Gene duplication dynamics and regulatory evolution shape the diversification of Asteraceae

Erika R. Moore-Pollard[1,2*‡], Paige A. Ellestad[1*], Brannan R. Cliver[3,4*], Jakub Baczyński[5,6], Matthew D. Pollard[1,7,8], Zachary Meharg[9,10], Andrew C. Willoughby[11], Samantha Drewry[1], Alex Harkess[10], J. Mauricio Bonifacino[12], Zachary L. Nimchuk[11,13], John M. Burke[5,14], Daniel S. Jones[3,4], Jennifer R. Mandel[1,15‡]

[1]*Department of Biological Sciences, University of Memphis, 3700 Walker Ave, Memphis, Tennessee 38152, USA*

[2]*Department of Ecology and Evolutionary Biology, University of Connecticut, 75 N Eagleville Rd, Unit 3043, Storrs, CT, 06269, USA*

[3]*Department of Biological Sciences, Clemson University, 132 Long Hall Clemson, SC 29634, USA*

[4]*Department of Biological Sciences, Auburn University, 101 Rouse Life Sciences, Auburn, Alabama 36849, USA*

[5]*Department of Plant Biology, Miller Plant Sciences, University of Georgia, Athens, Georgia 30602, USA*

[6]*Institute of Evolutionary Biology, Faculty of Biology, Biological and Chemical Research Centre, University of Warsaw, Żwirki i Wigury 101, 02-089 Warsaw, Poland*

[7]*Department of Systems Biology, UMass Chan Medical School, 368 Plantation Street, Worcester, Massachusetts 01605, USA*

[8]*Department of Genomics and Computational Biology, UMass Chan Medical School, 368 Plantation Street, Worcester, Massachusetts 01605, USA*

[9]*Department of Crop, Soil, and Environmental Sciences, Auburn University, Funchess Hall, 350 S College St Suite 201, Auburn, AL 36849, USA*

[10]*HudsonAlpha Institute for Biotechnology, Huntsville, Alabama 35806, USA*

[11]*Department of Biology, University of North Carolina at Chapel Hill, 250 Bell Tower Dr Chapel Hill, NC 27708, USA*

[12]*Laboratorio de Botánica, Facultad de Agronomía, Universidad de la República, Av. Garzón 12 780, Montevideo C.P. 12900, Uruguay*

[13]*Curriculum in Genetics and Molecular Biology, University of North Carolina at Chapel Hill, 250 Bell Tower Dr Chapel Hill, NC 27599, USA*

[14]*The Plant Center, University of Georgia, Athens, Georgia 30602, USA*

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

* These authors contributed equally to this work

‡ Corresponding authors

Authors' contact information:

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

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

Brannan R. Cliver, bcliver@clemson.edu, 0009-0002-4183-2789

Jakub Baczyński, jjb67768@uga.edu, 0000-0001-5272-9053

Matthew D. Pollard, matthew.pollard95@gmail.com, 0000-0002-5171-1211

Zach Meharg, zmeharg@hudsonalpha.org, 0009-0004-7382-8527

Andrew C. Willoughby, andrew.willoughby@duke.edu, 0000-0003-0355-1214

Samantha Drewry, sgdrewry@memphis.edu, 0009-0002-0278-6804

Alex Harkess, aharkess@hudsonalpha.org, 0000-0002-2035-0871

J. Mauricio Bonifacino, bonifacinoj@fagro.edu.uy, 0000-0003-0377-3853

Zachary L. Nimchuk, zackn@email.unc.edu, 0000-0003-1826-0671

John M. Burke, jmburke@uga.edu, 0000-0002-1412-5539

Daniel S. Jones, dsj2@clemson.edu, 0000-0002-9241-1813

‡ Jennifer Mandel, jmandel@memphis.edu, ORCID: 0000-0003-3539-2991

Abstract

The flowering plant order Asterales exhibits a striking disparity in species richness, with >30,000 species in Asteraceae compared to <50 in its sister family Calyceraceae. To investigate the genomic basis of this imbalance, we assembled three new chromosome-level genomes, including the first for Calyceraceae, and re-annotated five additional Asterales genomes. Comparative analyses revealed exceptionally high repeat content in both Asteraceae and Calyceraceae, pervasive chromosomal rearrangements, and evidence for shared and lineage-specific whole-genome duplications. In Asteraceae, tandem and dispersed duplications disproportionately drove expansions of gene families linked to secondary metabolism and stress response, while segmental duplicates bore signatures of adaptive selection for the regulation of biosynthetic and metabolic processes. Selective pressures on regulators of flowering time, but not on core flowering time genes, suggest an evolved balance between regulatory flexibility and developmental constraint in floral diversification. Overall, these patterns suggest that, beyond ancient polyploidy, small-scale duplications and selective fine-tuning of regulatory networks underpinned the ecological versatility in Asteraceae, fueling its extraordinary diversification.

Keywords: Asterales, comparative genomics, Compositae, episodic positive selection, gene family evolution, lineage-specific duplications, whole-genome duplication

Introduction

A recurring pattern in flowering plant evolution is the uneven distribution of species richness across the angiosperm tree of life, whereby species-poor lineages are often sister to highly diverse ones. While extinction and gaps in the fossil record may obscure some histories, the overall pattern remains clear[1]. Nowhere is this contrast more striking than in Asterales. The order includes ~35,800 species across 11 families, but >30,000 species belong to Asteraceae alone. Accounting for ~10% of all flowering plants and spanning nearly every continent and ecological niche (Fig. 1a), Asteraceae ranks among the most diverse and economically important plant families worldwide[2–8]. Over 2,000 of its members are used by humans for food, medicine, and ornamentals, yet it also ranks as the second most invasive plant family, exacting heavy ecological and economic costs[5,9]. In sharp contrast, its sister family, Calyceraceae, comprises fewer than 50 species confined to southern South America (Fig. 1a)[10–12]. This extreme disparity raises a central question: what genomic innovations fueled Asteraceae's extraordinary radiation?

Asteraceae boasts a long evolutionary history shaped by rapid radiation, polyploidization, hybridization, widespread ecological adaptation, and morphological innovation[3–7]. Its extensive diversification during the late Cretaceous has been largely attributed to enhanced reproductive success through pollinator attraction to its unique head-like inflorescence comprised of many individual florets, the capitulum[4,13,14]. Although Calyceraceae inflorescences also resemble capitula, important differences distinguish the two families, including the presence/absence of a terminal flower, fused/un-fused anthers, partially fused filaments attached to the corolla versus free filaments (Calyceraceae versus Asteraceae, respectively), and distinct mechanisms of pollen presentation[13,15,16]. Additionally, the production of diverse phytochemicals in Asteraceae facilitated ecological adaptation, particularly through defense responses[7]. These traits, coupled with a diversity of inflorescence types derived from distinct morphological and organizational modifications of the capitulum (Fig. 1b)[14], likely contributed to the disparity in species richness between the two families. The genomic mechanisms underlying their origin, however, have yet to be elucidated.

Acting across multiple scales, from individual genes to entire genomes, gene duplications have been a major driver of plant diversification by providing the raw genetic material for functional innovation[17]. Among these processes, whole-genome duplications (WGDs) are particularly pervasive throughout the evolutionary history of angiosperms and are widely recognized as a central force contributing to both their expansive diversity and widespread distribution[18–20]. The triplicate genome structure of Asteraceae suggests that the

family's diversification was catalyzed by two successive WGD events (i.e., paleohexaploidy) at its core which was then followed by lineage-specific WGDs in some of its most species-rich groups, e.g., tribe Cichorieae and the Heliantheae alliance[21–24]. Although limited by available genomic data, evidence also suggests that the sister family Calyceraceae experienced the same paleohexaploidy event[2–4]. This finding raises further questions about the role of mechanisms beyond WGDs in plant diversification. Resolving the precise phylogenetic placement of these events within Asterales is key to understanding the enormous evolutionary success of Asteraceae.

Beyond WGDs, transposable elements (TEs) represent another dynamic force in plant diversification, driving genomic change through chromosomal rearrangements and altering gene expression via insertions or chromatin remodeling[25–27]. Asteraceae genomes, notable for their exceptionally high repeat content dominated by TEs, highlight the likely role of TE activity in promoting lineage-specific adaptations to diverse environments[28]. While both WGDs and TE-mediated small-scale duplications have profoundly shaped plant evolution, their relative contributions to diversification remain unresolved.

Despite the evolutionary and economic importance of Asteraceae, our understanding of genome evolution in the family, and in the broader Asterales order, remains limited due to sparse genomic and transcriptomic resources, beyond a handful of well-studied species (e.g., *Helianthus* [29] and *Lactuca* [24]). To date, assembled genomes are mostly concentrated to the highly derived clades of Asteraceae with no representation of the early South American lineages[30]. Similarly, the small sister family, Calyceraceae, lacks genomic data, leaving a major gap in our understanding of the origin and early evolution of Asteraceae.

To investigate the genomic drivers of diversification across Asterales, we generated three new chromosome-level assemblies of *Acicarpha procumbens* (the first Calyceraceae genome), *Scaevola aemula* (Goodeniaceae), and *Platycodon grandiflorus* (Campanulaceae), and re-annotated five publicly available genomes in Menyanthaceae (*Nymphoides indica*) and Asteraceae (*Arctium lappa*, *Lactuca sativa*, *Chrysanthemum lavandulifolium*, and *Helianthus annuus*) (Fig. 1c). In doing so, we address a central evolutionary question: How have WGDs and small-scale gene duplications contributed to the origin and expansion of species diversity within Asterales? Our analyses highlight lineage-specific patterns shaped by genome duplications, chromosomal rearrangements, gene family dynamics, and signatures of selection to provide a comparative framework for understanding the genomic basis of Asteraceae's remarkable diversification.

Results

Extensive genome rearrangements yet similar composition in Asteraceae and Calyceraceae

We assembled chromosome-level genomes of *Platycodon*, *Scaevola*, and *Acicarpha*, with *Scaevola* and *Acicarpha* assemblies being haplotype-resolved. The assemblies were highly contiguous and complete based on BUSCO analysis, and comparable with the re-annotated genomes (range = 96.6%–98%, mean = 97.6%; Supplementary Figs. 1 and 2, Supplementary Tables 1 and 2). Genome rearrangements were widespread, with variations in chromosome number indicating multiple chromosomal fission/fusion events across all Asterales taxa (Fig. 1c,d).

As expected, repeat content was high in Asteraceae (range = 80.9%–89.3%)[28], and similarly elevated in *Scaevola* (Goodeniaceae, 82.1%) and *Acicarpha* (Calyceraceae). Unexpectedly, *Acicarpha* exhibited the highest repeat proportion, reaching 90.1% (Fig. 1e, Supplementary Table S2). In *Acicarpha*, the most abundant repeat type was *Ty1* (e.g., *copia*), comprising 28.63% of the genome, though *Arctium* had the highest and *Nymphoides* had the lowest *Ty1* content overall (33.82% and 10.84%, respectively). In contrast, *Helianthus* exhibited the highest proportion of *Ty3* (e.g., *gypsy*) elements (45.4%), while *Chrysanthemum* had the lowest (11.1%). The proportions of *Ty1*/*Ty3* retroelements show inverse patterns across the phylogeny, with earlier-diverging species having higher percentages of *Ty1* and lower *Ty3* (e.g., *Acicarpha*), while more recently derived taxa show the opposite (e.g., *Helianthus*; Fig. 1e), a pattern reported previously[28]. *Chrysanthemum* is the exception to this due to its low *Ty3* percentage; however, it has the highest percentage of Unclassified LTRs, potentially causing a decreased percentage in *Ty3* elements. Surprisingly, *Platycodon* and *Nymphoides* had an abundance of DNA transposons (26.9% and 23.3%, respectively), which are not shared with the rest of the species.

The proportions of duplicated gene types varied across Asterales genomes (Fig. 1d). Segmental duplicates, likely derived from WGDs[32], were most common in *Acicarpha* (18.2%), *Arctium* (19.1%), and *Helianthus* (19.8%) and least in *Scaevola* (2.5%). In contrast, small-scale duplicates (tandem, proximal, and dispersed), likely reflecting localized transposition or recombination events[32], were most frequent in *Nymphoides* (12.9%) and showed modest variation across other genomes (range = 7.6–10.8%, mean = 9.2%). Overall, *Acicarpha*'s duplication frequencies fell within the range observed across Asteraceae (Fig. 1e).

Shared and lineage-specific WGD in Asteraceae and Calyceraceae

173 Asteraceae is hypothesized to have experienced one to multiple paleopolyploidization
174 events during its evolution[2–4,23,24], though the phylogenetic position of WGD events within the
175 family and related lineages remains uncertain. Syntenic gene depths corroborate the presence
176 of ancient WGDs for all taxa assessed, ranging from two (*Platycodon, Nymphoides,* and
177 *Scaevola*) to five (*Helianthus*), as compared to *Vitis vinifera* (an ideal reference due to its lack of
178 WGD since the gamma [γ] event shared by all eudicots[33]; Fig. 2a, Supplementary Table 3). In
179 Calyceraceae, we observed syntenic duplications and, to a lesser extent, triplications between
180 *Acicarpha* and *Scaevola* (Calyceraceae versus Goodeniaceae), and between *Acicarpha* and
181 *Arctium* (Calyceraceae versus Asteraceae), indicating lineage-specific duplications within
182 Asteraceae and Calyceraceae (Fig. 2b, Supplementary Table 4).
183 Comparing synonymous substitution rate (K_S) distributions across Asterales, we inferred
184 multiple duplications: an ancient WGD around K_S=2 (likely the γ event); a WGD shared by
185 Goodeniaceae, Calyceraecae, and Asteraceae; small-scale duplications in *Scaevola* and
186 *Acicarpha*; and separate WGDs in *Chrysanthemum* and *Helianthus* (Fig. 2c; Supplementary
187 Figs. 3 and 4). Alternatively, using K_S distributions adjusted to the rate of a focal taxon: three
188 small-scale duplications in *Nymphoides*; a WGD shared by Goodeniaceae, Calyceraceae, and
189 Asteraceae; two WGDs in *Scaevola*; two WGDs in *Acicarpha*; one WGD at the crown of
190 Asteraceae; one WGD in *Lactuca*; two in *Chrysanthemum*; and one in *Helianthus.* Prior to the
191 divergence of Asteraceae, both analyses indicated a WGD shared by Goodeniaceae,
192 Calyceraceae, and Asteraceae, but only the rate-adjusted method identified the Asteraceae-
193 specific WGD (Supplemental Figs. 5 and 6).
194 To further investigate WGDs, we phased the genomes and examined relationships
195 among duplicated subgenomes. Using phased sub-genomes that retained the most collinear
196 genes homologous to *Vitis* (Supplementary Datasets 1 and 2; Supplemental Figs. 7-13),
197 phylogenetic analyses generally revealed that *Acicarpha* and Asteraceae share at least one
198 duplication event, while each lineage also experienced independent duplications (Fig. 2d).
199 These findings were further supported by gene fractionation results, which indicated that
200 *Acicarpha* sub-genomes differ in age from those of Asteraceae, and are generally more recent,
201 although this trend was not always consistent (e.g., patterns of gene fractionation are distinct in
202 *Lactuca* but overlap in *Chrysanthemum*; Fig. 2e).
203 Given the variation in WGD signals across analyses, we performed hypothesis testing to
204 better place WGDs under various models. Under the critical model[34], which allows duplication
205 and loss rates to vary among branches, we found support for WGDs before and after the
206 Calyceraceae-Asteraceae split and lineage-specific events within *Acicarpha*, *Chrysanthemum*,

207 and *Helianthus*. Under the relaxed model[34], which allows rate heterogeneity among and within
208 branches, we found support for events in all tested lineages (Supplementary Fig.14). These
209 results indicate multiple WGD scenarios (Figs. 2f, 3), underscoring the difficulty of placing
210 Asteraceae WGDs. Some findings support earlier work that inferred two events: one shared
211 between Calyceraceae and Asteraceae and a second, independent duplication in each lineage
212 (Fig. 2f). However, other analyses support an older event predating the split of Goodeniaceae
213 and Calyceraceae+Asteraceae, followed by lineage-specific duplications in Asteraceae and
214 Calyceraceae (Fig. 2f), a pattern not previously reported.

