,
851 Rieseberg L.H. 2016. Most compositae (Asteraceae) are descendants of a
852 paleohexaploid and all share a paleotetraploid ancestor with the calyceraceae.
853 Am. J. Bot. 103:1203–1211.

854 Barkley T.M. 1962. A Revision of Senecio aureus. Trans. Kans. Acad. Sci. 65:318–364.
855 Barkley T.M. 1988. Variation among the Aureoid Senecios of North America: A
856 geohistorical interpretation. Bot. Rev. 54:82–106.
857 Benton M.J., Donoghue P.C.J. 2007. Paleontological evidence to date the tree of life.
858 Mol. Biol. Evol. 24:26–53.
859 Bentz P.C., Leebens-Mack J. 2024. Developing Asparagaceae1726: An Asparagaceae-
860 specific probe set targeting 1726 loci for Hyb-Seq and phylogenomics in the
861 family. Appl. Plant Sci. 12:e11597.
862 Betrán E. 2015. The “life histories” of genes. J. Mol. Evol. 80:186–188.
863 Bezanson J., Chen J., Chung B., Karpinski S., Shah V.B., Vitek J., Zoubritzky L. 2018.
864 Julia: dynamism and performance reconciled by design. Proc. ACM Program.
865 Lang. 2:1–23.
866 Bolger A.M., Lohse M., Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina
867 sequence data. Bioinformatics. 30:2114–2120.
868 Braasch I., Postlethwait J.H. 2012. Polyploidy in Fish and the Teleost Genome
869 Duplication. Polyploidy and Genome Evolution. Springer Berlin Heidelberg. p.
870 341–383.
871 Brown J.W., Walker J.F., Smith S.A. 2017. Phyx: Phylogenetic tools for unix.
872 Bioinformatics. 33:1886–1888.
873 Costa F.P., Schrago C.G., Mello B. 2022. Accessing the relative performance of fast
874 molecular dating methods for phylogenomic data. :1–14.
875 Crameri S., Fior S., Zoller S., Widmer A. 2022. A target capture approach for
876 phylogenomic analyses at multiple evolutionary timescales in rosewoods
877 (Dalbergia spp.) and the legume family (Fabaceae). Mol. Ecol. Resour. 22:3087–

878 3105.
879 Danecek P., Bonfield J.K., Liddle J., Marshall J., Ohan V., Pollard M.O., Whitwham A.,
880 Keane T., McCarthy S.A., Davies R.M., Li H. 2021. Twelve years of SAMtools
881 and BCFtools. Gigascience. 10.
882 Degnan J.H., Rosenberg N.A. 2009. Gene tree discordance, phylogenetic inference and
883 the multispecies coalescent. Trends Ecol. Evol. 24:332–340.
884 Di Tommaso P., Chatzou M., Floden E.W., Barja P.P., Palumbo E., Notredame C. 2017.
885 Nextflow enables reproducible computational workflows. Nat. Biotechnol.
886 35:316–319.
887 Doyle J.J. 1992. Gene trees and species trees: Molecular systematics as one-character
888 taxonomy. Syst. Bot. 17:144.
889 Elven R., Murray D.F., Volodya Yu. Razzhivin B.A.Y. 2011. Annotated Checklist of the
890 Panarctic Flora (PAF) Vascular plants version 1.0. Available from
891 http://panarcticflora.org/results?biogeographic=&bioclimatic=&region=&name=pa
892 ckera+heterophylla#paf-863101.
893 Evans B.J., Kelley D.B., Tinsley R.C., Melnick D.J., Cannatella D.C. 2004. A
894 mitochondrial DNA phylogeny of African clawed frogs: Phylogeography and
895 implications for polyploid evolution. Mol. Phylogenet. Evol. 33:197–213.
896 Faircloth B.C. 2016. PHYLUCE is a software package for the analysis of conserved
897 genomic loci. Bioinformatics. 32:786–788.
898 Freeman C.C., Barkley T.M. 1995. A synopsis of the genus Packera (Asteraceae:
899 Senecioneae) in Mexico. SIDA Contrib. Bot. 16:699–709.
900 Fu Z.X., Jiao B.H., Nie B., Zhang G.J., Gao T.G., Chen Z.D., Lu A.M., Kong H.Z., Wang
901 X.Q., Wang Y.Z., Zhou S.L., Zhang S.Z., Wang X.M., Liu Z.J., Wang Q.F., Li

902 J.H., Li D.Z., Yi T.S., Hong M.A., Soltis D.E., Soltis P.S., Li J.H., Fu C.X., Liu
903 Q.X. 2016. A comprehensive generic-level phylogeny of the sunflower family:
904 Implications for the systematics of Chinese Asteraceae. J. Syst. Evol. 54:416–
905 437.
906 Funk V.A. 1985. Phylogenetic Patterns and Hybridization. Ann. Mo. Bot. Gard. 72:681.
907 Gallardo M.H., Kausel G., Jiménez A., Bacquet C., González C. 2004. Biological
908 relevance of polyploidy : ecology to genomics Whole-genome duplications in
909 South American desert rodents (Octodontidae). Society.:443–451.
910 Galtier N., Daubin V. 2008. Dealing with incongruence in phylogenomic analyses.
911 Philos. Trans. R. Soc. Lond. B Biol. Sci. 363:4023–4029.
912 Geldmacher J., Van Den Bogaard P., Hoernle K., Schmincke H.U. 2000. The 40Ar/39Ar
913 age dating of the Madeira Archipelago and hotspot track (eastern North Atlantic).
914 Geochem. Geophys. Geosyst. 1.
915 Gramling A. 2006. A conservation assessment of Packera millefolium, a Southern
916 Appalachian Endemic. .
917 Grant V. 1971. Plant Speciation. New York: Columbia University Press.
918 Hejase H.A., Liu K.J. 2016. A scalability study of phylogenetic network inference
919 methods using empirical datasets and simulations involving a single reticulation.
920 BMC Bioinformatics. 17:422.
921 Hobbs C.R., Baldwin B.G. 2013. Asian origin and upslope migration of Hawaiian
922 Artemisia (Compositae-Anthemideae). J. Biogeogr. 40:442–454.
923 Huang C.H., Zhang C., Liu M., Hu Y., Gao T., Qi J., Ma H. 2016. Multiple
924 Polyploidization Events across Asteraceae with Two Nested Events in the Early
925 History Revealed by Nuclear Phylogenomics. Mol. Biol. Evol. 33:2820–2835.

926 Huson D.H., Scornavacca C. 2012. Dendroscope 3: an interactive tool for rooted
927 phylogenetic trees and networks. Syst. Biol. 61:1061–1067.
928 Jackson C., McLay T., Schmidt-Lebuhn A.N. 2023. hybpiper-nf and paragone-nf:
929 Containerization and additional options for target capture assembly and paralog
930 resolution. APPS.
931 Jin J.-J., Yu W.-B., Yang J.-B., Song Y., dePamphilis C.W., Yi T.-S., Li D.-Z. 2020.
932 GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of
933 organelle genomes. Genome Biol. 21:241.
934 Johnson M.G., Gardner E.M., Liu Y., Medina R., Goffinet B., Shaw A.J., Zerega N.J.C.,
935 Wickett N.J. 2016. HybPiper: Extracting coding sequence and introns for
936 phylogenetics from high-throughput sequencing reads using target enrichment.
937 Appl. Plant Sci. 4:1600016.
938 Jones K.E., Alfredo Reyes-Betancort J., Hiscock S.J., Carine M.A. 2014. Allopatric
939 diversification, multiple habitat shifts, and hybridization in the evolution of
940 Pericallis (Asteraceae), a Macaronesian endemic genus. Am. J. Bot. 101:637–
941 651.
942 Katoh K., Standley D.M. 2013. MAFFT multiple sequence alignment software version 7:
943 Improvements in performance and usability. Mol. Biol. Evol. 30:772–780.
944 Kück P., Longo G.C. 2014. FASconCAT-G: Extensive functions for multiple sequence
945 alignment preparations concerning phylogenetic studies. Front. Zool. 11:1–8.
946 Kumar S., Stecher G., Li M., Knyaz C., Tamura K. 2018. MEGA X: Molecular
947 evolutionary genetics analysis across computing platforms. Mol. Biol. Evol.
948 35:1547–1549.
949 Kumar S., Stecher G., Peterson D., Tamura K. 2012. MEGA-CC: Computing core of