215 *Asteraceae gene family expansions comprised of mostly tandem duplicates*

216 To assess how duplications contributed to Asteraceae diversification, we examined gene
217 orthogroup expansions across the family. Orthogroups that underwent substantial expansion at
218 the base of Asteraceae (n = 115; Supplementary Fig. 15) were significantly enriched in
219 processes related to stress response, signaling, and metabolic regulation, including cellular
220 response to amino acid stimulus, sesquiterpene biosynthesis, and cell surface receptor
221 signaling (Fig. 4a, Supplementary Fig. 16). Among the most expanded gene families were those
222 related to stress response and signaling (e.g. *GLP9*[35] and *WAKL2*[36]; Table 1), and to
223 biosynthesis of primary and secondary metabolites (e.g., *FAD2*, involved in fatty acid
224 synthesis[37]; and *CYP71B23*, involved in sesquiterpene synthesis and modification[38]; Table 1).
225 These were primarily tandem duplicates (30%), followed by dispersed (28%), proximal (23%),
226 and then segmental (18%) duplicates (Supplementary Fig. 17). Alternatively, *Acicarpha*
227 expansions were not enriched in macromolecule biosynthesis, pathways responding to
228 hormones like auxin and to environmental stimuli such as wounding or organic substances (Fig.
229 4b). Of these, multiple orthogroups included genes related to *CRK8*, a cysteine-rich receptor-
230 like kinase typically involved abiotic and biotic stress responses[39] (Table 1). Over half (62%) of
231 *Acicarpha*'s expanded genes were dispersed duplicates, while 17% were tandem duplicates,
232 and only 4% were segmental duplicates (Supplementary Fig. 17).

233 The expansion of *FAD2* homologs in Asteraceae is hypothesized to have catalyzed
234 functional divergence[40] and supported the evolution of a unique N-C balance system in
235 conjunction with the loss of *PII,* a regulator of nitrogen-carbon assimilation[21]. The absence of *PII*
236 in *Acicarpha* and *Scaevola* suggests that this loss predates the Goodeniaceae split, clarifying
237 previous hypotheses about its timing. In contrast, *FAD2* expansion occurred only in Asteraceae
238 (Fig. 4c; Supplementary Dataset 3). Duplication patterns consisting of a few segmental genes,
239 with much larger proportions of proximal and tandem genes, mirrored previous research

240 suggesting that this expansion in Asteraceae arose via tandem duplications following genome
241 duplications[21] (Fig. 4c,d; Supplementary Fig. 18). This contrasts with Calyceraceae which has
242 fewer tandem and proximal duplicates (Fig. 4d). Also involved in metabolism, the expansion of
243 Asteraceae *CYP71B23* homologs involved a few segmental genes, but with higher proportions
244 of tandem, proximal, and dispersed genes (Fig. 4e,f; Supplementary Fig. 19). Overall, these
245 patterns indicate that gene family expansions, particularly those arising through tandem
246 duplications, have shaped metabolic and stress-response diversity in Asteraceae, linking
247 genome duplication and remodeling to the uneven diversification of Asterales.

248 *Site-specific selection in Asteraceae acting on mostly segmental duplicates*

249 Genes containing positively selected sites in all four Asteraceae taxa (n=586
250 orthogroups) were significantly enriched for the regulation of biosynthetic and metabolic
251 processes (Supplementary Fig. 20). Of these genes, segmental and dispersed duplicates
252 comprised 45% and 37%, respectively (Supplementary Fig. 17). In addition to the expansions in
253 the CYP superfamily—a large and functionally diverse gene family involved in secondary
254 metabolite synthesis, hormone regulation, and stress responses[41,42]—we observed positively
255 selected sites in all Asteraceae taxa for orthogroups related to *CYP96A1*, *CYP72A7*,
256 *CYP706A4*, *CYP706A6*, *CYP72A13*, and *CYP72A15* (Supplementary Dataset 3). Previous
257 studies also revealed strong lineage-specific diversification of CYP genes in Asteraceae,
258 indicating rapid evolutionary shifts, likely reflecting a central role in metabolic innovation[43].
259 To test for lineage-specific positive selection, we used the resulting 586 orthogroups with sites
260 under positive selection in Asteraceae species as input for branch model analyses. Of these, 47
261 showed evidence for episodic positive selection in an Asteraceae lineage (Supplementary
262 Dataset 3). Genes involved in the predominant biological processes associated with this gene
263 set (stress response and regulation of metabolic processes) include: *TMAC2*, a regulator of
264 abscisic acid and salt stress[44]; *WRKY27,* a transcription factor (TF) involved in plant defense
265 and floral development[45]; and *TCP5*, a TF that controls thermomorphogenesis[46], regulates
266 ethylene-mediated petal development and growth[47], and impacts leaf margin patterning[48] (Table
267 1). Other class II TCP TFs, *TCP2* and *TCP24*, also exhibited episodic positive selection in
268 Asteraceae. These genes regulate key developmental transitions and have been linked to floral
269 diversification in sunflower[49].

270 *Lineage-specific selection in Asteraceae supports fine-tuning of flowering time*

271 To focus on genes and pathways involved in floral innovation in Asteraceae, we
272 performed branch model selection analyses on a curated set of 335 flowering-associated

273 orthogroups (Supplementary Table 5). We identified 27 orthogroups with episodic positive
274 selection specific to all Asteraceae taxa (Supplementary Dataset 3) Of these, several
275 orthogroups contained genes involved in related pathways (photoperiod, autonomous, and
276 gibberellin) that regulate flowering time (Fig. 5). These genes function to directly or indirectly
277 suppress or activate *SOC1*, a floral integrator[50] that activates floral meristem identity genes,
278 such as *LFY* (a floral meristem identity transcriptional regulator that promotes flowering
279 transitions[51]) and *AP1* (a floral meristem identity MADS box gene[52]). Genes under selection in
280 Asteraceae were: *GI*, which promotes flowering via *FT* through the photoperiod pathway[53,54];
281 *FVE*, which promotes flowering by suppressing *FLC* in the autonomous pathway[55,56]; *CGA1* and
282 *GNC,* which delay flowering by suppressing *SOC1* in the gibberellin pathway[57,58]; *FD*, which
283 interacts with *FT* to activate *SOC1*; and *AGL24,* which integrates with *SOC1* to link floral signals
284 and promote flowering[59] (Fig. 5, Table 1). Of these, episodic positive selection was only
285 identified in *Acicarpha* homologs of *FD* and *CGA1*. Overall, selection on multiple genes across
286 interconnected pathways likely fine-tuned flowering time regulation in Asteraceae.

287 **Discussion**

288 While the role of WGD in driving gene expansion, diversification, and selection is evident
289 across our analyses, the number and timing of WGD events within Asterales remain difficult to
290 resolve despite having high-quality, contiguous genome assemblies. Contrasting results from
291 different approaches (Figs. 2f, 3) highlight the inherent challenges in detecting ancient WGDs,
292 particularly within large and diverse plant families. This complexity is compounded by processes
293 such as biased gene retention and loss, genome rearrangements, chromosomal fission and
294 fusion, incomplete lineage sorting, and hybridization, including allopolyploidy[60,61]. Synteny-
295 based methods, while powerful, often yield incomplete reconstructions of ancient events,
296 especially across distantly related taxa. Similarly, K_S-based approaches face limitations: signals
297 from ancient WGDs may be obscured by subsequent small-scale duplications and losses,
298 attenuated with time, or confounded with speciation signals, especially in cases of allopolyploidy
299 where increased variation in K_S values blurs inference[62].

300 Combining multiple forms of inference, our results support multiple hypotheses on WGD
301 placement in Asterales. Triplicate syntenic gene depths echo previous studies indicating a
302 shared paleohexaploidy event between Asteraceae and Calyceraceae [2–4] (Fig. 3a). However,
303 this inference was refined through sub-genome phasing and hypothesis testing to indicate that
304 Asteraceae and Calyceraceae shared a WGD, then each experienced their own (Fig. 3b). Some
305 K_S distribution and gene family expansion results suggest that the core WGD is shared not only

306 by Asteraceae and Calyceraceae, but also Goodeniaceae (Fig. 3c). However, this finding is
307 unlikely given the comparative syntenic gene depths and is probably an artefact of unaccounted
308 biological processes, such as small-scale duplications and biased gene retention[62].
309 Nonetheless, it is evident that both large-scale and small-scale duplications played a
310 formative role in the divergence and diversification of Asterales. While WGDs account for most
311 positively selected genes in Asteraceae, duplications originating from transposable element
312 activity (i.e., tandem, proximal, and dispersed duplicates) explained most gene family
313 expansions (Supplementary Fig. 17). A biased retention of segmental duplicates from
314 polyploidy, particularly those supporting complex processes and protein interactions, has been
315 observed in other systems, and the high proportion of segmental duplicates in Asteraceae and
316 *Acicarpha* may reflect this[32,63,64]. Rather than being lost or pseudogenized, retained duplicate
317 genes likely underwent subfunctionalization, neofunctionalization, or dosage amplification to
318 increase adaptive fitness. These findings highlight the substantial roles tandem duplication and
319 transposition, alongside WGD, have in generating genomic novelty[65].
320 In Asteraceae, post-duplication outcomes may have enhanced ecologically beneficial
321 processes such as secondary metabolite biosynthesis and cellular responses to environmental
322 stimuli, which are enriched in expanded and positively selected gene families (Supplementary
323 Fig. 20). Given Asteraceae's rich array of secondary compounds used in defense, attraction,
324 and environmental interaction[66], the lineage-specific retention and expansion genes involved in
325 metabolite production and regulation likely reflect specialization in metabolite biosynthesis.
326 Secondary chemistry contributes to key ecological interactions, including herbivore deterrence,
327 pathogen resistance, and pollinator attraction, and its diversification may have enhanced
328 environmental responsiveness and niche breadth[66]. Conversely, in *Acicarpha* the limited
329 expansion of metabolite biosynthesis genes may have limited their geographic range through
330 outcompetition. These patterns of gene family expansion, retention, and selection thus support
331 the broader view that chemical diversification, driven by underlying genomic innovation, has
332 been a major driver of Asteraceae's evolutionary success[5].
333 Advanced reproductive systems and strategies have also played a key role in
334 Asteraceae's evolutionary and ecological success. In a capitulum, flowering time and individual
335 floret development is precise and highly coordinated[13,14]. Our findings suggest that selective
336 pressures on duplicated genes and transposon activity may have fine-tuned gene regulation
337 associated with flowering time, conferring a key reproductive advantage in Asteraceae. Among
338 positively selected flowering time genes, patterns of evolution varied. For example, *FD*
339 homologs exhibited a two-step episode of positive selection: first on Asteraceae/*Acicarpha*

340 dispersed duplicates, then again on Asteraceae and *Acicarpha* duplicates individually. The *FVE*
341 homologs suggest that Asteraceae and *Acicarpha* segmental duplicates were derived from a
342 shared WGD, followed by positive selection on only one Asteraceae lineage. The patterns of *FD*
343 and *FVE* evolution reflect a shared history between Asteraceae and *Acicarpha*. Conversely, *GI*
344 and *CGA1* patterns reflect an Asteraceae-specific WGD. *GI* homologs showed two episodes of
345 diversifying selection in Asteraceae segmental duplicates, likely indicating sub-functionalization
346 or dosage compensation (Fig. 5). Among *CGA1* homologs, however, episodic positive selection
347 occurred only along one of the segmentally duplicated Asteraceae lineages. While likely
348 originating from transposon activity, rather than WGD, the dispersed duplicates of *Acicarpha*
349 *CGA1* homologs exhibited a similar pattern of selection on one duplicated lineage. These
350 findings underscore distinct mechanisms of functional divergence and the fate of duplicates
351 between Asteraceae and *Acicarpha* that may have conferred unique adaptive advantages.
352 Furthermore, the absence of positive selection on core integrators (e.g. *SOC1*) and
353 meristem identity genes (e.g., *LFY* and *AP1*) suggests that strong purifying selection preserves
354 central flowering functions, while upstream regulators offer an evolutionary interface for
355 adjusting developmental timing. This pattern suggests that adaptive evolution in Asteraceae has
356 primarily targeted regulatory modifications within the flowering gene network, enabling flexibility
357 without disrupting deeply conserved functions. These patterns imply lineage-specific fine-tuning
358 of flowering time, likely in response to environmental heterogeneity. Because some genes in
359 traditional plant model systems, e.g., *FLC* in *Arabidopsis* (Supplementary Fig. 21), are lineage-
360 specific, these species may not serve as broadly representative models for flowering time
361 regulation. Instead, groups like Asteraceae and other diverse clades offer additional insights into
362 the evolutionary dynamics of these pathways.
363 Genome duplication has influenced not only speciation, but also the phenotypic and
364 ecological breadth of plants, shaping habitat use, life history strategies, competitive interactions,
365 responses to herbivores and pathogens, and pollinator relationships[67–72]. In both Asteraceae
366 and Calyceraceae, ancient large- and small-scale duplications left a strong genomic imprint, yet
367 distinct evolutionary pressures acting on duplicated genes, particularly those involved in
368 metabolite biosynthesis and flowering time regulation, drove the emergence of key ecological
369 traits and ultimately enabled the extraordinary diversification of Asteraceae.

370 **Materials and Methods**

371 Taxon selection

372 For this study, we generated de novo chromosome-level assemblies of *Acicarpha*
373 *procumbens* Less. (Calyceraceae), hereafter *Acicarpha*; *Scaevola aemula* R.Br.
374 (Goodeniaceae), hereafter *Scaevola*; and *Platycodon grandiflorus* (Jacq.) A.DC. 'Komachi'
375 (Campanulaceae), hereafter *Platycodon*, using the materials and methods described below. We
376 included additional publicly available Asterales species for comparison: *Arctium lappa* L.[73]
377 (Asteraceae; NCBI: PRJNA764011), hereafter *Arctium*; *Lactuca sativa* L.[24] (Asteraceae; NCBI:
378 PRJNA173551), hereafter *Lactuca*; *Chrysanthemum lavandulifolium* (Fisch. ex Trautv.)
379 Makino[74], hereafter *Chrysanthemum* (Asteraceae; NCBI: PRJNA681093); *Helianthus annuus*
380 L.[29,75,76] (Asteraceae; HA412-HO v2: https://sunflowergenome.org/), hereafter *Helianthus*; and
381 *Nymphoides indica* (L.) Kuntze[77] (Menyanthaceae; CNGB: CNP0002767), hereafter
382 *Nymphoides*. Additionally, for broader evolutionary and functional comparisons, we included the
383 well-studied genomes: *Vitis vinifera* L. cv. Cabernet Sauvignon cl. 08 v 1.1[78] (Vitaceae) ,
384 hereafter *Vitis,* and *Arabidopsis thaliana* (L.) Heynh. (Brassicaceae; TAIR10;
385 https://v2.arabidopsis.org/), hereafter *Arabidopsis* (Supplementary Table 6).
386 We collected seeds of *Acicarpha* in Uruguay, purchased seeds of *Platycodon* through
387 Park Seed (https://parkseed.com/), and acquired cuttings of *Scaevola* through North Carolina
388 Farms Inc. (https://www.ncfarmsinc.com/). We grew all plants in greenhouses at the University
389 of Memphis (Memphis, Tennessee, USA) and Auburn University (Auburn, Alabama, USA). A
390 single individual of *Platycodon* was taken through two successive rounds of inbreeding prior to
391 use. *Acicarpha* and *Scaevola* individuals were not inbred before sequencing. Vouchers for each
392 specimen of *Acicarpha*, *Scaevola*, and *Platycodon* are available at the University of Memphis
393 herbarium (MEM) and Auburn University Herbarium (AUA) (see Supplementary Table 7 for
394 details).

395 Sequencing

396 For *Acicarpha*, we generated short- and long-read DNA, Omni-C, and RNA sequencing
397 data. For short reads, genomic DNA was extracted from ~25 mg of fresh leaf tissue from a
398 single individual ground with a Bead Mill 24 Homogenizer (ThermoFisher Scientific, Atlanta,
399 Georgia, USA) using the E.Z.N.A. SP Plant DNA Kit (Atlanta, Georgia, USA). High molecular
400 weight (HMW) DNA for long-read sequencing was isolated from 2.5–3 g of flash-frozen young
401 leaf tissue ground in liquid nitrogen with an autoclaved mortar and pestle using the TaKaRa
402 NucleoBond HMW DNA kit (Düren, Germany). RNA was extracted from young leaf (YL), shoot
403 apical meristem (SAM), and mature leaf (ML) tissues using the E.Z.N.A. Plant RNA kit (Omega
404 Bio-tek, cat #R6827-01) and treated with RNase-free DNase (Omega Bio-tek, cat #E1091).

405 Flash-frozen tissue was sent to the HudsonAlpha Institute for Biotechnology (Huntsville,
406 Alabama, USA) for Omni-C library preparation using the Dovetail Omni-C kit (Dovetail
407 Genomics, Inc.). The gDNA, Omni-C libraries, and the RNA were sequenced using an Illumina
408 NovaSeq 6000 with PE150 chemistry, the HMW DNA was sequenced on a PacBio Sequel II at
409 HudsonAlpha Institute for Biotechnology.
410 For *Scaevola*, short- and long-read DNA, Hi-C, and RNA data were generated using
411 equivalent tissue amounts and extraction kits as for *Acicarpha*. Hi-C libraries were prepared with
412 the Proximo Hi-C Kit-Plant (Phase Genomics, cat #KT3040) and sequenced at the University of
413 North Carolina's High Throughput Sequencing Facility (UNC HTSF) on a NovaSeq 6000 (PE-50
414 and PE-100). The gDNA and HMW DNA were sequenced on NovaSeq PE150 and PacBio
415 Sequel II sequencers, respectively, at HudsonAlpha Institute for Biotechnology.
416 For *Platycodon*, we generated short- and long-read DNA, Hi-C, and RNA sequencing
417 data. Genomic DNA was extracted from ~100 mg of homogenized tissue using a modified
418 CTAB protocol. Short-read libraries were prepared at UNC's HTSF with the KAPA Hyper DNA
419 kit and sequenced at BGI-Genomics on the DNBseq platform (PE100). Total RNA from YL and
420 SAM tissues of five individuals was extracted using the E.Z.N.A. Plant RNA kit (Omega Bio-tek,
421 #R6827-01), treated with RNase-free DNase (#E1091), and used for library preparation (KAPA
422 stranded mRNA-seq kit, UNC HTSF). Hi-C libraries were prepared with the Proximo Hi-C Kit-
423 Plant (Phase Genomics, #KT3040) following the v4.0 protocol (October 2020) and sequenced at
424 UNC HTSF on a NovaSeq 6000 (PE100). For long reads, HMW DNA was extracted using the
425 ONT "High molecular weight gDNA extraction from plant leaves" (CTAB–Genomic-tip) protocol,
426 purified with Amicon Ultra filters (#UFC510008), and size-selected using the Circulomics Short
427 Read Eliminator Kit (#SS-100-101-01, v2). ONT libraries were prepared with the Ligation
428 Sequencing Kit (#SQK-LSK109) following previously described modifications[124] and sequenced
429 on a MinION Mk1B (MIN-101B) using FLO-MIN106D flow cells (R9.4.1 pore) controlled by
430 MinKNOW v3.6.5. Two flow cells were used; each washed three times with the Flow Cell Wash
431 Kit (#EXP-WSH003) and reloaded with the same library.
432 We checked raw long-read and short-read sequence data quality using FastQC v.
433 0.11.9, a program designed to highlight potential issues with sequence data. Additionally, we
434 checked raw long-read data quality using PAUVRE[82] v. 0.2.3, which plots the length and quality
435 distribution of raw Oxford Nanopore (ONT) or PacBio reads (Fig. S19). We then trimmed raw
436 reads to filter out inadequate reads and remove adapters using fastp[83] v. 0.23.2.

437 <u>Genome assembly and scaffolding</u>

438 We assembled long-read data for *Scaevola* and *Acicarpha* using HifiASM[84] v. 0.19.5 and
439 v. 0.18.5, respectively, with Hi-C data for *Scaevola* and Omni-C data for *Acicarpha* integrated
440 into the assemblies, resulting in two haplotypes per species. We base-called the *Platycodon*
441 MinION data using the python package guppy v. 3.1.6 (ONT) with model R9.4.1_450bps in
442 high-accuracy mode. We then assembled the genome with a hybrid approach utilizing ONT
443 long-read and DNBseq (BGI) short-read genomic DNA sequencing as inputs. For the hybrid
444 assembly, we used MaSuRCA[85] v. 3.4.2, and polished with short-read data using bowtie2[86] v.
445 2.1.4 and pilon[87] v. 1.22.
446 We aligned Omni-C and Hi-C reads to the assembly using BWA[88,89] v. 0.7.17 and used
447 Samtools[90] v1.11 to convert the output sam file into a bam file, sorted the bam file, and created
448 an index for the assembly fasta file to prepare for scaffolding. We conducted scaffolding,
449 followed by manual curation of the *Scaevola*, *Acicarpha*, and *Platycodon* assemblies, using Yet
450 another Hi-C Scaffolding tool (YaHS)[91] v1.1 for *Acicarpha* and v. 1.2a.2 for *Platycodon* and
451 *Scaevola*, with the Omni-C data for *Acicarpha*, Hi-C data for the *Scaevola*, and *Platycodon* used
452 as input. We prepared the resulting assemblies for Juicebox[92] (v. 2.17.00 for *Acicarpha* and v.
453 2.15.00 for *Scaevola* and *Platycodon*) visualization using the Juicer Pre tools v1.1 included in
454 YaHS before inputting them into the Juicebox GUI for manual scaffolding and correction to
455 produce pseudochromosomes. We assessed genome assembly quality and completeness
456 using Benchmarking Universal Single-Copy Orthologs (BUSCO)[93] v. 5.4.7 with embryophyta
457 from OrthoDB (www.orthodb.org) as the reference lineage. We then generated assembly
458 statistics of the finalized assemblies using the publicly available python script
459 (https://github.com/lexnederbragt/sequencetools/blob/master/assemblathon_stats.py).

460 Genome annotation

461 We screened the *Acicarpha*, *Scaevola*, and *Platycodon* genomes for bacterial
462 contaminants using StandardDB from Kraken2[94] v. 2.0.8 and removed all contigs with positive
463 hits. Using the resulting files as input, we then used the Extensive de-novo TE Annotator
464 (EDTA)[95] pipeline v. 1.9.3 to run RepeatModeler2[96] and identify remaining TEs utilizing the
465 default parameters, except --overwrite, --sensitive, --anno, and --evaluate were set to '1'. Using
466 the decontaminated genome file and the resulting TE library file from EDTA, we then ran
467 RepeatMasker[97] v. 4.1.5 to soft mask the genomes. We annotated the soft masked genomes
468 with RNA sequences using Braker3[98] v. 3.0.3 and labelled functional gene annotations using
469 InterProScan[99] v. 5.66-98.0. We then used OMArk v. 0.3.0 (https://omark.omabrowser.org/,

470 accessed June 3, 2024) to assess the proteome quality of *Acicarpha*, *Scaevola*, and *Platycodon*
471 with Asterids (NCBI ID: 71274) as the ancestral clade.
472 To ensure consistency between annotations and comparative genomic analyses, we re-
473 annotated five publicly available genomes from *Nymphoides, Arctium, Chrysanthemum,*
474 *Helianthus, Lactuca* following the steps detailed above. We assessed annotation quality across
475 all eight taxa using the gff3_stats python script
476 (https://github.com/HuffordLab/GenomeQC/blob/master/scripts/gff3_stats.py) from
477 GenomeQC[100].

478 Plastid and mitochondrial genome assembly
479 We extracted chloroplast sequences from all trimmed paired end short read data using
480 the GetOrganelle toolkit[101] following the recommended parameters for embryophyte plant
481 plastomes but increasing the number of rounds and the disentangle time limit (-R 75 --
482 disentangle-time-limit 7200). For long read data, PacBio and ONT minion, we used ptGAUL to
483 extract chloroplast sequences[102]. We then annotated complete circular sequences using
484 GeSeq[103] and assessed collinearity between plastomes using the progressive alignment in
485 Mauve[104]. In addition, we assembled mitochondrial genomes of *Platycodon, Scaevola,* and
486 *Acicarpha* using OatK v. 1.0 (https://github.com/c-zhou/oatk).

487 Genome structure
488 We characterized genomic sequences' size, heterozygosity, and repetitiveness by k-mer
489 frequency analyses (k=21) using Jellyfish[105] and GenomeScope[106] 2.0. We assessed
490 heterozygous k-mer pairs, obtained with KMC3[107], using Smudgeplot[106] to estimate ploidy levels
491 and infer genomic complexity in terms of sub-genome composition. We identified lower (L) and
492 upper (U) end cut-off values, below and above which all k-mers were discarded as errors, using
493 k-mer coverage output from Genomescope 2.0, as recommended in the Smudgeplot
494 documentation (https://github.com/KamilSJaron/smudgeplot), where L = (kcov/2)-5.

495 Gene duplications
496 We investigated gene duplications of all eight target Asterales genomes, plus genomes
497 from *Vitis* and *Arabidopsis*, by assessing homology and synteny to identify duplicated genes
498 and genomic regions, then by analyzing K_S values to estimate the relative age of duplication
499 events. To quantify and classify types of gene duplications within each genome, we used the
500 "duplicate gene classifier" program in MCScanX[32]. Through this program, duplicated genes
501 were classified as: segmental, proximal, tandem, or dispersed, as defined by [32].

502 We utilized OrthoFinder v. 2.5.5[108] to identify conserved orthologs and gene duplications
503 across the ten genomes, using the -M msa parameter to generate a multiple species alignment.
504 We then identified the best evolutionary model and performed a maximum likelihood
505 phylogenetic analysis on the alignment using IQTree[109] v. 2.1.3 with 1,000 bootstraps to
506 generate a species tree. We used the orthologous gene groups identified by OrthoFinder for
507 downstream analyses to investigate evolutionary processes, however, we filtered the groups to
508 exclude multiple isoforms of the same gene by keeping only the longest isoform.

509 Syntenic gene depths

510 To investigate putative paleopolyploidy, we assessed synteny among the eight Asterales
511 genomes. We used the JCVI[110111] v. 1.3.8 python function "compara catalog ortholog" with the
512 default c-score cutoff of 0.7 to identify syntenic regions, then generated genome-wide synteny
513 plots of the full eight genomes following phylogenomic placement. We evaluated pairwise
514 syntenic gene depth between all genomes using the "compara synteny depth" option. We then
515 used these inferred values as input for subsequent analyses to phase subgenomes and further
516 elucidate polyploid origins (detailed in "Phasing subgenomes" below).

517 Phasing subgenomes

518 We implemented the two most prominent software tools to identify and separate
519 subgenomes, SubPhaser[112] and Whole-Genome Duplication Integrated (WGDI[111]). However,
520 given that the tested taxa have not experienced recent polyploidization events and are currently
521 recognized as diploids, SubPhaser could not be used with our data as it is unable to accurately
522 phase genomes that experienced paleopolyploidy events. Instead, we used WGDI v. 0.6.5 to
523 further investigate polyploidization events across Asterales using *Vitis* as a reference. Steps to
524 investigate polyploidy followed the subgenome aware analyses available on GitHub
525 (https://github.com/SunPengChuan/wgdi-example/blob/main/phase_subgenomes.md). Multiple
526 chromosomes of *Vitis* were tested as a reference for phasing, though only the chromosome with
527 the largest gene retention and collinearity, chromosome 11, was used for further analyses. Age
528 estimates from K_S values (e.g., Fig. S23) were used to help separate subgenomes if they were
529 one the same chromosome or were not distinct. We determined phylogenomic relationships
530 among the subgenomes using WGDI's '-at' parameter to generate gene trees with IQTree and
531 then ASTRAL-III[113] v. 5.7.8 to generate a final subgenome coalescent tree. We also assessed
532 gene fractionation within each subgenome to determine the evolutionary age of each
533 subgenome using the 'r' parameter of WGDI. We then reanalyzed the resulting subgenomes

534 within Asteraceae, Calyceraceae, and Goodeniaceae in JCVI to investigate synteny across
535 these phased regions using methods detailed above.

536 Accelerated gene evolution in *Acicarpha* and Asteraceae

537 To identify major targets of accelerated evolution within conserved genomic regions, we
538 used Cactus[114–116] v2.5.1, to generate a reference-free alignment of the Asterales subgenomes
539 that are syntenic to chromosome 11 of *Vitis*. Using the inferred phylogenomic relationship
540 between subgenomes as the guide tree to inform hierarchical alignment, we performed three
541 separate tests for accelerated evolution at different positions in the subgenome tree: across all
542 Asteraceae and *Acicarpha* subgenomes (clade 1, Fig. S11), in the *Acicarpha*-specific
543 duplication event (clade 2, Fig. S11), and in *Acicarpha* within the Subgenome 1 clade (clade 3,
544 Fig. S11).