950 molecular evolutionary genetics analysis program for automated and iterative
951 data analysis. Bioinformatics. 28:2685–2686.
952 Kumar S., Stecher G., Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics
953 Analysis Version 7.0 for Bigger Datasets. Mol. Biol. Evol. 33:1870–1874.
954 Kurtzer G.M., cclerget, Bauer M., Kaneshiro I., Trudgian D., Godlove D. 2021.
955 hpcng/singularity: Singularity 3.7.3. Zenodo.
956 Kurtzer G.M., Sochat V., Bauer M.W. 2017. Singularity: Scientific containers for mobility
957 of compute. PLoS One. 12:e0177459.
958 Landis M.J., Matzke N.J., Moore B.R., Huelsenbeck J.P. 2013. Bayesian analysis of
959 biogeography when the number of areas is large. Syst. Biol. 62:789–804.
960 Lanfear R., Calcott B., Ho S.Y.W., Guindon S. 2012. PartitionFinder: Combined
961 selection of partitioning schemes and substitution models for phylogenetic
962 analyses. Mol. Biol. Evol. 29:1695–1701.
963 Langmead B., Salzberg S.L. 2012. Fast gapped-read alignment with Bowtie 2. Nat.
964 Methods. 9:357–359.
965 Li H., Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler
966 transform. Bioinformatics. 25:1754–1760.
967 Linder C.R., Rieseberg L.H. 2004. Reconstructing patterns of reticulate evolution in
968 plants. Am. J. Bot. 91:1700–1708.
969 Maddison W.P. 1997. Gene trees in species trees. Systematic Bio. 46:523–536.
970 Mahoney A.M., Kowal R.R. 2008. Three new varieties of Packera paupercula
971 (Asteraceae, Senecioneae) in midwestern and southeastern North America.
972 Novon St. Louis Mo. 18:220–228.
973 Mandel J.R., Dikow R.B., Funk V.A., Masalia R.R., Staton S.E., Kozik A., Michelmore

974 R.W., Rieseberg L.H., Burke J.M. 2014. A Target Enrichment Method for
975 Gathering Phylogenetic Information from Hundreds of Loci: An Example from the
976 Compositae. Appl. Plant Sci. 2:1300085.
977 Mandel J.R., Dikow R.B., Siniscalchi C.M., Thapa R., Watson L.E., Funk V.A. 2019. A
978 fully resolved backbone phylogeny reveals numerous dispersals and explosive
979 diversifications throughout the history of Asteraceae. Proc. Natl. Acad. Sci. U. S.
980 A. 116:14083–14088.
981 Masterson J. 1994. Stomatal Size. Science. 264:421–424.
982 Matzke N.J. 2014. Model selection in historical biogeography reveals that founder-event
983 speciation is a crucial process in island clades. Syst. Biol. 63:951–970.
984 McDougall I., Schmincke H.U. 1976. Geochronology of Gran Canaria, Canary Islands:
985 Age of shield building volcanism and other magmatic phases. Bulletin
986 Volcanologique. 40:57–77.
987 Mello B. 2018. Estimating timetrees with MEGA and the timetree resource. Mol. Biol.
988 Evol. 35:2334–2342.
989 Mendes F.K., Hahn Y., Hahn M.W. 2016. Gene tree discordance can generate patterns
990 of diminishing convergence over time. Mol. Biol. Evol. 33:3299–3307.
991 Minh B.Q., Schmidt H.A., Chernomor O., Schrempf D., Woodhams M.D., Von Haeseler
992 A., Lanfear R., Teeling E. 2020. IQ-TREE 2: New Models and Efficient Methods
993 for Phylogenetic Inference in the Genomic Era. Mol. Biol. Evol. 37:1530–1534.
994 Mitchell N., Lewis P.O., Lemmon E.M., Lemmon A.R., Holsinger K.E. 2017. Anchored
995 phylogenomics improves the resolution of evolutionary relationships in the rapid
996 radiation of Protea L. Am. J. Bot. 104:102–115.
997 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.

Ortiz E.M., Höwener A., Shigita G., Raza M., Maurin O., Zuntini A., Forest F., Baker W.J., Schaefer H. 2023. A novel phylogenomics pipeline reveals complex pattern of reticulate evolution in Cucurbitales. bioRxiv.

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

1022 Palazzesi L., Pellicer J., Barreda V.D., Loeuille B., Mandel J.R., Pokorny L., Siniscalchi
1023 C.M., Tellería M.C., Leitch I.J., Hidalgo O. 2022. Asteraceae as a model system
1024 for evolutionary studies: from fossils to genomes. Bot. J. Linn. Soc. 200:143–164.
1025 Panchy N., Lehti-Shiu M., Shiu S.H. 2016. Evolution of gene duplication in plants. Plant
1026 Physiol. 171:2294–2316.
1027 Panero J.L. 2005. New combinations and infrafamilial taxa in the Asteraceae.
1028 Phytologia. 87:1–14.
1029 Paradis E., Schliep K. 2019. Ape 5.0: An environment for modern phylogenetics and
1030 evolutionary analyses in R. Bioinformatics. 35:526–528.
1031 Parham J.F., Donoghue P.C.J., Bell C.J., Calway T.D., Head J.J., Holroyd P.A., Inoue
1032 J.G., Irmis R.B., Joyce W.G., Ksepka D.T., Patané J.S.L., Smith N.D., Tarver
1033 J.E., Van Tuinen M., Yang Z., Angielczyk K.D., Greenwood J.M., Hipsley C.A.,
1034 Jacobs L., Makovicky P.J., Müller J., Smith K.T., Theodor J.M., Warnock R.C.M.,
1035 Benton M.J. 2012. Best practices for justifying fossil calibrations. Syst. Biol.
1036 61:346–359.
1037 Pease J.B., Brown J.W., Walker J.F., Hinchliff C.E., Smith S.A. 2018. Quartet Sampling
1038 distinguishes lack of support from conflicting support in the green plant tree of
1039 life. Am. J. Bot. 105:385–403.
1040 Peirson J.A., Reznicek A.A., Semple J.C. 2012. Polyploidy, infraspecific cytotype
1041 variation, and speciation in Goldenrods: The cytogeography of Solidago subsect.
1042 Humiles (Asteraceae) in North America. Taxon. 61:197–210.
1043 Pelser P.B., Kennedy A.H., Tepe E.J., Shidler J.B., Nordenstam B., Kadereit J.W.,
1044 Watson L.E. 2010. Patterns and causes of incongruence between plastid and
1045 nuclear Senecioneae (Asteraceae) phylogenies. Am. J. Bot. 97:856–873.