545 Specifically, we created a neutral model of sequence evolution based on ancestral
546 repetitive sequences[117]. Repetitive elements were identified within the most ancestral sequence
547 in the subgenome alignment using RepeatModeler and RepeatMasker[97]. A random sample of
548 100,000 sites from these ancestral repeats was then selected to generate a multiple alignment
549 file (MAF) of repetitive sites across all subgenomes. This MAF was used to construct the neutral
550 model with PHAST phyloFit[118,119], using default settings and the same phylogeny used to
551 generate the subgenome alignment (Fig. S11).

552 With this neutral model, we identified conserved regions within the Asterales
553 subgenomes using PHAST phastCons[118,119], with *Vitis* as the reference. This was performed
554 separately for each of the three clade-specific analyses, with conserved regions detected in the
555 outgroup subgenomes. For each analysis, we obtained a list of the most-conserved elements
556 and converted their log-odds scores to length-normalized log-odds ratios. Elements with
557 normalized log-odds ratios below the 30th percentile were excluded from further analyses[120].
558 This filtering step produced three distinct lists of conserved regions in outgroup subgenomes,
559 allowing us to assess accelerated evolution within our clades of interest. We then tested for GC-
560 biased gene conversion (gBGC) within our ingroups using PHAST phastBias[118] and considered
561 any region with a score >0.5 to be impacted by gBGC. We filtered our lists of conserved regions
562 to remove any elements that were impacted by gBGC.

563 We used the likelihood ratio test of PHAST phyloP[118,121] to test the conserved elements
564 for accelerated evolution within our clades of interest. We corrected for multiple hypothesis
565 testing using Storey's correction method[122] (FDR=0.05). This produced a separate list of
566 accelerated regions (ARs) for each of our clade-specific analyses. To assess the biological

567 impacts of the accelerated evolution detected in each of our analyses, we identified the genes
568 most likely to be impacted by each AR. We associated an AR with a gene if it overlapped an
569 exon. Exon identities and locations were obtained from annotations of the *Vitis* genome[78].
570 ARs can be distributed unevenly across the genome, often clustering into regions of high
571 density, known as AR hotspots. To detect these hotspots, we employed the permutation method
572 of Cahill et al.[123]. Specifically, we used a sliding window approach, counting ARs in 100 kb
573 windows along *Vitis* chromosome 11, with windows offset by 10 kb steps. We conducted 100
574 permutations, each time randomly designating N conserved elements as accelerated, where N
575 corresponds to the actual number of ARs on the chromosome. For each permutation, we
576 repeated the sliding window approach to determine the maximum simulated AR density in a 100
577 kb window. Regions were designated as AR hotspots if their true AR density exceeded the 95th
578 percentile of maximum AR densities observed across all permutations. This approach prevents
579 regions with high conserved element density from being falsely identified as hotspots. We
580 identified separate hotspots for each of our clade-specific analyses.

581 K_S distributions

582 To investigate the influence of major genome duplication events and positive selection
583 on Asterales genomes, we assessed the non-synonymous (K_a) and synonymous (K_S)
584 substitution rates between gene pairs. Peaks in K_S value distributions, indicating the presence
585 of duplicated genes, were used to infer histories of WGD. For homologous gene pairs, we
586 estimated K_S values using an established pipeline[124] in which orthologs were aligned using
587 MUSCLE[125], then converted into nucleotides using PAL2NAL[126], then K_S values were calculated
588 between gene pairs using the F3X4 model in codeml within PAML[127]. We plotted K_S values of
589 paralogs and pairwise orthologous gene pairs in R[128,129] with kernel density estimations using
590 geom_density() from ggplot2[130]. Additionally, we used output from these analyses to identify
591 signatures of positive selection ($K_a/K_S > 1$), neutral selection ($K_a/K_S \approx 1$), and purifying selection
592 ($K_a/K_S < 1$) of paralogous gene pairs. To more accurately compare estimations of divergence
593 derived from orthologous K_S distributions with estimations of WGD derived from paralogous K_S
594 distributions, we then re-scaled ortholog distributions to the paralog scale of each focal species
595 using *ksrates*[131]. For these analyses, *Platycodon* was used as the outgroup, therefore, WGD
596 events were not assessed for this lineage.

597 WGD hyphothesis testing

598 Directly following previously published methods and scripts[132,133]
599 (https://gitlab.com/barker-lab/ajb_special_issue_2024.git), we tested predicted WGD events,

600 hypothesized by previous literature and identified by *ksrates*, using WHALE[34]. WHALE is a tool
601 designed to test WGD hypotheses by analyzing the relative support for different evolutionary
602 scenarios. It uses gene tree topologies and considers the expected patterns of gene retention
603 and loss following WGD events. By applying a likelihood-based framework, WHALE evaluates
604 whether observed gene tree discordance is best explained by WGD or alternative models,
605 making it a valuable approach for investigating genome evolution. Given results from *ksrates*
606 and primary literature, we tested nine total WGD events split into three separate analyses: one
607 based solely off findings from previous studies, the second analysis used only the results of
608 *ksrates*, and the third analysis combined results of the first two analyses (Fig. S15).

609 Evolution of gene family size

610 To investigate the evolution of gene family sizes in Asterales, we reconstructed ancestral
611 states of the orthologous groups identified by OrthoFinder by employing the fast algorithm[134]
612 through the 'ace' function in the *phytools*[135] R package. From the reconstructed ancestral states,
613 we applied targeted filters to classify orthogroups as: 1) specific to *Acicarpha*, 2) lost in
614 *Acicarpha*, 3) specific to Asteraceae (evolved in the ancestral branch of Asteraceae and absent
615 from other lineages), 4) lost in the last common ancestor of Asteraceae, 5) specific to both
616 Asteraceae and *Acicarpha*, and 6) lost in the last common ancestor of both Asteraceae and
617 *Acicarpha*. Filtering was based on a Bayes factor threshold of 10, indicating that the likelihood of
618 an orthogroup being present or absent at a given node was ten times higher than the alternative
619 scenario.

620 Next, we used CAFE[136] v. 5.0 to further investigate the dynamics of orthogroup
621 expansion and contraction. Since CAFE requires an ultrametric tree, we used TreePL[137,138] to
622 calibrate the phylogeny reconstructed by OrthoFinder. We selected four calibration points based
623 on previous work[6], constraining the minimum ages of ancestral nodes for the following groups:
624 Asteraceae and Calyceraceae (*Tubulifloridites lillei* Type A[139]; 72.1 Mya), all Asteraceae
625 included in the study (*Raiguenrayun cura*[140]; 47.5 Mya), *Helianthus* and *Lactuca* (*Cichorium*
626 *intybus* type pollen[141]; 22 Mya), and *Helianthus* and *Chrysanthemum* (*Artemisia* type pollen[142];
627 31 Mya). The root age was constrained between 121 and 126 Mya using a secondary
628 calibration point for Pentapetalae[143] (Fig. S24). The optimal smoothing parameter was
629 determined through cross-validation, testing 20 values across a range of magnitudes (from 1e-
630 10 to 1e+09) to ensure the best fit.

631 The tested CAFE models varied in the number of gamma parameters (ranging from 1 to
632 3), reflecting assumed differences in expansion and contraction rates across orthogroups.

633 Orthogroups identified by the best-fitting model as significantly expanded or contracted (Viterbi
634 ≤ 0.05) in Asteraceae, Acicarpha, and both Asteraceae and Acicarpha were retained for
635 downstream analyses, in addition to those identified as lost or specific to these groups. Genes
636 from significantly expanded orthogroups were then blasted to the *Arabidopsis thaliana*
637 database, TAIR10, using BLASTP[144] to determine closest gene similarity and trees were then
638 visualized in R.

639 Positive selection

640 We further investigated positive selection across Asterales genomes, focusing on
641 Asteraceae, using site-specific and branch-specific methods. For all taxa, we assessed
642 orthology and subsequent positive selection of single-copy orthologous groups using site-
643 specific methods within the AlexandrusPS pipeline[145]. For the orthogroups which were found to
644 exhibit positive selection for genes pertaining to all four Asteraceae taxa and no others, we
645 analyzed further by using branch-specific methods to determine if the found signatures of
646 positive selection originated from taxon-specific or family-specific lineages. Additionally, we
647 assessed a subset of orthogroups relating to floral development for lineage-specific positive
648 selection using branch-specific methods. These floral-related orthogroups were chosen based
649 on GO annotations for *Arabidopsis thaliana* genome (TAIR10: available at
650 https://v2.arabidopsis.org/) to include 18 terms related to meristem maintenance and
651 organization (GO:0010582, GO:0009933, GO:0009934, GO:0048507, GO:0010074,
652 GO:0019827), flower development (GO:0009908, GO:0010076, GO:0009910, GO:009911,
653 GO:0048438, GO:0010093, GO:2000032), vegetative shoot system/inflorescence development
654 (GO:0048367, GO:0010077, GO:0010492, GO:0090506,), and vegetative-reproductive phase
655 transition (GO:0048510). For both subsets of orthogroups, we generated codon alignments from
656 coding sequences by adding gaps for frameshifts and replacing stop codons with gaps using
657 MACSE[146]. Then we used the aBSREL (adaptive Branch-Site Random Effects Likelihood)
658 function in HyPhy[147] to test if positive selection has occurred on a proportion of branches. We
659 inspected results using HyPhy Vision (http://vision.hyphy.org/aBSREL), then visualized resulting
660 gene trees in R using Ape 5.0[148] and ggtree[149].

661 GO enrichment

662 For inferences on the functional role of genes identified in these analyses, we used the
663 basic local alignment search tool (BLAST[144]) to compare target proteins to a TAIR10 protein
664 database and identify matches based on an e-value threshold of 1e-10. For groups of gene
665 groups identified to be expanded or contracted or under positive selection, we performed gene

666 ontology (GO) enrichment analyses to summarize the significant GO terms for biological
667 processes[150]. We used the clusterProfiler[151] package in R, with annotations based on BLASTP
668 best-hit matches to the TAIR10 database. To account for multiple testing, p-values were
669 adjusted using the Benjamini–Hochberg false discovery rate (FDR) correction. Enrichment was
670 considered significant at an adjusted p-value (P-adjust) threshold of < 0.01.

671 **Acknowledgements**
672 We thank Jane Grimwood (HudsonAlpha), Eric Lyons (University of California, Berkley) for
673 assistance and consultation on CoGe, and Eric Spangler and Kristian Skjervold of the University
674 of Memphis High-Performance Computing (HPC) cluster.

675 **Author contributions**
676 ERMP, PE, JMB, DSJ, and JRM conceived the study. ERMP, PE, BRC, JB, MDP, and ZM
677 performed data curation, formal analyses, and developed methodologies and software. ERMP,
678 PE, JB, and JRM led the investigation, data validation, visualization, and drafting of the
679 manuscript. ACW contributed to data visualization. MDP, SD, and ZM provided additional
680 analytical and software support. AH, MB, ZN, JMB, DSJ, and JRM contributed to manuscript
681 review and editing. MB and ZN provided resources, and ZN, JMB, DSJ, and JRM acquired
682 funding and supervised the project. All authors read and approved of the final manuscript.

683 **Data availability**
684 Raw data and genome assembly data for *Acicarpha*, *Scaevola*, and *Platycodon* are uploaded to
685 NCBI BioProject PRJNA1253301. Accession numbers of these newly developed and publicly
686 available data can be found in Supplementary Table S6. Hardmasked, annotated genome
687 assemblies for *Acicarpha, Scaevola,* and *Platycodon,* as well as the re-annotated assemblies
688 for *Helianthus, Chrysanthemum, Lactuca, Arctium,* and *Nymphoides* were uploaded and made
689 publicly accessible on the Comparative Genomics Platform[152] (id67898, id67904, id67903,
690 id67900, id67899, id67901, id67867, id67902, respectively).

691 **Funding**
692 This work was supported by the National Science Foundation under grants IOS-2214472, DBI-
693 2319905, and DEB-745197, and by the National Institutes of Health through an R35 MIRA
694 award (R35GM119614).

References

1. Vamosi, J. C., Magallón, S., Mayrose, I., Otto, S. P. & Sauquet, H. Macroevolutionary Patterns of Flowering Plant Speciation and Extinction. *Annu Rev Plant Biol* **69**, 685–706 (2018).
2. Barker, M. S. *et al.* Most Compositae (Asteraceae) are descendants of a paleohexaploid and all share a paleotetraploid ancestor with the Calyceraceae. *Am J Bot* **103**, 1203–1211 (2016).
3. Huang, C.-H. *et al.* Multiple Polyploidization Events across Asteraceae with Two Nested Events in the Early History Revealed by Nuclear Phylogenomics. *Mol Biol Evol* **33**, 2820–2835 (2016).
4. Zhang, C. *et al.* Phylotranscriptomic insights into Asteraceae diversity, polyploidy, and morphological innovation. *J Integr Plant Biol* **63**, 1273–1293 (2021).
5. Palazzesi, L. *et al.* Asteraceae as a model system for evolutionary studies: from fossils to genomes. *Botanical Journal of the Linnean Society* **200**, 143–164 (2022).
6. Mandel, J. R. *et al.* A fully resolved backbone phylogeny reveals numerous dispersals and explosive diversifications throughout the history of Asteraceae. *Proc Natl Acad Sci U S A* **116**, 14083–14088 (2019).
7. Panero, J. L. & Crozier, B. S. Macroevolutionary dynamics in the early diversification of Asteraceae. *Mol Phylogenet Evol* **99**, 116–132 (2016).
8. Susanna, A. *et al.* The classification of the Compositae: A tribute to Vicki Ann Funk (1947–2019). *Taxon* **69**, 807–814 (2020).
9. Diazgranados, M. *et al.* World Checklist of Useful Plant Species. *KNB Data Repository* https://knb.ecoinformatics.org/view/doi:10.5063/F1CV4G34 (2020).
10. Lundberg, J. G. Asteraceae and relationships within Asterales. in (2009).
11. Brignone, N. F., Mazet, N., Pozner, R. & Denham, S. S. Calyceraceae: Unexpected diversification pattern in the Southern Andes. *Perspect Plant Ecol Evol Syst* **60**, 125744 (2023).
12. Denham, S. S., Zavala-Gallo, L., Johnson, L. A. & Pozner, R. E. Insights into the phylogeny and evolutionary history of Calyceraceae. *Taxon* **65**, 1328–1344 (2016).
13. Zhang, T. & Elomaa, P. Development and evolution of the Asteraceae capitulum. *New Phytologist* **242**, 33–48 (2024).
14. Gurung, V., Muñoz-Gómez, S. & Jones, D. S. Putting heads together: Developmental genetics of the Asteraceae capitulum. *Curr Opin Plant Biol* **81**, 102589 (2024).
15. Pozner, R., Zanotti, C. & Johnson, L. A. Evolutionary origin of the Asteraceae capitulum: Insights from Calyceraceae. *Am J Bot* **99**, 1–13 (2012).
16. Teeri, T. H. *et al.* Reproductive meristem fates in Gerbera. *J Exp Bot* **57**, 3445–3455 (2006).
17. Flagel, L. E. & Wendel, J. F. Gene duplication and evolutionary novelty in plants. *New Phytologist* **183**, 557–564 (2009).
18. Doyle, J. J. Polyploidy in Legumes. in *Polyploidy and Genome Evolution* 147–180 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2012). doi:10.1007/978-3-642-31442-1_9.