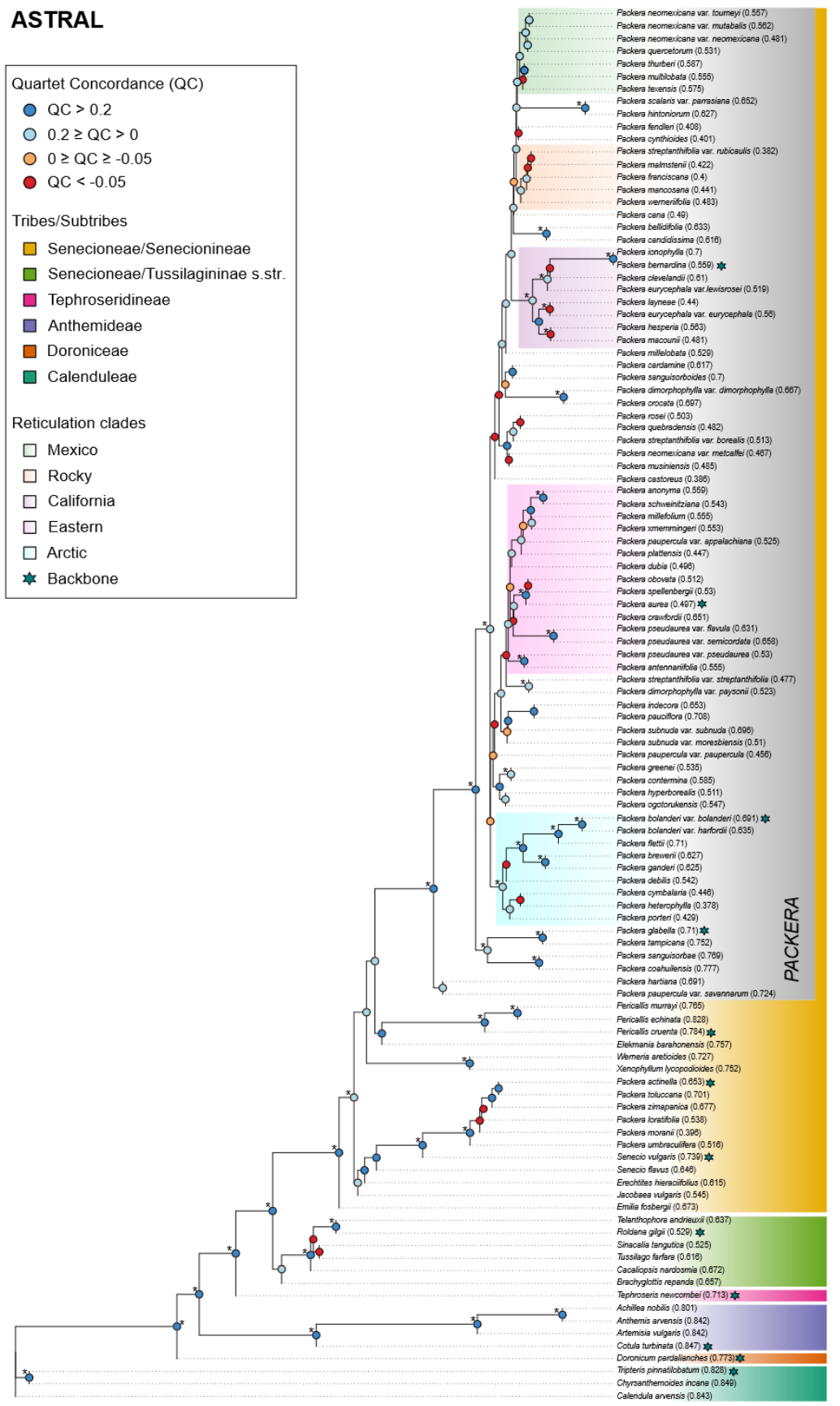
1046 R Core Team. 2016. R: A Language and Environment for Statistical Computing. Vienna:
1047 R Foundation for Statistical Computing.
1048 Ree R.H., Smith S.A. 2008. Maximum likelihood inference of geographic range
1049 evolution by dispersal, local extinction, and cladogenesis. Syst. Biol. 57:4–14.
1050 Revell L.J. 2012. phytools: An R package for phylogenetic comparative biology (and
1051 other things). Methods Ecol. Evol. 3:217–223.
1052 Ronquist F. 1997. Dispersal-vicariance analysis: A new approach to the quantification of
1053 historical biogeography. Syst. Biol. 46:195–203.
1054 Rose J.P., Toledo C.A.P., Moriarty Lemmon E., Lemmon A.R., Sytsma K.J. 2021. Out
1055 of Sight, Out of Mind: Widespread Nuclear and Plastid-Nuclear Discordance in
1056 the Flowering Plant Genus Polemonium (Polemoniaceae) Suggests Widespread
1057 Historical Gene Flow Despite Limited Nuclear Signal. Syst. Biology. 70:162–180.
1058 RStudio. 2020. RStudio: Integrated Development for R. Boston, MA: RStudio, PBC.
1059 Salmon A., Udall J.A., Jeddeloh J.A., Wendel J. 2012. Targeted capture of
1060 homoeologous coding and noncoding sequence in polyploid cotton. G3
1061 (Bethesda). 2:921–930.
1062 Sayyari E., Mirarab S. 2016. Fast Coalescent-Based Computation of Local Branch
1063 Support from Quartet Frequencies. Mol. Biol. Evol. 33:1654–1668.
1064 Schliep K.P. 2011. phangorn: phylogenetic analysis in R. Bioinformatics. 27:592–593.
1065 Shang H.-Y., Jia K.-H., Li N.-W., Zhou M.-J., Yang H., Tian X.-L., Ma Y.-P., Zhang R.-G.
1066 2025. Phytop: a tool for visualizing and recognizing signals of incomplete lineage
1067 sorting and hybridization using species trees output from ASTRAL. Hortic. Res.
1068 12:uhae330.
1069 Siu-Ting K., Torres-Sánchez M., San Mauro D., Wilcockson D., Wilkinson M., Pisani D.,

1070 O'Connell M.J., Creevey C.J. 2019. Inadvertent paralog inclusion drives
1071 artifactual topologies and timetree estimates in phylogenomics. Mol. Biol. Evol.
1072 36:1344–1356.
1073 Smith S.A., Moore M.J., Brown J.W., Yang Y. 2015. Analysis of phylogenomic datasets
1074 reveals conflict, concordance, and gene duplications with examples from animals
1075 and plants. BMC Evol. Biol. 15:150.
1076 Smith S.A., O'Meara B.C. 2012. TreePL: Divergence time estimation using penalized
1077 likelihood for large phylogenies. Bioinformatics. 28:2689–2690.
1078 Solís-Lemus C., Ané C. 2016. Inferring Phylogenetic Networks with Maximum
1079 Pseudolikelihood under Incomplete Lineage Sorting. PLoS Genet. 12:1–21.
1080 Solís-Lemus C., Bastide P., Ané C. 2017. PhyloNetworks: A package for phylogenetic
1081 networks. Mol. Biol. Evol. 34:3292–3298.
1082 Soltis D.E., Soltis P.S. 1999. Polyploidy: Recurrent formation and genome evolution.
1083 Trends Ecol. Evol. 14:348–352.
1084 Soltis P.S., Soltis D.E. 2009. The role of hybridization in plant speciation. Annu. Rev.
1085 Plant Biol. 60:561–588.
1086 Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-
1087 analysis of large phylogenies. Bioinformatics. 30:1312–1313.
1088 Steel M.A., Penny D. 1993. Distributions of tree comparison metrics--some new results.
1089 Syst. Biol. 42:126–141.
1090 Steenwyk J.L., Buida T.J., Labella A.L., Li Y., Shen X.-X., Rokas A. 2021. PhyKIT: a
1091 broadly applicable UNIX shell toolkit for processing and analyzing phylogenomic
1092 data. Bioinformatics. 37:2325–2331.
1093 Struck T.H. 2013. The impact of paralogy on phylogenomic studies - a case study on

1094 annelid relationships. PLoS One. 8:e62892.
1095 Stull G.W., Soltis P.S., Soltis D.E., Gitzendanner M.A., Smith S.A. 2020. Nuclear
1096 phylogenomic analyses of asterids conflict with plastome trees and support novel
1097 relationships among major lineages. Am. J. Bot. 107:790–805.
1098 Takhtajan A. 1986. Floristic Regions of the World. Berkley and Los Angeles, California:
1099 University of California Press.
1100 Tamura K., Battistuzzi F.U., Billing-Ross P., Murillo O., Filipski A., Kumar S. 2012.
1101 Estimating divergence times in large molecular phylogenies. Proc. Natl. Acad.
1102 Sci. U. S. A. 109:19333–19338.
1103 Tamura K., Tao Q., Kumar S. 2018. Theoretical foundation of the RelTime method for
1104 estimating divergence times from variable evolutionary rates. Mol. Biol. Evol.
1105 35:1770–1782.
1106 Tao Q., Tamura K., Kumar S. 2020a. The Molecular Evolutionary Clock. The Molecular
1107 Evolutionary Clock.
1108 Tao Q., Tamura K., Mello B., Kumar S. 2020b. Reliable Confidence Intervals for
1109 RelTime Estimates of Evolutionary Divergence Times. Mol. Biol. Evol. 37:280–
1110 290.
1111 Thapa R., Bayer R.J., Mandel J.R. 2020. Phylogenomics Resolves the Relationships
1112 within Antennaria (Asteraceae, Gnaphalieae) and Yields New Insights into its
1113 Morphological Character Evolution and Biogeography. Syst. Bot. 45:387–402.
1114 Tillich M., Lehwark P., Pellizzer T., Ulbricht-Jones E.S., Fischer A., Bock R., Greiner S.
1115 2017. GeSeq - Versatile and accurate annotation of organelle genomes. Nucleic
1116 Acids Res. 45:W6–W11.
1117 Tricou T., Tannier E., de Vienne D.M. 2022. Ghost lineages highly influence the

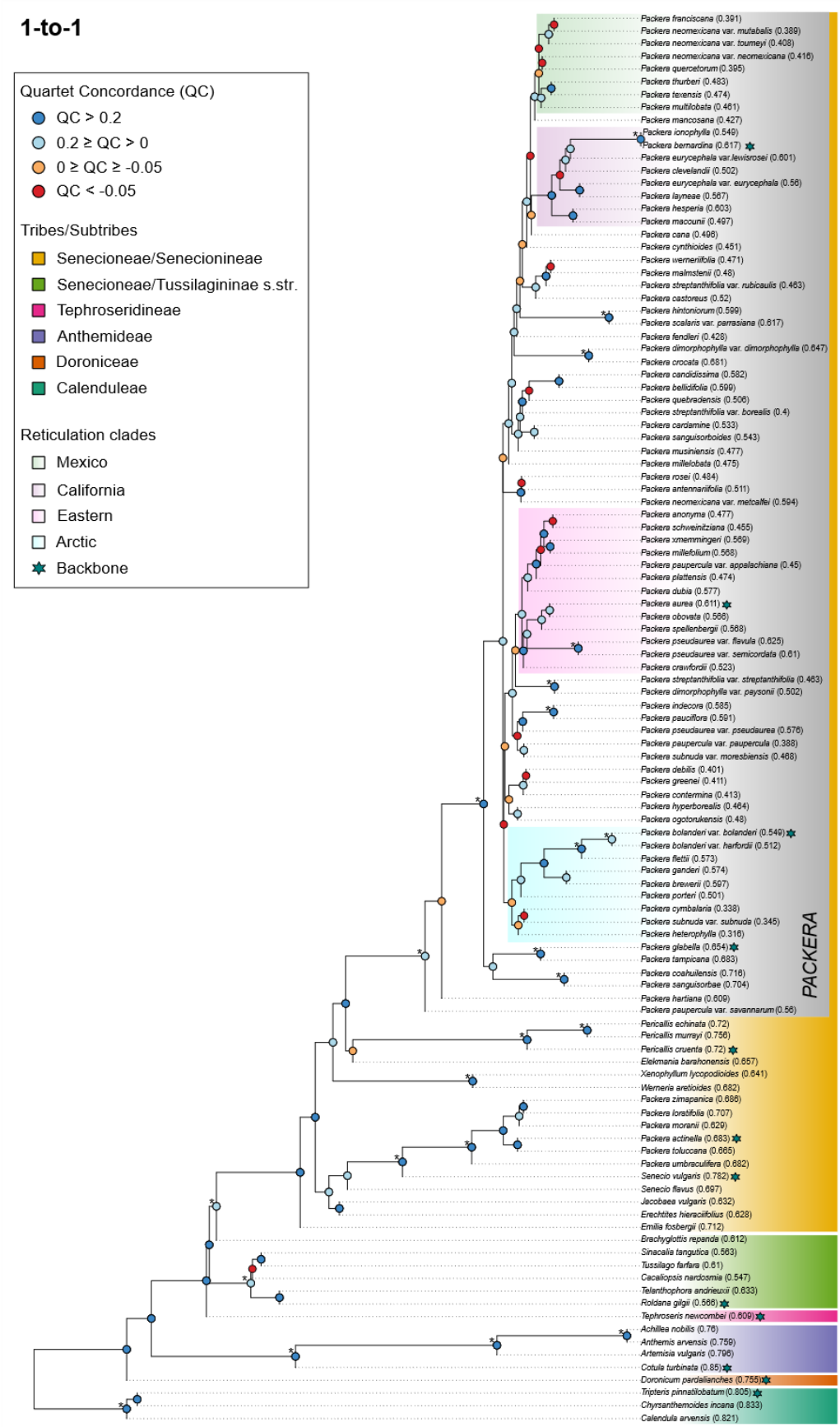
1118 interpretation of introgression tests. Syst. Biol. 71:1147–1158.
1119 Trock D.K. 1999. A revisionary synthesis of the genus Packera (Asteraceae:
1120 Senecioneae). .
1121 Trock D.K. 2006. Packera Á. Löve & D. Löve. Flora of North America Editorial
1122 Committee, eds. 1993+. New York and Oxford: p. 570–602.
1123 Vision T.J., Brown D.G., Tanksley S.D. 2000. The origins of genomic duplications in
1124 Arabidopsis. Science. 290:2114–2117.
1125 Wang W.M. 2004. On the origin and development of Artemisia (Asteraceae) in the
1126 geological past. Bot. J. Linn. Soc. 145:331–336.
1127 Wendel J.F., Cronn R.C. 2003. Polyploidy and the Evolutionary History of Cotton.
1128 Botany Publication and Papers. 23:139–186.
1129 Yang Y., Smith S.A. 2014. Orthology inference in nonmodel organisms using
1130 transcriptomes and low-coverage genomes: Improving accuracy and matrix
1131 occupancy for phylogenomics. Mol. Biol. Evol. 31:3081–3092.
1132 Yeatts L., Schneider A., Schneider B. 2011. Packera mancosana (Asteraceae:
1133 Senecioneae), a new species and shale barren endemic of southwestern
1134 colorado. Phytoneuron.:1–8.
1135 Yu G., Smith D.K., Zhu H., Guan Y., Lam T.T.-Y. 2017. Ggtree: An r package for
1136 visualization and annotation of phylogenetic trees with their covariates and other
1137 associated data. Methods Ecol. Evol. 8:28–36.
1138 Zhang C., Huang C.-H., Liu M., Hu Y., Panero J.L., Luebert F., Gao T., Ma H. 2021.
1139 Phylotranscriptomic insight into Asteraceae diversity, polyploidy, and
1140 morphological innovation. Journal of Integrative Plant Biolog. 63:1273–1293.
1141 Zhang C., Mirarab S. 2022a. ASTRAL-Pro 2: ultrafast species tree reconstruction from

1142 multi-copy gene family trees. Bioinformatics. 38:4949–4950.
1143 Zhang C., Mirarab S. 2022b. Weighting by gene tree uncertainty improves accuracy of
1144 quartet-based species trees. Mol. Biol. Evol. 39.
1145 Zhang C., Rabiee M., Sayyari E., Mirarab S. 2018. ASTRAL-III: Polynomial time species
1146 tree reconstruction from partially resolved gene trees. BMC Bioinformatics.
1147 19:15–30.
1148 Zhang C., Scornavacca C., Molloy E.K., Mirarab S. 2020. ASTRAL-pro: Quartet-based
1149 species-tree inference despite paralogy. Mol. Biol. Evol. 37:3292–3307.
1150