19. Eric Schranz, M., Mohammadin, S. & Edger, P. P. Ancient whole genome duplications, novelty and diversification: the WGD Radiation Lag-Time Model. *Curr Opin Plant Biol* **15**, 147–153 (2012).
20. Soltis, D. E. *et al.* Polyploidy and angiosperm diversification. *Am J Bot* **96**, 336–348 (2009).
21. Shen, F. *et al.* Comparative genomics reveals a unique nitrogen-carbon balance system in Asteraceae. *Nat Commun* **14**, (2023).
22. Kong, X. *et al.* Two-step model of paleohexaploidy, ancestral genome reshuffling and plasticity of heat shock response in Asteraceae. *Hortic Res* **10**, (2023).
23. Barker, M. S. *et al.* Multiple Paleopolyploidizations during the Evolution of the Compositae Reveal Parallel Patterns of Duplicate Gene Retention after Millions of Years. *Mol Biol Evol* **25**, 2445–2455 (2008).
24. Reyes-Chin-Wo, S. *et al.* Genome assembly with in vitro proximity ligation data and whole-genome triplication in lettuce. *Nat Commun* **8**, 14953 (2017).
25. Oliver, K. R., McComb, J. A. & Greene, W. K. Transposable Elements: Powerful Contributors to Angiosperm Evolution and Diversity. *Genome Biol Evol* **5**, 1886–1901 (2013).
26. Fedoroff, N. Transposons and genome evolution in plants. *Proceedings of the National Academy of Sciences* **97**, 7002–7007 (2000).
27. Lippman, Z. *et al.* Role of transposable elements in heterochromatin and epigenetic control. *Nature* **430**, 471–476 (2004).
28. Staton, S. E. & Burke, J. M. Evolutionary transitions in the Asteraceae coincide with marked shifts in transposable element abundance. *BMC Genomics* **16**, 623 (2015).
29. Badouin, H. *et al.* The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature* **546**, 148–152 (2017).
30. Schwacke, R., Bolger, M. E. & Usadel, B. PubPlant: a continuously updated online resource for sequenced and published plant genomes. Preprint at https://doi.org/10.1101/2025.03.12.642823 (2025).
31. Lundberg, J. & Bremer, K. A Phylogenetic Study of the Order Asterales Using One Morphological and Three Molecular Data Sets. *Int J Plant Sci* **164**, 553–578 (2003).
32. Wang, Y. *et al.* MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res* **40**, e49–e49 (2012).
33. Jiao, Y. *et al.* A genome triplication associated with early diversification of the core eudicots. *Genome Biol* **13**, R3 (2012).
34. Zwaenepoel, A. & Van de Peer, Y. Inference of Ancient Whole-Genome Duplications and the Evolution of Gene Duplication and Loss Rates. *Mol Biol Evol* **36**, 1384–1404 (2019).
35. Jin, Y. *et al.* Functional divergence of GLP genes between G. barbadense and G. hirsutum in response to Verticillium dahliae infection. *Genomics* **114**, 110470 (2022).
36. Wang, Z. *et al.* Comparative genomics analysis of WAK/WAKL family in Rosaceae identify candidate WAKs involved in the resistance to Botrytis cinerea. *BMC Genomics* **24**, 337 (2023).
37. Dar, A. A., Choudhury, A. R., Kancharla, P. K. & Arumugam, N. The FAD2 Gene in Plants: Occurrence, Regulation, and Role. *Front Plant Sci* **8**, (2017).

780 38. Li, Y. & Wei, K. Comparative functional genomics analysis of cytochrome P450 gene
781 superfamily in wheat and maize. *BMC Plant Biol* **20**, 93 (2020).
782 39. Quezada, E.-H. *et al.* Cysteine-Rich Receptor-Like Kinase Gene Family Identification in
783 the Phaseolus Genome and Comparative Analysis of Their Expression Profiles Specific
784 to Mycorrhizal and Rhizobial Symbiosis. *Genes (Basel)* **10**, 59 (2019).
785 40. Feng, T. *et al.* FAD2 Gene Radiation and Positive Selection Contributed to Polyacetylene
786 Metabolism Evolution in Campanulids. *Plant Physiol* **181**, 714–728 (2019).
787 41. Pandian, B. A., Sathishraj, R., Djanaguiraman, M., Prasad, P. V. V. & Jugulam, M. Role
788 of Cytochrome P450 Enzymes in Plant Stress Response. *Antioxidants* **9**, 454 (2020).
789 42. Chakraborty, P., Biswas, A., Dey, S., Bhattacharjee, T. & Chakrabarty, S. Cytochrome
790 P450 Gene Families: Role in Plant Secondary Metabolites Production and Plant Defense.
791 *J Xenobiot* **13**, 402–423 (2023).
792 43. Chen, Y., Klinkhamer, P. G. L., Memelink, J. & Vrieling, K. Diversity and evolution of
793 cytochrome P450s of Jacobaea vulgaris and Jacobaea aquatica. *BMC Plant Biol* **20**, 342
794 (2020).
795 44. Alshareef, N. O. *et al.* Root remodeling mechanisms and salt tolerance trade-offs: The
796 roles of HKT1, TMAC2, and TIP2;2 in Arabidopsis. *PLoS Genet* **21**, e1011713 (2025).
797 45. Mukhtar, M. S., Deslandes, L., Auriac, M., Marco, Y. & Somssich, I. E. The Arabidopsis
798 transcription factor WRKY27 influences wilt disease symptom development caused by
799 *Ralstonia solanacearum*. *The Plant Journal* **56**, 935–947 (2008).
800 46. Han, X. *et al.* Arabidopsis Transcription Factor TCP5 Controls Plant
801 Thermomorphogenesis by Positively Regulating PIF4 Activity. *iScience* **15**, 611–622
802 (2019).
803 47. van Es, S. W. *et al.* Novel functions of the Arabidopsis transcription factor TCP5 in petal
804 development and ethylene biosynthesis. *The Plant Journal* **94**, 867–879 (2018).
805 48. Yu, H. *et al.* TCP5 controls leaf margin development by regulating KNOX and BEL-like
806 transcription factors in Arabidopsis. *J Exp Bot* **72**, 1809–1821 (2021).
807 49. Chapman, M. A., Leebens-Mack, J. H. & Burke, J. M. Positive Selection and Expression
808 Divergence Following Gene Duplication in the Sunflower CYCLOIDEA Gene Family. *Mol*
809 *Biol Evol* **25**, 1260–1273 (2008).
810 50. Lee, J. & Lee, I. Regulation and function of SOC1, a flowering pathway integrator. *J Exp*
811 *Bot* **61**, 2247–2254 (2010).
812 51. Blackman, B. K. *et al.* Contributions of Flowering Time Genes to Sunflower Domestication
813 and Improvement. *Genetics* **187**, 271–287 (2011).
814 52. Li, D. *et al.* Arabidopsis Class II TCP Transcription Factors Integrate with the FT–FD
815 Module to Control Flowering. *Plant Physiol* **181**, 97–111 (2019).
816 53. Sawa, M. & Kay, S. A. GIGANTEA directly activates Flowering Locus T in Arabidopsis
817 thaliana. *Proceedings of the National Academy of Sciences* **108**, 11698–11703 (2011).
818 54. Sawa, M., Nusinow, D. A., Kay, S. A. & Imaizumi, T. FKF1 and GIGANTEA Complex
819 Formation Is Required for Day-Length Measurement in Arabidopsis. *Science (1979)* **318**,
820 261–265 (2007).
821 55. Xie, H. *et al.* DNA Methylation of the Autonomous Pathway Is Associated with Flowering
822 Time Variations in Arabidopsis thaliana. *Int J Mol Sci* **25**, 7478 (2024).

823 56. Simpson, G. G. The autonomous pathway: epigenetic and post-transcriptional gene
824 regulation in the control of Arabidopsis flowering time. *Curr Opin Plant Biol* **7**, 570–574
825 (2004).
826 57. Richter, R., Bastakis, E. & Schwechheimer, C. Cross-Repressive Interactions between
827 SOC1 and the GATAs GNC and GNL/CGA1 in the Control of Greening, Cold Tolerance,
828 and Flowering Time in Arabidopsis . *Plant Physiol* **162**, 1992–2004 (2013).
829 58. Bellucci, E. *et al.* Selection and adaptive introgression guided the complex evolutionary
830 history of the European common bean. *Nat Commun* **14**, 1908 (2023).
831 59. Huang, N., Tien, H. & Yu, T. Arabidopsis leaf-expressed AGAMOUS-LIKE 24 mRNA
832 systemically specifies floral meristem differentiation. *New Phytologist* **241**, 504–515
833 (2024).
834 60. Van de Peer, Y., Mizrachi, E. & Marchal, K. The evolutionary significance of polyploidy.
835 *Nat Rev Genet* **18**, 411–424 (2017).
836 61. Soltis, P. S., Marchant, D. B., Van de Peer, Y. & Soltis, D. E. Polyploidy and genome
837 evolution in plants. *Curr Opin Genet Dev* **35**, 119–125 (2015).
838 62. Dunn, T. & Sethuraman, A. Accurate Inference of the Polyploid Continuum Using
839 Forward-Time Simulations. *Mol Biol Evol* **41**, (2024).
840 63. Freeling, M. Bias in Plant Gene Content Following Different Sorts of Duplication: Tandem,
841 Whole-Genome, Segmental, or by Transposition. *Annu Rev Plant Biol* **60**, 433–453
842 (2009).
843 64. Freeling, M. & Thomas, B. C. Gene-balanced duplications, like tetraploidy, provide
844 predictable drive to increase morphological complexity. *Genome Res* **16**, 805–814
845 (2006).
846 65. Qiao, X. *et al.* Different Modes of Gene Duplication Show Divergent Evolutionary Patterns
847 and Contribute Differently to the Expansion of Gene Families Involved in Important Fruit
848 Traits in Pear (Pyrus bretschneideri). *Front Plant Sci* **9**, (2018).
849 66. *Systematics, Evolution, and Biogeography of Compositae*. (International Association for
850 Plant Taxonomy, Vienna, 2009).
851 67. Ramsey, J. & Ramsey, T. S. Ecological studies of polyploidy in the 100 years following its
852 discovery. *Philosophical Transactions of the Royal Society B: Biological Sciences* **369**,
853 20130352 (2014).
854 68. Strong, D. R. & Ayres, D. R. Ecological and Evolutionary Misadventures of *Spartina*.
855 *Annu Rev Ecol Evol Syst* **44**, 389–410 (2013).
856 69. Martin, S. L. & Husband, B. C. Adaptation of diploid and tetraploid Chamerion
857 angustifolium to elevation but not local environment. *Evolution (N Y)* **67**, 1780–1791
858 (2013).
859 70. Boalt, E., Arvanitis, L., Lehtilä, K. & Ehrlén, J. The association among herbivory
860 tolerance, ploidy level, and herbivory pressure in cardamine pratensis. *Evol Ecol* **24**,
861 1101–1113 (2010).
862 71. Thompson, J. N. & Merg, K. F. Evolution of polyploidy and the diversification of plant-
863 pollinator interactions. *Ecology* **89**, 2197–2206 (2008).
864 72. Oswald, B. P. & Nuismer, S. L. Neopolyploidy and pathogen resistance. *Proceedings of*
865 *the Royal Society B: Biological Sciences* **274**, 2393–2397 (2007).

866 73. Fan, W. *et al.* The genomes of chicory, endive, great burdock and yacon provide insights
867 into Asteraceae palaeo-polyploidization history and plant inulin production. *Mol Ecol*
868 *Resour* **22**, 3124–3140 (2022).
869 74. Wen, X. *et al.* The Chrysanthemum lavandulifolium genome and the molecular
870 mechanism underlying diverse capitulum types. *Hortic Res* **9**, (2022).
871 75. Todesco, M. *et al.* Massive haplotypes underlie ecotypic differentiation in sunflowers.
872 *Nature* **584**, 602–607 (2020).
873 76. Huang, K. *et al.* The genomics of linkage drag in inbred lines of sunflower. *Proceedings of*
874 *the National Academy of Sciences* **120**, (2023).
875 77. Yang, J.-S., Qian, Z.-H., Shi, T., Li, Z.-Z. & Chen, J.-M. Chromosome-level genome
876 assembly of the aquatic plant Nymphoides indica reveals transposable element bursts
877 and NBS-LRR gene family expansion shedding light on its invasiveness. *DNA Research*
878 **29**, (2022).
879 78. Minio, A., Cochetel, N., Vondras, A. M., Massonnet, M. & Cantu, D. Assembly of
880 complete diploid-phased chromosomes from draft genome sequences. *G3*
881 *Genes|Genomes|Genetics* **12**, (2022).
882 79. Doyle, J. J. & Doyle, J. L. A Rapid DNA Isolation Procedure for Small Quantities of Fresh
883 Leaf Tissue. *Phytochemical Bulletin* **19**, 11–15 (1987).
884 80. Vaillancourt, B. & Buell, C. R. High molecular weight DNA isolation method from diverse
885 plant species for use with Oxford Nanopore sequencing. Preprint at
886 https://doi.org/10.1101/783159 (2019).
887 81. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data.
888 Preprint at (2010).
889 82. Schultz, D., Ebbert, M. & De Coster, W. PAUVRE. Preprint at (2019).
890 83. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor.
891 *Bioinformatics* **34**, i884–i890 (2018).
892 84. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo
893 assembly using phased assembly graphs with hifiasm. *Nat Methods* **18**, 170–175 (2021).
894 85. Zimin, A. V. *et al.* Hybrid assembly of the large and highly repetitive genome of Aegilops
895 tauschii, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. *Genome*
896 *Res* **27**, 787–792 (2017).
897 86. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods*
898 **9**, 357–359 (2012).
899 87. Walker, B. J. *et al.* Pilon: An Integrated Tool for Comprehensive Microbial Variant
900 Detection and Genome Assembly Improvement. *PLoS One* **9**, e112963 (2014).
901 88. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
902 (2013).
903 89. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler
904 transform. *Bioinformatics* **25**, 1754–1760 (2009).
905 90. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, (2021).
906 91. Zhou, C., McCarthy, S. A. & Durbin, R. YaHS: yet another Hi-C scaffolding tool.
907 *Bioinformatics* **39**, (2023).
908 92. Durand, N. C. *et al.* Juicebox Provides a Visualization System for Hi-C Contact Maps with
909 Unlimited Zoom. *Cell Syst* **3**, 99–101 (2016).