ASTRAL
Quartet Concordance (QC)
QC > 0.2
0.2 ≥ QC > 0
0 ≥ QC ≥ -0.05
QC < -0.05
Tribes/Subtribes
Senecioneae/Senecionineae
Senecioneae/Tussilagininae s.str.
Tephroseridineae
Anthemideae
Doroniceae
Calenduleae
Reticulation clades
Mexico
Rocky
California
Eastern
Arctic
Backbone
Packera neomexicana var. tourneyi (0.557)
Packera neomexicana var. mutabalis (0.562)
Packera neomexicana var. neomexicana (0.481)
Packera quercetorum (0.531)
Packera thurberi (0.587)
Packera multilobata (0.555)
Packera texensis (0.575)
Packera scalaris var. parrasiana (0.652)
Packera hintoniorum (0.627)
Packera fendleri (0.408)
Packera cynthioides (0.401)
Packera streptanthifolia var. rubicaulis (0.382)
Packera malmstenii (0.422)
Packera franciscana (0.4)
Packera mancosana (0.441)
Packera werneriifolia (0.483)
Packera cana (0.49)
Packera bellidifolia (0.633)
Packera candidissima (0.616)
Packera ionophylla (0.7)
Packera bernardina (0.559)
Packera clevelandii (0.61)
Packera eurycephala var.lewisrosei (0.519)
Packera layneae (0.44)
Packera eurycephala var. eurycephala (0.56)
Packera hesperia (0.563)
Packera macounii (0.481)
Packera millelobata (0.529)
Packera cardamine (0.617)
Packera sanguisorboides (0.7)
Packera dimorphophylla var. dimorphophylla (0.667)
Packera crocata (0.697)
Packera rosei (0.503)
Packera quebradensis (0.482)
Packera streptanthifolia var. borealis (0.513)
Packera neomexicana var. metcalfei (0.467)
Packera musiniensis (0.485)
Packera castoreus (0.386)
Packera anonyma (0.559)
Packera schweinitziana (0.543)
Packera millefolium (0.555)
Packera xmemmingeri (0.553)
Packera paupercula var. appalachiana (0.525)
Packera plattensis (0.447)
Packera dubia (0.496)
Packera obovata (0.512)
Packera spellenbergii (0.53)
Packera aurea (0.497)
Packera crawfordii (0.651)
Packera pseudaurea var. flavula (0.631)
Packera pseudaurea var. semicordata (0.658)
Packera pseudaurea var. pseudaurea (0.53)
Packera antennariifolia (0.555)
Packera streptanthifolia var. streptanthifolia (0.477)
Packera dimorphophylla var. paysonii (0.523)
Packera indecora (0.653)
Packera pauciflora (0.708)
Packera subnuda var. subnuda (0.696)
Packera subnuda var. moresbiensis (0.51)
Packera paupercula var. paupercula (0.456)
Packera greenei (0.535)
Packera contermina (0.585)
Packera hyperborealis (0.511)
Packera ogotorukensis (0.547)
Packera bolanderi var. bolanderi (0.691)
Packera bolanderi var. harfordii (0.635)
Packera flettii (0.71)
Packera breweri (0.627)
Packera ganderi (0.625)
Packera debilis (0.542)
Packera cymbalaria (0.446)
Packera heterophylla (0.378)
Packera porteri (0.429)
Packera glabella (0.71)
Packera tampicana (0.752)
Packera sanguisorbae (0.769)
Packera coahuilensis (0.777)
Packera hartiana (0.691)
Packera paupercula var. savannarum (0.724)
PACKERA
Pericaillis murrayi (0.765)
Pericaillis echinata (0.828)
Pericaillis cruenta (0.784)
Elekmania barahonensis (0.757)
Werneria aretioides (0.727)
Xenophyllum lycopodioides (0.752)
Packera actinella (0.653)
Packera toluccana (0.701)
Packera zimapanica (0.677)
Packera loratifolia (0.538)
Packera moranii (0.396)
Packera umbraculifera (0.516)
Senecio vulgaris (0.739)
Senecio flavus (0.646)
Erechtites hieraciifolius (0.615)
Jacobaea vulgaris (0.545)
Emilia fosbergii (0.673)
Telanthophora andrieuxii (0.637)
Roldana gilgii (0.529)
Sinacalia tangutica (0.525)
Tussilago farfara (0.616)
Cacaliopsis nardosmia (0.672)
Brachyglottis repanda (0.657)
Tephroseris newcombei (0.713)
Achillea nobilis (0.801)
Anthemis arvensis (0.842)
Artemisia vulgaris (0.842)
Cotula turbinata (0.847)
Doronicum pardalianches (0.773)
Tripteris pinnatilobatum (0.828)
Chyrsanthemoides incana (0.849)
Calendula arvensis (0.843)

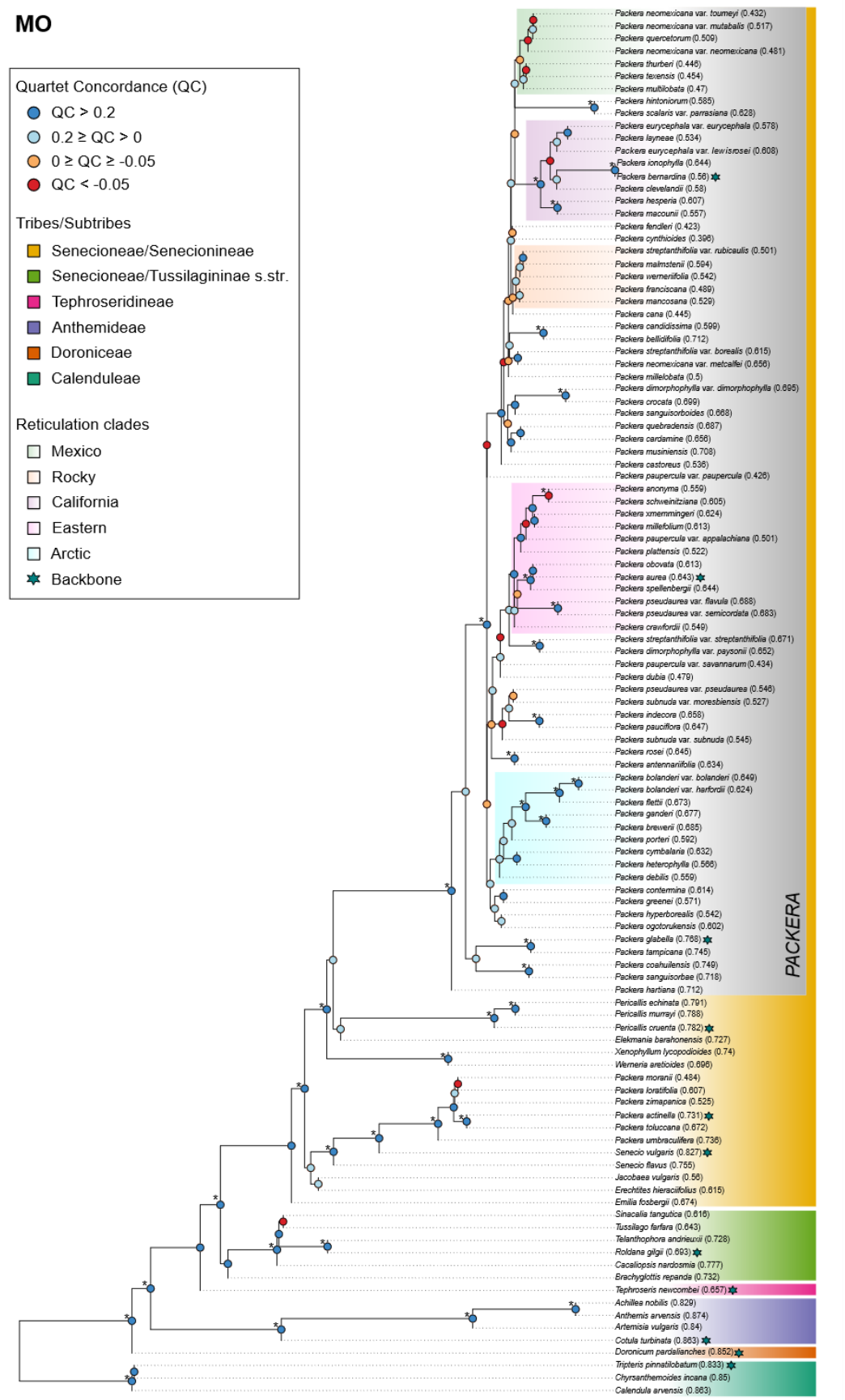
1151

1152 **Figure 1.** Full phylogeny of the ASTRAL analysis. Tribal/subtribal status is indicated on
1153 the right of the phylogeny and follows the color code as indicated in the key. Numbers
1154 next to tip labels correspond to the Quartet Fidelity (QF) score from Quartet Sampling,
1155 which ranges from 0 to 1, with values closer to 0 indicating erroneous taxa. Circles at
1156 nodes show the Quartet Concordance (QC) factors from Quartet Sampling, where red
1157 and orange indicate counter-support nodes. Asterisks next to QC circles indicate high
1158 support (≥95 LPP) at that node. Highlighted clades, and teal stars next to some tip
1159 labels, indicate clades where reticulation events were tested, following the color key on
1160 the left.
1161