910 93. Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A. & Zdobnov, E. M. BUSCO Update:
911 Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic
912 Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Mol Biol Evol* **38**,
913 4647–4654 (2021).
914 94. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2.
915 *Genome Biol* **20**, 257 (2019).
916 95. Ou, S. *et al.* Benchmarking transposable element annotation methods for creation of a
917 streamlined, comprehensive pipeline. *Genome Biol* **20**, 275 (2019).
918 96. Flynn, J. M. *et al.* RepeatModeler2 for automated genomic discovery of transposable
919 element families. *Proceedings of the National Academy of Sciences* **117**, 9451–9457
920 (2020).
921 97. Smit, A. F. A., Hubley, R. & Green, P. RepeatMasker Open-4.0. Preprint at
922 http://www.repeatmasker.org (2013).
923 98. Gabriel, L. *et al.* BRAKER3: Fully automated genome annotation using RNA-seq and
924 protein evidence with GeneMark-ETP, AUGUSTUS and TSEBRA. Preprint at
925 https://doi.org/10.1101/2023.06.10.544449 (2023).
926 99. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification.
927 *Bioinformatics* **30**, 1236–1240 (2014).
928 100. Manchanda, N. *et al.* GenomeQC: a quality assessment tool for genome assemblies and
929 gene structure annotations. *BMC Genomics* **21**, 193 (2020).
930 101. Jin, J. J. *et al.* GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of
931 organelle genomes. *Genome Biol* **21**, 1–31 (2020).
932 102. Zhou, W. *et al.* Plastid Genome Assembly Using Long-read data. *Mol Ecol Resour* **23**,
933 1442–1457 (2023).
934 103. Tillich, M. *et al.* GeSeq- versatile and accurate annotation of organelle genomes. *Nucleic*
935 *Acids Res* **45**, (2017).
936 104. Darling, A. C. E., Mau, B., Blattner, F. R. & Perna, N. T. Mauve: Multiple Alignment of
937 Conserved Genomic Sequence With Rearrangements. *Genome Res* **14**, 1394–1403
938 (2004).
939 105. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of
940 occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).
941 106. Ranallo-Benavidez, T. R., Jaron, K. S. & Schatz, M. C. GenomeScope 2.0 and
942 Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun* **11**, (2020).
943 107. Kokot, M., Dlugosz, M. & Deorowicz, S. KMC 3: counting and manipulating k-mer
944 statistics. *Bioinformatics* **33**, 2759–2761 (2017).
945 108. Emms, D. M. & Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative
946 genomics. *Genome Biol* **20**, (2019).
947 109. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic
948 Inference in the Genomic Era. *Mol Biol Evol* **37**, 1530–1534 (2020).
949 110. Tang, H. *et al.* JCVI: A versatile toolkit for comparative genomics analysis. *iMeta* **3**,
950 (2024).
951 111. Sun, P. *et al.* WGDI: A user-friendly toolkit for evolutionary analyses of whole-genome
952 duplications and ancestral karyotypes. *Mol Plant* **15**, 1841–1851 (2022).

112. Jia, K. *et al.* Sub Phaser: a robust allopolyploid subgenome phasing method based on subgenome-specific k-mers. *New Phytologist* **235**, 801–809 (2022).
113. Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19**, 15–30 (2018).
114. Zhang, R.-G., Shang, H.-Y., Jia, K.-H. & Ma, Y.-P. Subgenome phasing for complex allopolyploidy: case-based benchmarking and recommendations. *Brief Bioinform* **25**, (2023).
115. Paten, B. *et al.* Cactus: Algorithms for genome multiple sequence alignment. *Genome Res* **21**, 1512–1528 (2011).
116. Armstrong, J. *et al.* Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature* **587**, 246–251 (2020).
117. Sullivan, P. F. *et al.* Leveraging base-pair mammalian constraint to understand genetic variation and human disease. *Science (1979)* **380**, (2023).
118. Hubisz, M. J., Pollard, K. S. & Siepel, A. PHAST and RPHAST: phylogenetic analysis with space/time models. *Brief Bioinform* **12**, 41–51 (2011).
119. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* **15**, 1034–1050 (2005).
120. Keough, K. C. *et al.* Three-dimensional genome rewiring in loci with human accelerated regions. *Science (1979)* **380**, (2023).
121. Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res* **20**, 110–121 (2010).
122. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences* **100**, 9440–9445 (2003).
123. Cahill, J. A. *et al.* Positive selection in noncoding genomic regions of vocal learning birds is associated with genes implicated in vocal learning and speech functions in humans. *Genome Res* **31**, 2035–2049 (2021).
124. Yocca, A. *et al.* A chromosome-scale assembly for ‘d’Anjou’ pear. *G3: Genes, Genomes, Genetics* **14**, (2024).
125. Edgar, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792–1797 (2004).
126. Suyama, M., Torrents, D. & Bork, P. PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res* **34**, (2006).
127. Yang, Z. *PAML: A Program Package for Phylogenetic Analysis by Maximum Likelihood*. *CABIOS APPLICATIONS NOTE* vol. 13 http://mw511.biol.berkeley.edu/ziheng/paml.html. (1997).
128. R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org. Preprint at https://stat.ethz.ch/pipermail/r-help/2008-May/161481.html.
129. R Core Team. R: A language and environment for statistical computing. Preprint at https://www.r-project.org/ (2017).
130. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag, New York, 2016).

997 131. Sensalari, C., Maere, S. & Lohaus, R. ksrates: positioning whole-genome duplications
998 relative to speciation events in Ks distributions. *Bioinformatics* **38**, 530–532 (2022).
999 132. McKibben, M. T. W., Finch, G. & Barker, M. S. Species-tree topology impacts the
1000 inference of ancient whole-genome duplications across the angiosperm phylogeny. *Am J*
1001 *Bot* **111**, (2024).
1002 133. Chen, H. *et al.* Revisiting ancient polyploidy in leptosporangiate ferns. *New Phytologist*
1003 **237**, 1405–1417 (2023).
1004 134. Pupko, T., Pe, I., Shamir, R. & Graur, D. A Fast Algorithm for Joint Reconstruction of
1005 Ancestral Amino Acid Sequences. *Mol Biol Evol* **17**, 890–896 (2000).
1006 135. Revell, L. J. phytools: An R package for phylogenetic comparative biology (and other
1007 things). *Methods Ecol Evol* **3**, 217–223 (2012).
1008 136. De Bie, T., Cristianini, N., Demuth, J. P. & Hahn, M. W. CAFE: a computational tool for
1009 the study of gene family evolution. *Bioinformatics* **22**, 1269–1271 (2006).
1010 137. Maurin, K. J. L. An empirical guide for producing a dated phylogeny with treePL in a
1011 maximum likelihood framework. (2020).
1012 138. Smith, S. A. & O'Meara, B. C. treePL: divergence time estimation using penalized
1013 likelihood for large phylogenies. *Bioinformatics* **28**, 2689–2690 (2012).
1014 139. Barreda, V. D. *et al.* Early evolution of the angiosperm clade Asteraceae in the
1015 Cretaceous of Antarctica. *Proceedings of the National Academy of Sciences* **112**, 10989–
1016 10994 (2015).
1017 140. Barreda, V. D. *et al.* An extinct Eocene taxon of the daisy family (Asteraceae):
1018 evolutionary, ecological and biogeographical implications. *Ann Bot* **109**, 127–134 (2012).
1019 141. Hochuli, P. A. *Palynologische Untersuchungen Im Oligozän Und Untermiozän Der*
1020 *Zentralen Und Westlichen Paratethys*. (Geologischen Institut der Eidg. Technischen
1021 Hochschule und der Universtät Zürich, Zurich, 1978).
1022 142. Beentje, H. J. & Caligari, P. D. S. *Compositae: Proceedings of the International*
1023 *Compositae Conference, Kew, 1994*. (Royal Botanic Gardens, 1996).
1024 143. Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L. L. & Hernández-Hernández, T. A
1025 metacalibrated time-tree documents the early rise of flowering plant phylogenetic
1026 diversity. *New Phytologist* **207**, 437–453 (2015).
1027 144. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment
1028 search tool. *J Mol Biol* **215**, 403–410 (1990).
1029 145. Ceron-Noriega, A., Schoonenberg, V. A. C., Butter, F. & Levin, M. AlexandrusPS: A User-
1030 Friendly Pipeline for the Automated Detection of Orthologous Gene Clusters and
1031 Subsequent Positive Selection Analysis. *Genome Biol Evol* **15**, (2023).
1032 146. Ranwez, V., Harispe, S., Delsuc, F. & Douzery, E. J. P. MACSE: Multiple Alignment of
1033 Coding SEquences Accounting for Frameshifts and Stop Codons. *PLoS One* **6**, e22594
1034 (2011).
1035 147. Kosakovsky Pond, S. L. *et al.* HyPhy 2.5—A Customizable Platform for Evolutionary
1036 Hypothesis Testing Using Phylogenies. *Mol Biol Evol* **37**, 295–299 (2020).
1037 148. Paradis, E. & Schliep, K. Ape 5.0: An environment for modern phylogenetics and
1038 evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).

149. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* **8**, 28–36 (2017).
150. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat Genet* **25**, 25–29 (2000).
151. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS* **16**, 284–287 (2012).
152. Lyons, E., Pedersen, B., Kane, J. & Freeling, M. The Value of Nonmodel Genomes and an Example Using SynMap Within CoGe to Dissect the Hexaploidy that Predates the Rosids. *Trop Plant Biol* **1**, 181–190 (2008).
153. GBIF.org User. Occurrence Download. Preprint at https://doi.org/10.15468/DL.RJ2PNP (2025).
154. GBIF.org User. Occurrence Download. Preprint at https://doi.org/10.15468/DL.BRQ5TF (2025).
155. GBIF.org User. Occurrence Download. Preprint at https://doi.org/10.15468/DL.66G87F (2025).

1056 **Table 1.** Representative genes from analyses aiming to understand the genomic mechanisms underpinning diversification in
1057 Asterales. Gene symbols were obtained from BLASTP search results to a TAIR10 protein database. Functions were summarized
1058 from representative publications. Asterisk indicates that this gene was the best match for six different orthogroups. Complete results
1059 are listed in Supplementary Dataset S3.

Category	Gene	Description	Involved in
Top ten expanded orthogroups in Asteraceae	*GLP9*	germin-like protein 9	disease resistance (Jin et al., 2022)
	WAKL2	wall associated kinase-like 2	cellular communication, disease resistance (Wang et al., 2023)
	FAD2	fatty acid desaturase 2	fatty acid synthesis (Dar et al., 2017; Feng et al., 2019)
	CYP71B23	cytochrome P450, family 71, subfamily B, polypeptide 37	sisquiterpene synthesis and modification (Y. Li & Wei, 2020)
	AT1G43760	DNAse I-like superfamily protein	NA
	RPP13	NB-ARC domain-containing disease resistance protein	disease resistance (Yuan et al., 2025)
	WIP4	Wound-responsive family protein	root development (Diaz-Ramirez et al., 2022)
	HYR1	UDP-Glycosyltransferase superfamily protein	growth regulation (Tan et al., 2022)
	CYP72A15	cytochrome P450, family 72, subfamily A, polypeptide 15	merabolic resistance (Zhou et al., 2025)
	HTB9	Histone superfamily protein	chromatin remodeling (Liu et al., 2007)
Top ten expanded orthogroups in *Acicarpha*	*NFU3*	NFU domain protein 3	plant growth, development, and flowering (Nath et al., 2017)
	*CRK8**	cysteine-rich RECEPTOR-like kinase	abiotic and biotic stress response (Quezada et al., 2019)
	LRR4	NB-ARC domain-containing disease resistance protein	disease resistance (Cao et al., 2024)
	AT2G01050	zinc ion binding; nucleic acid binding	NA
	SAUR2	SAUR-like auxin-responsive protein family	stress response (C. Li et al., 2024)
Lost in Goodeniaceae, Calyceraceae, and Asteraceae	*GLB1*(*PII*)	Nitrogen regulatory protein P-II homolog	nitrogen and metabolic regulation (Shen et al., 2023)
Positive Selection in Asteraceae	*CYP72A15*	cytochrome P450, family 72, subfamily A, polypeptide 15	secondary metabolism (Prall et al., 2016)
	TMAC2	AFP2 (ABI five-binding protein 2) family protein	stress response (Alshareef et al., 2025)
	TCP2,5,24	TEOSINTE BRANCHED 1, cycloidea and PCF transcription factor 2,5,24)	plant growth and development (D. Li et al., 2019; van Es et al., 2018; Yu et al., 2021)
	WRKY6,21,22,27, 31,39,42,57,74,75	WRKY DNA-binding proteins	stress response (Bakshi & Oelmüller, 2014)
	GI	gigantea protein	flowering time regulation (Sawa et al., 2007; Sawa & Kay, 2011)
	FVE	Transducin family protein / WD-40 repeat family protein	regulation of flowering time and cold response (Baek et al., 2008)
	CGA1	cytokinin-responsive gata factor 1	regulation of chloroplast development, stress response, and flowering time (Chiang et al., 2012; Richter et al., 2013)
	FD	Basic-leucine zipper (bZIP) transcription factor family protein	regulation of floral transition, plant development, and environmental response (Yang et al., 2025)
	AGL24	AGAMOUS-like 24	regulation of flowering time (Huang et al., 2024)

Table 1 References:

Alshareef, N. O., Melino, V. J., Saber, N., De Rosa, A., Rey, E., Wang, J. Y., AlBabili, S., Byrt, C., Tester, M. A., & Julkowska, M. M. (2025). Root remodeling mechanisms and salt tolerance trade-offs: The roles of HKT1, TMAC2, and TIP2;2 in Arabidopsis. PLOS Genetics, 21(6), e1011713. https://doi.org/10.1371/journal.pgen.1011713

Baek, I.-S., Park, H.-Y., You, M. K., Lee, J. H., & Kim, J.-K. (2008). Functional Conservation and Divergence of FVE Genes that Control Flowering Time and Cold Response in Rice and Arabidopsis. Molecules and Cells, 26(4), 368–372. https://doi.org/10.1016/S1016-8478(23)14010-6

Bakshi, M., & Oelmüller, R. (2014). WRKY transcription factors. Plant Signaling & Behavior, 9(2), e27700. https://doi.org/10.4161/psb.27700

Cao, Y., Mo, W., Li, Y., Xiong, Y., Wang, H., Zhang, Y., Lin, M., Zhang, L., & Li, X. (2024). Functional characterization of NBS-LRR genes reveals an NBS-LRR gene that mediates resistance against Fusarium wilt. BMC Biology, 22(1), 45. https://doi.org/10.1186/s12915-024-01836-x

Chiang, Y.-H., Zubo, Y. O., Tapken, W., Kim, H. J., Lavanway, A. M., Howard, L., Pilon, M., Kieber, J. J., & Schaller, G. E. (2012). Functional Characterization of the GATA Transcription Factors GNC and CGA1 Reveals Their Key Role in Chloroplast Development, Growth, and Division in Arabidopsis . Plant Physiology, 160(1), 332–348. https://doi.org/10.1104/pp.112.198705

Dar, A. A., Choudhury, A. R., Kancharla, P. K., & Arumugam, N. (2017). The FAD2 Gene in Plants: Occurrence, Regulation, and Role. Frontiers in Plant Science, 8. https://doi.org/10.3389/fpls.2017.01789

Diaz-Ramirez, D., Diaz-Garcia, U. S., Magdaleno-Garcia, G., Huep, G., Appelhagen, I., Sagasser, M., & Marsch-Martinez, N. (2022). Expression and Functional Analyses of the WIP Gene Family in Arabidopsis. Plants, 11(15), 2010. https://doi.org/10.3390/plants11152010

Feng, T., Yang, Y., Busta, L., Cahoon, E. B., Wang, H., & Lü, S. (2019). FAD2 Gene Radiation and Positive Selection Contributed to Polyacetylene Metabolism Evolution in Campanulids. Plant Physiology, 181(2), 714–728. https://doi.org/10.1104/pp.19.00800

Huang, N., Tien, H., & Yu, T. (2024). Arabidopsis leaf-expressed AGAMOUS-LIKE 24 mRNA systemically specifies floral meristem differentiation. New Phytologist, 241(1), 504–515. https://doi.org/10.1111/nph.19293

Jin, Y., Fan, L., Zhang, Y., Hu, W., Han, X., Yan, Q., Yang, J., Li, F., & Yang, Z. (2022). Functional divergence of GLP genes between G. barbadense and G. hirsutum in response to Verticillium dahliae infection. Genomics, 114(5), 110470. https://doi.org/10.1016/j.ygeno.2022.110470

Li, C., Jiang, C., Pan, X., Qi, Y., Zhao, W., Dong, H., & Liu, Q. (2024). OsSAUR2, a small auxin-up RNA gene, is crucial for arsenic tolerance and accumulation in rice. Environmental and Experimental Botany, 226, 105894. https://doi.org/10.1016/j.envexpbot.2024.105894

Li, D., Zhang, H., Mou, M., Chen, Y., Xiang, S., Chen, L., & Yu, D. (2019). Arabidopsis Class II TCP Transcription Factors Integrate with the FT–FD Module to Control Flowering. Plant Physiology, 181(1), 97–111. https://doi.org/10.1104/pp.19.00252

Li, Y., & Wei, K. (2020). Comparative functional genomics analysis of cytochrome P450 gene superfamily in wheat and maize. BMC Plant Biology, 20(1), 93. https://doi.org/10.1186/s12870-020-2288-7

Liu, Y., Koornneef, M., & Soppe, W. J. J. (2007). The Absence of Histone H2B Monoubiquitination in the Arabidopsis hub1 (rdo4) Mutant Reveals a Role for Chromatin Remodeling in Seed Dormancy. The Plant Cell, 19(2), 433–444. https://doi.org/10.1105/tpc.106.049221

Nath, K., O'Donnell, J. P., & Lu, Y. (2017). Chloroplastic iron-sulfur scaffold protein NFU3 is essential to overall plant fitness. Plant Signaling & Behavior, 12(2), e1282023. https://doi.org/10.1080/15592324.2017.1282023

Prall, W., Hendy, O., & Thornton, L. E. (2016). Utility of a Phylogenetic Perspective in Structural Analysis of CYP72A Enzymes from Flowering Plants. PLOS ONE, 11(9), e0163024. https://doi.org/10.1371/journal.pone.0163024

Quezada, E.-H., García, G.-X., Arthikala, M.-K., Melappa, G., Lara, M., & Nanjareddy, K. (2019). Cysteine-Rich Receptor-Like Kinase Gene Family Identification in the Phaseolus Genome and Comparative Analysis of Their Expression Profiles Specific to Mycorrhizal and Rhizobial Symbiosis. Genes, 10(1), 59. https://doi.org/10.3390/genes10010059

Richter, R., Bastakis, E., & Schwechheimer, C. (2013). Cross-Repressive Interactions between SOC1 and the GATAs GNC and GNL/CGA1 in the Control of Greening, Cold Tolerance, and Flowering Time in Arabidopsis . Plant Physiology, 162(4), 1992–2004. https://doi.org/10.1104/pp.113.219238

Sawa, M., & Kay, S. A. (2011). GIGANTEA directly activates Flowering Locus T in Arabidopsis thaliana. Proceedings of the National Academy of Sciences, 108(28), 11698–11703. https://doi.org/10.1073/pnas.1106771108

Sawa, M., Nusinow, D. A., Kay, S. A., & Imaizumi, T. (2007). FKF1 and GIGANTEA Complex Formation Is Required for Day-Length Measurement in Arabidopsis. Science, 318(5848), 261–265. https://doi.org/10.1126/science.1146994

Shen, F., Qin, Y., Wang, R., Huang, X., Wang, Y., Gao, T., He, J., Zhou, Y., Jiao, Y., Wei, J., Li, L., & Yang, X. (2023). Comparative genomics reveals a unique nitrogen-carbon balance system in Asteraceae. Nature Communications, 14(1). https://doi.org/10.1038/s41467-023-40002-9

Tan, Y., Yang, J., Jiang, Y., Wang, J., Liu, Y., Zhao, Y., Jin, B., Wang, X., Chen, T., Kang, L., Guo, J., Cui, G., Tang, J., & Huang, L. (2022). Functional Characterization of UDP-Glycosyltransferases Involved in Anti-viral Lignan Glycosides Biosynthesis in Isatis indigotica. Frontiers in Plant Science, 13. https://doi.org/10.3389/fpls.2022.921815

van Es, S. W., Silveira, S. R., Rocha, D. I., Bimbo, A., Martinelli, A. P., Dornelas, M. C., Angenent, G. C., & Immink, R. G. H. (2018). Novel functions of the Arabidopsis transcription factor TCP5 in petal development and ethylene biosynthesis. The Plant Journal, 94(5), 867–879. https://doi.org/10.1111/tpj.13904

Wang, Z., Ma, Y., Chen, M., Da, L., Su, Z., Zhang, Z., & Liu, X. (2023). Comparative genomics analysis of WAK/WAKL family in Rosaceae identify candidate WAKs involved in the resistance to Botrytis cinerea. BMC Genomics, 24(1), 337. https://doi.org/10.1186/s12864-023-09371-9

Yang, H., Zhong, M., Liu, Y., Chen, L., Liu, B., Kong, F., Li, H., & Yue, L. (2025). Multifaceted roles of FD gene family in flowering, plant architecture, and adaptation. Frontiers in Plant Science, 16. https://doi.org/10.3389/fpls.2025.1602756

Yu, H., Zhang, L., Wang, W., Tian, P., Wang, W., Wang, K., Gao, Z., Liu, S., Zhang, Y., Irish, V. F., & Huang, T. (2021). TCP5 controls leaf margin development by regulating KNOX and BEL-like transcription factors in Arabidopsis. Journal of Experimental Botany, 72(5), 1809–1821. https://doi.org/10.1093/jxb/eraa569

Yuan, B., Li, C., Wang, Q., Yao, Q., Guo, X., Zhang, Y., & Wang, Z. (2025). Identification and functional characterization of the RPP13 gene family in potato (Solanum tuberosum L.) for disease resistance. Frontiers in Plant Science, 15. https://doi.org/10.3389/fpls.2024.1515060

Zhou, F.-Y., Lu, H., Huang, X., Han, Y., Zhang, Y., Han, H., Tsykalchuk, L., Nyporko, A., & Yu, Q. (2025). Cytochrome CYP72A15 may play a role in metabolic resistance to mesotrione in wild radish. Pesticide Biochemistry and Physiology, 210, 106380. https://doi.org/10.1016/j.pestbp.2025.106380

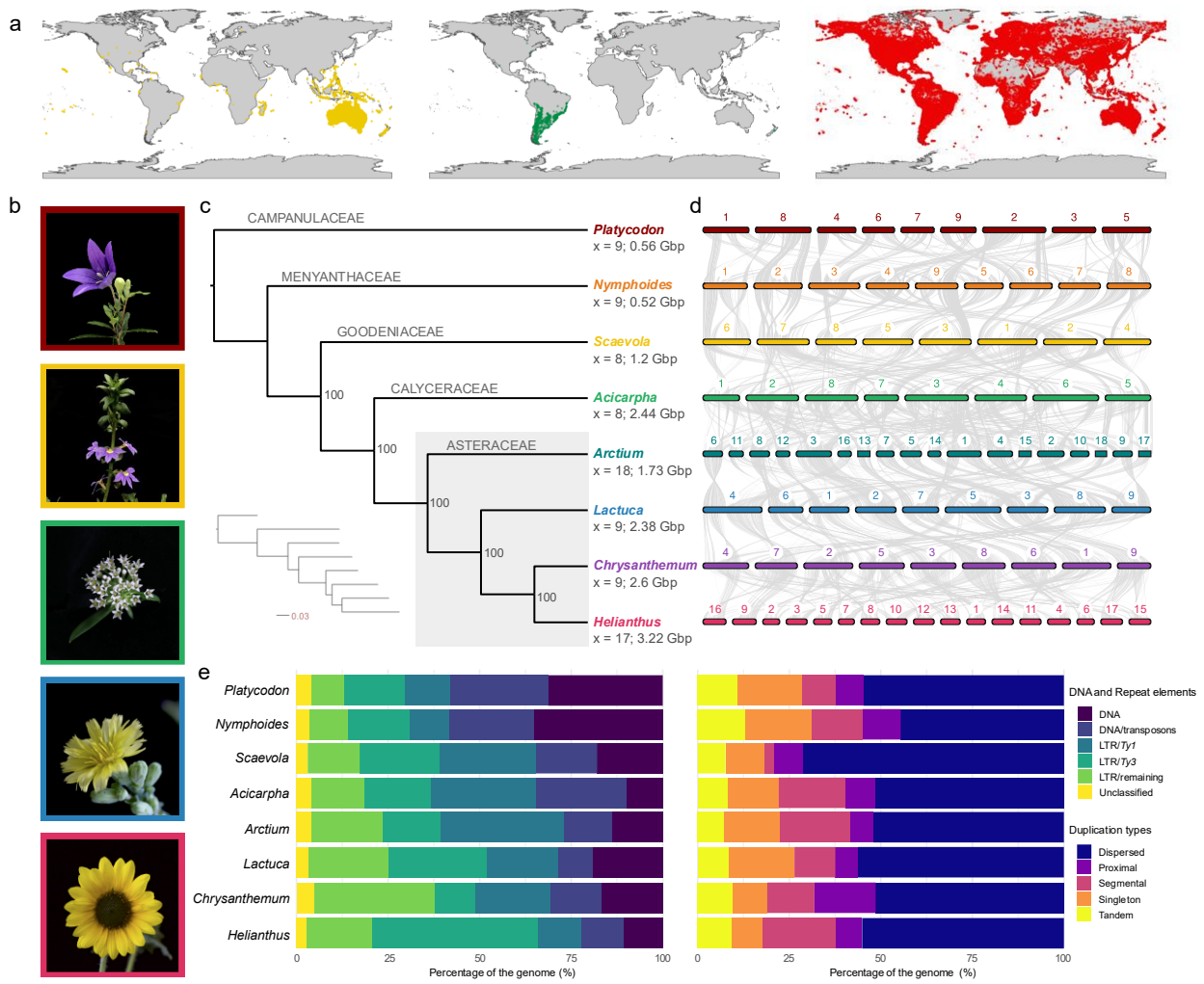

Fig. 1. Characteristics of the eight focal Asterales species. **a)** from left to right, distribution of Goodeniaceae (yellow), Calyceraceae (green), and Asteraceae (red), based on occurrence points of preserved specimens accessed from the Global Biodiversity Information Facility[153–155]. **b)** Photos of some tested taxa. Colored box around image matches species colors in **c** and **d**. From top to bottom: *Platycodon grandiflorus* (Campanulaceae); *Scaevola aemula* (Goodeniaceae); *Acicarpha procumbens* (Calyceraceae); *Lactuca sativa* (Asteraceae); *Helianthus annuus* (Asteraceae). Image of *Acicarpha procumbens* by D.S. Jones, remaining images by A.C. Willoughby. **c)** Nuclear phylogenetic cladogram of the eight species generated via the multiple-species alignment from OrthoFinder and IQTree. Values at nodes are bootstrap support values. Tree to the bottom left shows branch lengths by nucleotide substitutions per site. The four Asteraceae taxa are shaded gray. Values under each species name at the tip label correspond to the estimated hapoloid chromosome number (x) and genome assembly size in Gb, respectively. **d)** Synteny plots generated with JCVI ordered by phylogenetic placement of

1131 the eight target taxa. Each bar is representative of a single chromosome, with circled numbers
1132 above the bar indicating chromosome number. Gray lines between species' chromosomes
1133 indicate a single syntenic gene. **e)** Barplots indicating the percentage of repeat elements (left)
1134 and gene duplication type (right) across the genomes with percentage of the genome on the x-
1135 axis and species on the y-axis. Colors correspond to the legends to the right. Taxon order
1136 follows phylogenetic placement from **c.**