1-to-1
Quartet Concordance (QC)
QC > 0.2
0.2 ≥ QC > 0
0 ≥ QC ≥ -0.05
QC < -0.05
Tribes/Subtribes
Senecioneae/Senecionineae
Senecioneae/Tussilagininae s.str.
Tephroseridineae
Anthemideae
Doroniceae
Calenduleae
Reticulation clades
Mexico
California
Eastern
Arctic
Backbone
Packera franciscana (0.391)
Packera neomexicana var. mutabalis (0.389)
Packera neomexicana var. tournayi (0.408)
Packera neomexicana var. neomexicana (0.416)
Packera quercetorum (0.395)
Packera thurberi (0.483)
Packera texensis (0.474)
Packera multilobata (0.461)
Packera mancosana (0.427)
Packera ionophylla (0.549)
Packera bernardina (0.617)
Packera eurycephala var.lewisrosei (0.601)
Packera clevelandii (0.502)
Packera eurycephala var. eurycephala (0.56)
Packera layneae (0.567)
Packera hesperia (0.603)
Packera macounii (0.497)
Packera cana (0.496)
Packera cynthioides (0.451)
Packera werneriifolia (0.471)
Packera malmstenii (0.48)
Packera streptanthifolia var. rubicaulis (0.463)
Packera castoreus (0.52)
Packera hintoniorum (0.599)
Packera scalaris var. parrasiana (0.617)
Packera fendleri (0.428)
Packera dimorphophylla var. dimorphophylla (0.647)
Packera crocata (0.681)
Packera candidissima (0.582)
Packera bellidifolia (0.599)
Packera quebradensis (0.506)
Packera streptanthifolia var. borealis (0.4)
Packera cardamine (0.533)
Packera sanguisorboides (0.543)
Packera musiniensis (0.477)
Packera millelobata (0.475)
Packera rosei (0.484)
Packera antennariifolia (0.511)
Packera neomexicana var. metcalfei (0.594)
Packera anonyma (0.477)
Packera schweinitziana (0.455)
Packera xmemmingeri (0.569)
Packera millefolium (0.568)
Packera paupercula var. appalachiana (0.45)
Packera plattensis (0.474)
Packera dubia (0.577)
Packera aurea (0.611)
Packera obovata (0.566)
Packera spellenbergii (0.568)
Packera pseudaurea var. flavula (0.625)
Packera pseudaurea var. semicordata (0.61)
Packera crawfordii (0.523)
Packera streptanthifolia var. streptanthifolia (0.463)
Packera dimorphophylla var. paysonii (0.502)
Packera indecora (0.585)
Packera pauciflora (0.591)
Packera pseudaurea var. pseudaurea (0.576)
Packera paupercula var. paupercula (0.388)
Packera subnuda var. moresbiensis (0.468)
Packera debilis (0.401)
Packera greenei (0.411)
Packera contermina (0.413)
Packera hyperborealis (0.464)
Packera ogotorukensis (0.48)
Packera bolanderi var. bolanderi (0.549)
Packera bolanderi var. harfordii (0.512)
Packera flettii (0.573)
Packera ganderi (0.574)
Packera breweri (0.597)
Packera porteri (0.501)
Packera cymbalaria (0.338)
Packera subnuda var. subnuda (0.345)
Packera heterophylla (0.316)
Packera glabella (0.654)
Packera tampicana (0.683)
Packera coahuilensis (0.716)
Packera sanguisorbae (0.704)
Packera hartiana (0.609)
Packera paupercula var. savannarum (0.56)
PACKERA
Pericallis echinata (0.72)
Pericallis murrayi (0.756)
Pericallis cruenta (0.72)
Elekmania barahonensis (0.657)
Xenophyllum lycopodioides (0.641)
Werneria aretioides (0.682)
Packera zimapanica (0.686)
Packera loratifolia (0.707)
Packera moranii (0.629)
Packera actinella (0.683)
Packera toluccana (0.665)
Packera umbraculifera (0.682)
Senecio vulgaris (0.782)
Senecio flavus (0.697)
Jacobaea vulgaris (0.632)
Erechtites hieraciifolius (0.628)
Emilia fosbergii (0.712)
Brachyglottis repanda (0.612)
Sinacalia tangutica (0.563)
Tussilago farfara (0.61)
Cacaliopsis nardosmia (0.547)
Telanthophora andrieuxii (0.633)
Roldana gilgii (0.566)
Tephroseris newcombei (0.609)
Achillea nobilis (0.76)
Anthemis arvensis (0.759)
Artemisia vulgaris (0.796)
Cotula turbinata (0.85)
Doronicum pardalianches (0.755)
Tripteris pinnatilobatum (0.805)
Chrysanthemoides incana (0.833)
Calendula arvensis (0.821)

1162

Figure 2. Full phylogeny of the paragone-nf: 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.


MO
Quartet Concordance (QC)
QC > 0.2
0.2 ≥ QC > 0
0 ≥ QC ≥ -0.05
QC < -0.05
Tribes/Subtribes
Senecioneae/Senecionineae
Senecioneae/Tussilagininae s.str.
Tephroseridineae
Anthemideae
Doroniceae
Calenduleae
Reticulation clades
Mexico
Rocky
California
Eastern
Arctic
Backbone
Packera neomexicana var. toumeyi (0.432)
Packera neomexicana var. mutabalis (0.517)
Packera quercetorum (0.509)
Packera neomexicana var. neomexicana (0.481)
Packera thurberi (0.446)
Packera texensis (0.454)
Packera multilobata (0.47)
Packera hintoniorum (0.585)
Packera scalaris var. parrasiana (0.628)
Packera eurycephala var. eurycephala (0.578)
Packera layneae (0.534)
Packera eurycephala var. lewisrosei (0.608)
Packera ionophylla (0.644)
Packera bernardina (0.56)
Packera clevelandii (0.58)
Packera hesperia (0.607)
Packera macounii (0.557)
Packera fendleri (0.423)
Packera cynthioides (0.396)
Packera streptanthifolia var. rubicaulis (0.501)
Packera malmstenii (0.594)
Packera werneriifolia (0.542)
Packera franciscana (0.489)
Packera mancosana (0.529)
Packera cana (0.445)
Packera candidissima (0.599)
Packera bellidifolia (0.712)
Packera streptanthifolia var. borealis (0.615)
Packera neomexicana var. metcalfei (0.656)
Packera millelobata (0.5)
Packera dimorphophylla var. dimorphophylla (0.695)
Packera crocata (0.699)
Packera sanguisorboides (0.668)
Packera quebradensis (0.687)
Packera cardamine (0.656)
Packera musiniensis (0.708)
Packera castoreus (0.536)
Packera paupercula var. paupercula (0.426)
Packera anonyma (0.559)
Packera schweinitziana (0.605)
Packera xmemmingeri (0.624)
Packera millefolium (0.613)
Packera paupercula var. appalachiana (0.501)
Packera plattensis (0.522)
Packera obovata (0.613)
Packera aurea (0.643)
Packera spellenbergii (0.644)
Packera pseudaurea var. flavula (0.688)
Packera pseudaurea var. semicordata (0.683)
Packera crawfordii (0.549)
Packera streptanthifolia var. streptanthifolia (0.671)
Packera dimorphophylla var. paysonii (0.652)
Packera paupercula var. savannarum (0.434)
Packera dubia (0.479)
Packera pseudaurea var. pseudaurea (0.546)
Packera subnuda var. moresbiensis (0.527)
Packera indecora (0.658)
Packera pauciflora (0.647)
Packera subnuda var. subnuda (0.545)
Packera rosei (0.645)
Packera antennariifolia (0.634)
Packera bolanderi var. bolanderi (0.649)
Packera bolanderi var. harfordii (0.624)
Packera flettii (0.673)
Packera ganderi (0.677)
Packera breweri (0.685)
Packera porteri (0.592)
Packera cymbalaria (0.632)
Packera heterophylla (0.566)
Packera debilis (0.559)
Packera contermina (0.614)
Packera greenei (0.571)
Packera hyperborealis (0.542)
Packera ogotorukensis (0.602)
Packera glabella (0.768)
Packera tampicana (0.745)
Packera coahuilensis (0.749)
Packera sanguisorbae (0.718)
Packera hartiana (0.712)
PACKERA
Pericallis echinata (0.791)
Pericallis murrayi (0.788)
Pericallis cruenta (0.782)
Elekmania barahonensis (0.727)
Xenophyllum lycopodioides (0.74)
Werneria aretioides (0.696)
Packera moranii (0.484)
Packera loratifolia (0.607)
Packera zimapanica (0.525)
Packera actinella (0.731)
Packera toluccana (0.672)
Packera umbraculifera (0.736)
Senecio vulgaris (0.827)
Senecio flavus (0.755)
Jacobaea vulgaris (0.56)
Erechtites hieraciifolius (0.615)
Emilia fosbergii (0.674)
Sinacalia tangutica (0.616)
Tussilago farfara (0.643)
Telanthophora andrieuxii (0.728)
Roldana gilgii (0.693)
Cacaliopsis nardosmia (0.777)
Brachyglottis repanda (0.732)
Tephroseris newcombei (0.657)
Achillea nobilis (0.829)
Anthemis arvensis (0.874)
Artemisia vulgaris (0.84)
Cotula turbinata (0.863)
Doronicum pardalianches (0.852)
Tripteris pinnatilobatum (0.833)
Chrysanthemoides incana (0.85)
Calendula arvensis (0.863)

1173

1174 **Figure 3.** Full phylogeny of the paragone-nf: MO analysis. Tribal/subtribal status is
1175 indicated on the right of the phylogeny and follows the color code as indicated in the
1176 key. Numbers next to tip labels correspond to the Quartet Fidelity (QF) score from
1177 Quartet Sampling, which ranges from 0 to 1, with values closer to 0 indicating erroneous
1178 taxa. Circles at nodes show the Quartet Concordance (QC) factors from Quartet
1179 Sampling, where red and orange indicate counter-support nodes. Asterisks next to QC
1180 circles indicate high support (≥95 LPP) at that node. Highlighted clades, and teal stars
1181 next to some tip labels, indicate clades where reticulation events were tested, following
1182 the color key on the left.
1183

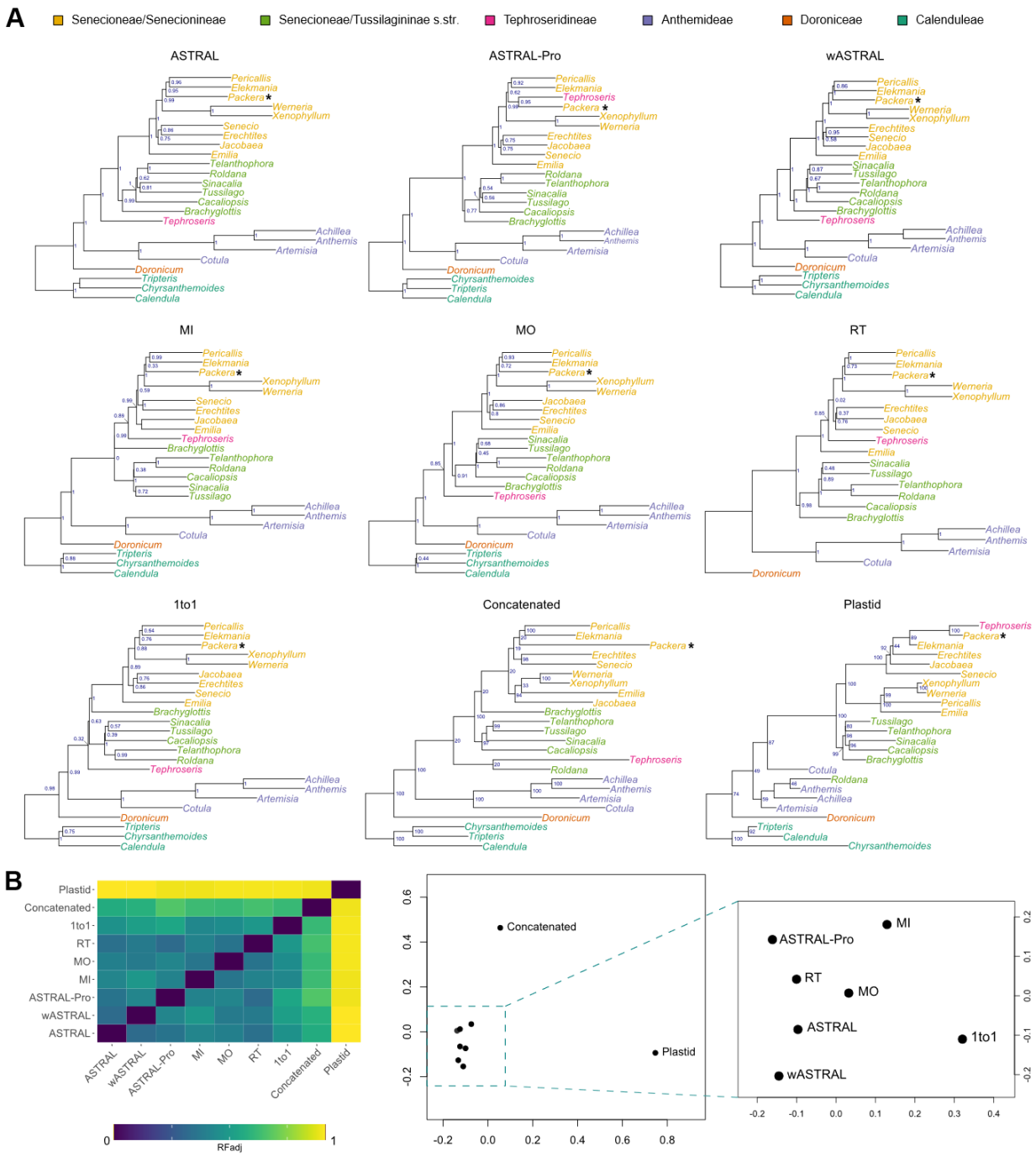


Figure 4. Genus-level phylogenies of all trees generated in this study. Genera are colored according to their tribal/subtribal classifications following the key on the top of the figure. Packera is marked with a black asterisk next to its placement in the trees. Navy blue text at node represents support values (LPP or BS). Bottom left is a matrix of

1189 RF_{adj} values between the nine different trees, with most similar topologies as dark blue
1190 squares (0) and least similar as yellow squares (1). Bottom right are the RF_{adj} values
1191 plotted as a principal components analysis. Points closer together indicate higher
1192 similarity, while dots further away indicate higher dissimilarity.
1193