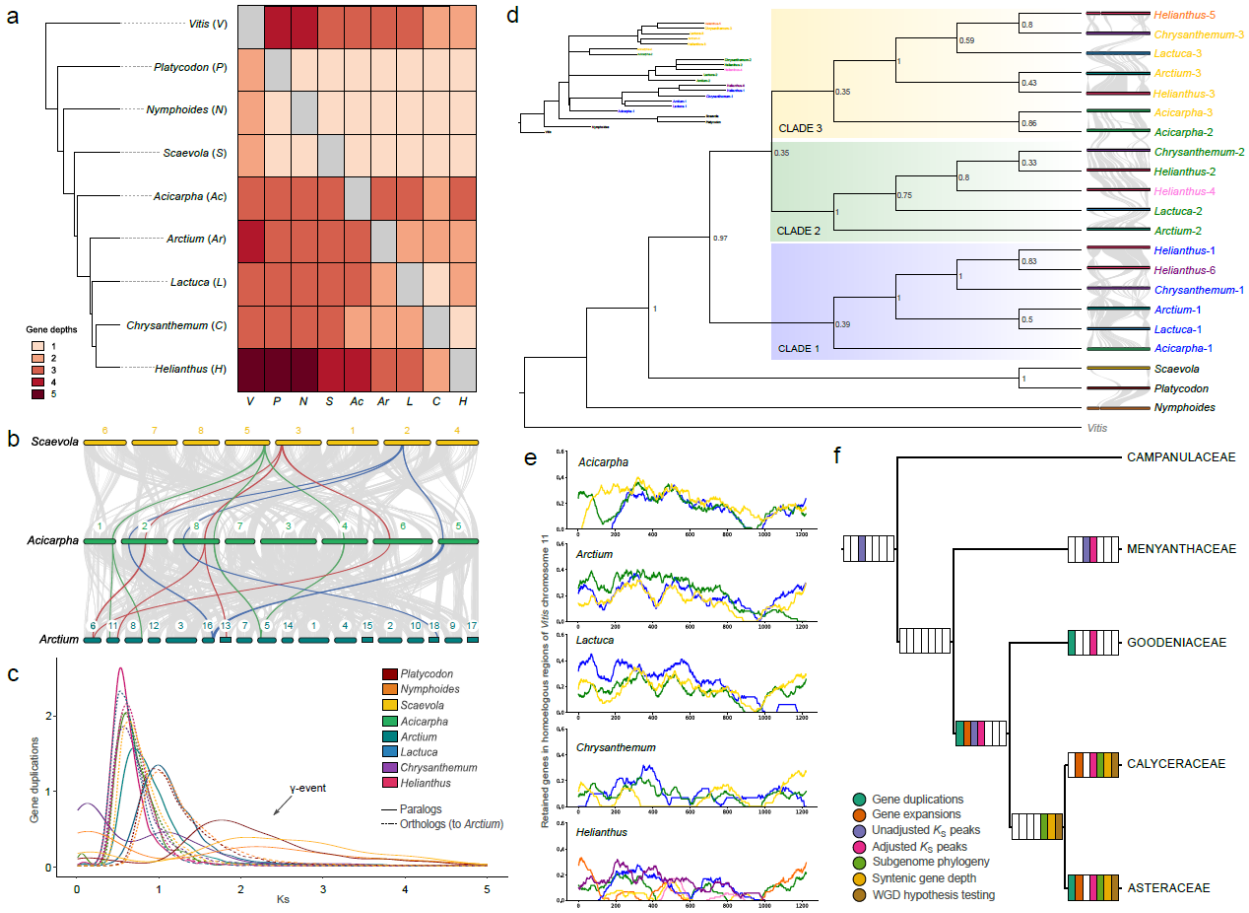

Fig. 2. Evidence of WGD events across the eight tested Asterales taxa. **a)** Pairwise gene depth ratios from JCVI between the eight Asterales species plus Vitis. The top right portion of the matrix indicates the gene depth ratio of the y-axis, while the bottom left portion is for the x-axis (e.g., *Scaevola:Acicarpha* is 1:3). Colors follow the legend in the bottom left of the panel. **b)** Syntenic regions between *Scaevola* (Goodeniacae), *Acicarpha* (Calyceraceae), and *Arctium* (Asteraceae) and as indicated by JCVI. Colored synteny ribbons are identified gene triplications between *Scaevola* to *Acicarpha* and follow those genes to *Arctium*. Bars represent chromosomes with circled numbers above the bars representing the chromosome number. **c)** K_S plot of paralogous gene pairs and orthologous gene pairs against Arctium for each species tested. Peaks closer to zero indicate a younger duplication event. Colors of lines correspond to the legend at the top right. **d)** Chromosome 11 subgenome phylogeny as determined by WGDI. Potential duplication events are indicated by clade, from Clades 1-3. Species names at the tips are colored according to subgenome number: blue: 1, green: 2, yellow: 3, pink: 4, orange: 5, purple: 6. Synteny plots at each tip were generated with JCVI and colored according to the synteny plots of Figure 1c. Local posterior probability (LPP) support values are at each node. **e)** Genome fractionation results from the subgenome analysis of chromosome 11 from WGDI for

1154 Acicarpha and four Asteraceae subgenomes. Gene retention is on the y-axis and base pair
1155 number on the x-axis. Colors in the graph correspond to the subgenome numbers which are
1156 also represented as the tip label color in d. **f)** Summary of WGD results across this study.
1157 Colored boxes indicate the findings, as indicated by the legend, that provide evidence for WGDs
1158 on that branch. White boxes indicate no evidence was found for that analysis. Campanulaceae
1159 has no boxes as WGDs were not directly tested for that family.

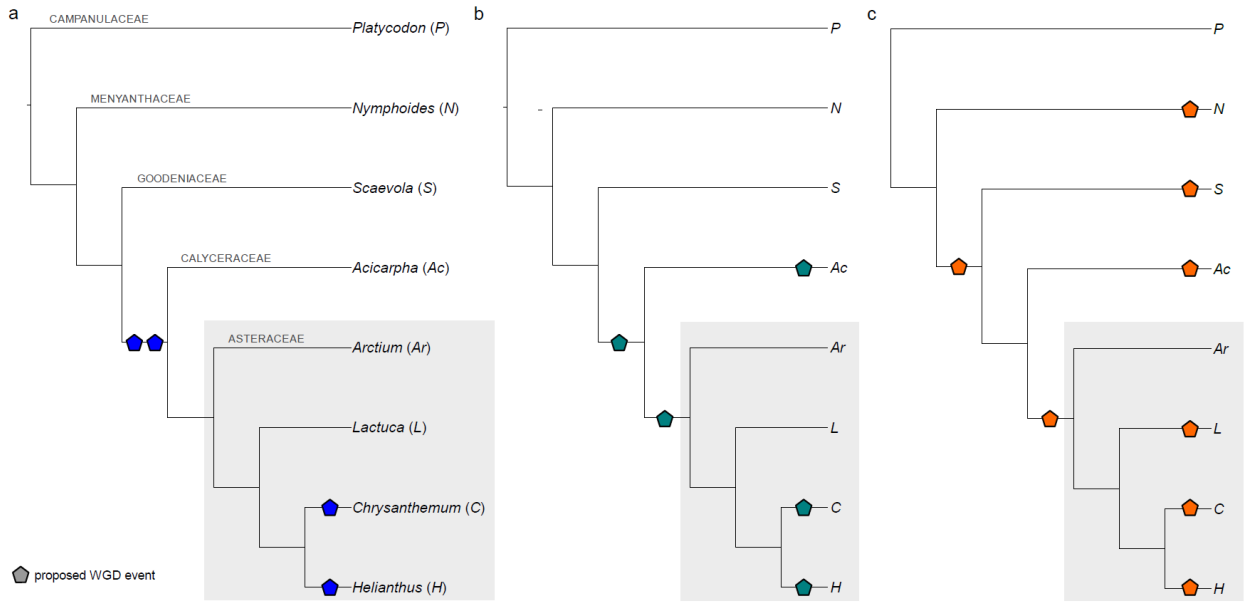

Fig. 3. Possible WGD scenarios in Asterales. A colored pentagon indicates a WGD placement. The scenarios are as follows: **a)** hypothesis for WGT at the crown of Asteraceae and Calyceraceae, followed by unique WGD in *Chrysanthemum* and *Helianthus* lineages, supported by previous literature, **b)** hypothesis for WGD at the crown of Asteraceae and Calyceraceae, followed by WGD at the base of Asteraceae, then unique WGD in *Acicarpha, Chrysanthemum,* and *Helianthus* lineages, supported by the critical model in WHALE, **c)** hypothesis for WGD at the crown of Asteraceae, Calyceraceae, and Goodeniaceae, as well as unique WGD in *Nymphoides, Scaevola, Acicarpha, Lactuca, Chrysanthemum,* and *Helianthus* lineages, supported by the relaxed model in WHALE.

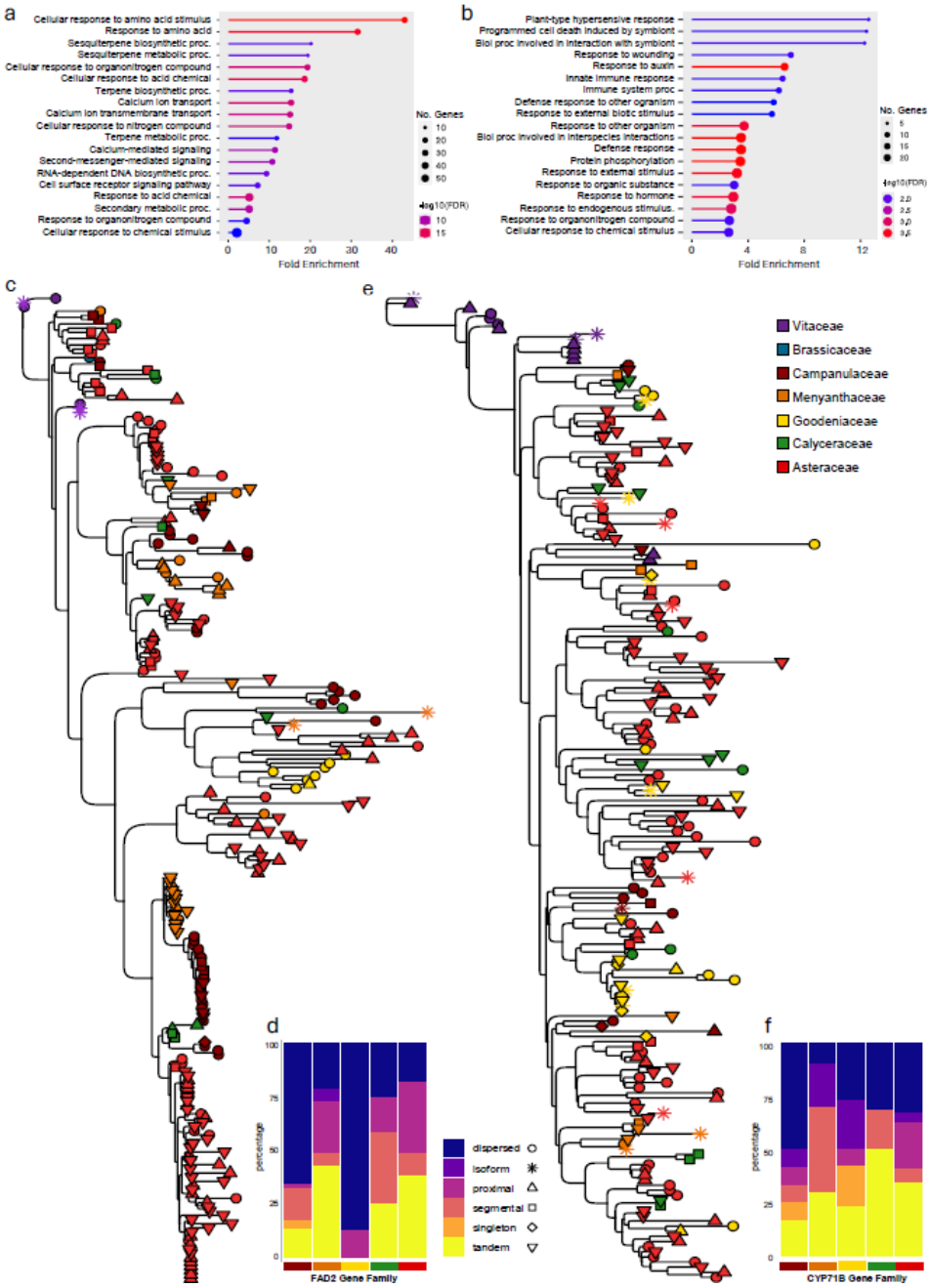

Fig. 4. Genes underpinning Asteraceae diversification. GO enrichment of biological processes associated with expanded orthogroups in **a)** Asteraceae and **b)** *Acicarpha;* **c)** Expanded FAD2 gene family tree where tip colors represent the taxonomic family of each gene as indicated by the legend at the right and tip shapes represent the duplication type of each gene as indicated by the legend in d**; d)** Composition of each duplication type by taxonomic family for the FAD2 gene family; **e)** Expanded CYP1B23 gene family tree with same tip shape and color designations as c. **f)** Composition of each duplication type by taxonomic family for the CYP1B23 gene family**.**

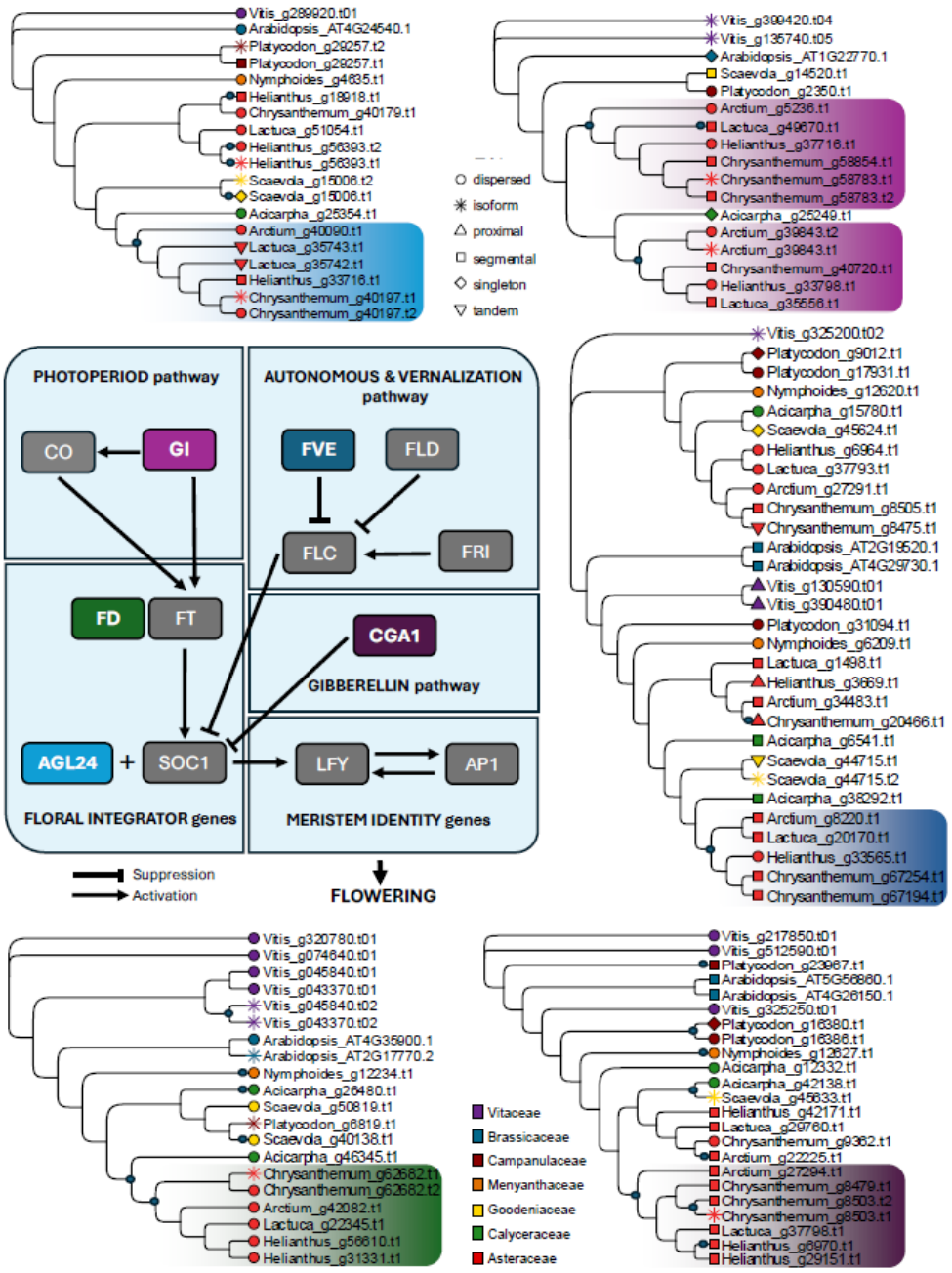

Fig. 5. Flowering time regulatory network involving: *CO, GI, FT, FD, FVE, FLD, FRI, FLC, CGA1, GNC, AGL24, SOC1, LFY,* and *AP1.* Colored gene boxes indicate orthogroups which experienced episodic positive selection in a lineage of Asteraceae homologs. Episodes of positive selection for only Asteraceae lineages are visualized as shaded blocks in phylogenetic trees and correspond to their respective gene color. Within the phylogenetic gene trees, additional episodes of positive selection are indicated as small black circles along nodes and tips, duplication types of genes are indicated by tip shapes, and plant families are indicated by color.