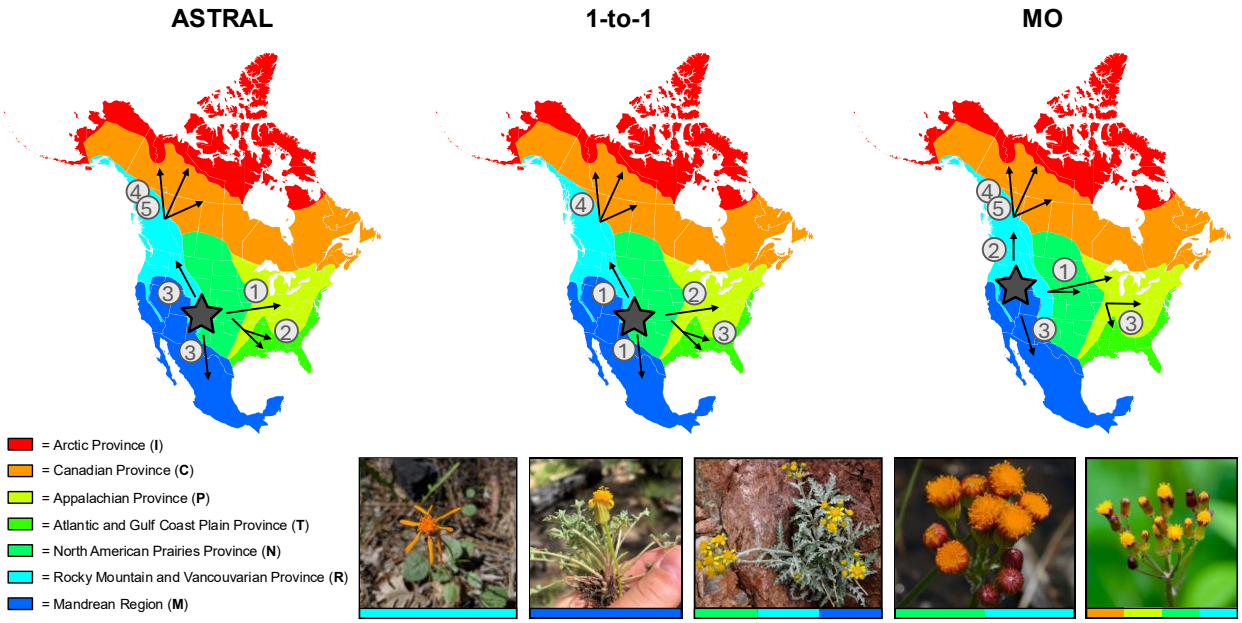


Figure 5. Predicted biogeographic region summaries generated from BioGeoBEARS of the HybPiper, paragone-nf: 1-to-1, and paragone-nf: MO analyses. The maps of North America are colored according to the ten geographic regions outlined by Takhtajan (1986): Arctic Province (I), Canadian Province (C), Appalachian 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, 5 being the most recent major expansion. All BioGeoBears results can be found in Supplemental Figure I. 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 &

1210 Á.Löve; and Rayless mountain ragwort, *P. indecora* (Greene) Á.Löve & D.Löve. Image

1211 credits: Erika R. Moore-Pollard and Robert Lagier.

1212

1213

1214 **Figure 6.** PhyloNetworks figures of the three tested analyses: ASTRAL, 1-to-1, and MO.
1215 The various areas of the phylogeny that were tested is shown on the left and can be
1216 read as a row. Dark green edges indicate the major hybrid edge, while the light blue
1217 shows the minor hybrid edge. The associated numbers next to each branch
1218 corresponds to the percentage of parentage given by that branch. Number of
1219 hybridizations correspond with the best log likelihood value. The 1-to-1 analysis does
1220 not have a Rocky analysis since the clade in that tree only had four taxa.

1221 **Table S1.** Voucher specimens of the 111 taxa used in this study. Publication status and
1222 authorities assigned by IPNI. FNA = Flora of North America (2006).
1223 **Table S2.** Raw sequence statistics using the pxlssq function in *Phyx* for each of the 111
1224 samples in this study.
1225 **Table S3.** Matrix and tree statistics of the nine trees generated with 111 taxa.
1226 **Table S4.** Best fitting models for each gene from PartitionFinder given the overall run
1227 (General) and the ASTRAL-Pro run where the paralogs and single-copy orthologs were
1228 run separately. The three models tested were GTR, GTR+G, and GTR+I+G. Numbers
1229 in parenthesis next to the model type corresponds to the number of genes for that
1230 model.
1231 **Table S5.** Dating scenarios with various age constraints (A-E) implemented in RelTime.
1232 y = age constraint was used in the scenario; n = age constraint was not used in the
1233 scenario.
1234 **Table S6.** Presence/Absence values of each species given the ten geographic regions
1235 outlined by Takhtajan (1986). A value of "0" indicates absence, "1" indicates presence.
1236 Letters to the right of each column heading relates to Fig. 5. The four column headings
1237 above the geographic regions refer to the four groupings of *Packera* used in the
1238 discussion, no groupings were designated for Europe or Central America since they
1239 represent outgroup distributions.
1240 **Table S7.** Log likelihood values and taxon lists on the five PhyloNetworks runs–
1241 Backbone, Arctic/Alpine, California/Mexico, Eastern, and Rocky Mountains–across the
1242 ASTRAL, 1-to-1, and MO trees. The number of reticulations with the lowest log
1243 likelihood value, indicated by blue font, is considered the best fit.
1244 **Table S8.** Assembly statistics generated from the HybPiper pipeline.

1245 **Table S9.** Percent missing data of each plastid assembly generated.

1246 **Table S10.** Pairwise adjusted Robinson-Foulds (RFadj) distances between the nine
1247 main trees compared in this study. The values range from zero to one, indicating
1248 whether the tree topologies are identical or completely dissimilar, respectively.

1249 **Table S11.** Resulting divergence rates of crown *Packera*, Senecioneae, and
1250 Anthemideae from ASTRAL, MO, 1-to-1 using RelTime.

1251 **Table S12.** Biogeographical model selection for all *Packera* taxa and four outgroup
1252 genera (*Pericallis, Elekmania, Werneria, Xenophyllum*). Ln L = log likelihood values;
1253 Num params = number of free parameters; d = “dispersal” parameter; e = “extinction”
1254 parameter; j = “jump” parameter which specifies the weight of each jump dispersal event
1255 in the cladogenesis matrix; AIC = Akaike information criterion; AICc = Akaike
1256 information criterion with corrections for small sample sizes. “wt” represents the
1257 resulting model weights with AIC and AICc. Blue font indicates the chosen BioGeoBears
1258 analysis.

1259 **Table S13.** List of species in each of the five tested networks for the ASTRAL, 1-to-1,
1260 and MO trees. Colored cells indicate that the species was present in the network, no
1261 color indicates absence. Colors are according to the network shading from Figs. 1, 2,
1262 and 3.

1263 **Figure S1.** Annotated plastids of 18 taxa with complete plastomes.

1264 **Figure S2.** Full chloroplast tree of 111 mapped plastid data. Colors at the end of tips
1265 corresponds to the percentage of missing data for each taxon.

1266 **Figure S3.** All tested trees containing 111 taxa (excluding the complete plastid tree).
1267 Scale bar under the tree on the left represents the mean number of nucleotide
1268 substitutions per site. Support values–bootstraps (BS, ranging from 0-100) or local

1269 posterior probabilities (LPP, ranging from 0-1)–are indicated next to the node. Species
1270 tree inference methods are detailed in Table 1. Phylogenies generated using phytools in
1271 R.
1272 **Figure S4.** Results of Quartet Sampling on the different, tested trees. Values at each
1273 node correspond to the Quartet Concordance (QC), Quartet Differential (QD), and
1274 Quartet Informativeness (QI) values (QC/QD/QI). Colored circles at tips correspond to
1275 the Quartet Fidelity (QF) score.
1276 **Figure S5.** PhyParts results of the different phylogenies. Numbers above the branch
1277 indicate the number of concordant gene trees, while the numbers below the branch
1278 indicate the number of conflicting gene trees. Blue: concordant gene trees, green: top
1279 alternative bipartition, red: all other alternative bipartitions; gray: uninformative or
1280 missing data.
1281 **Figure S6.** PhyTop results.
1282 **Figure S7.** Divergence rate estimates of Scenarios 1-6 using RelTime for the ASTRAL,
1283 1-to-1, and MO trees in millions of years. Left tree contains the mean age of that
1284 designated node, the right tree shows the 95% age range for the designated node.
1285 **Figure S8.** Predicted biogeographic regions generated from BioGeoBEARS on all
1286 *Packera* taxa and four outgroup genera (*Pericallis*, *Elekmania*, *Werneria*, and
1287 *Xenophyllum*) using all tested models for the ASTRAL, 1-to-1, and MO trees. Boxes and
1288 circles at nodes represent the geographic regions provided in Figure 5.
1289 **Figure S9.** Plots of loglikelihood values from Table S7 of each hybridization run (along
1290 the left y-axis) for each tree tested (along the top x-axis). The bottom x-axis indicates
1291 the 11 tested hybridization numbers (n=0-10). Colored points indicate the best fitting
1292 hybridization for that analysis.