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, Memphis, Tennessee, USA*
[2]*Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, Connecticut, USA*
[3]*Department of Biological Sciences, Clemson University, Clemson, South Carolina, USA*
[4]*Department of Biological Sciences, Auburn University, Auburn, Alabama, USA*
[5]*Department of Plant Biology, University of Georgia, Athens, Georgia, USA*
[6]*Institute of Evolutionary Biology, Biological and Chemical Research Centre, University of Warsaw, Warsaw, Poland*
[7]*Department of Systems Biology, UMass Chan Medical School, Worcester, Massachusetts, USA*
[8]*Department of Genomics and Computational Biology, UMass Chan Medical School, Worcester, Massachusetts, USA*
[9]*Department of Crop, Soil, and Environmental Sciences, Auburn University, Auburn, Alabama, USA*
[10]*HudsonAlpha Institute for Biotechnology, Huntsville, Alabama, USA*
[11]*Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA*
[12]*Facultad de Agronomía, Universidad de la República, Montevideo, Uruguay*
[13]*Curriculum in Genetics and Molecular Biology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA*
[14]*The Plant Center, University of Georgia, Athens, Georgia, USA*
[15]*ACRE: Institute for Agricultural and Conservation Research and Education, University of Memphis, Memphis, Tennessee, 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, ORCID: 0000-0003-3930-3239

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

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

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

Zachary Meharg, zmeharg@hudsonalpha.org, ORCID: 0009-0004-7382-8527

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

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

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

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

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

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

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

‡ Jennifer R. 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 genomes. Comparative analyses revealed exceptionally high repeat content in both Asteraceae and Calyceraceae, pervasive chromosomal rearrangements, and evidence for shared and lineage-specific WGDs. 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 flowering time regulators suggest an evolved balance between regulatory flexibility and developmental constraint in floral diversification. These patterns reveal 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

Main

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, with >30,000 species belonging to Asteraceae alone[2,3]. 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[4–10]. 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[7,11]. In sharp contrast, its sister family, Calyceraceae, comprises fewer than 50 species confined to southern South America (Fig. 1a)[12–14]. Both Asteraceae and Calyceraceae originated in South America during the late Cretaceous[8,9], but only Asteraceae subsequently diversified extensively and dispersed globally[7]. 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[5–9]. Its extensive diversification has been largely attributed to enhanced reproductive success through pollinator attraction to its unique head-like inflorescence composed of many individual florets, the capitulum[6,15,16]. Although Calyceraceae inflorescences also resemble capitula, they did not spur analogous radiations. Key differences distinguishing the two families include the absence/presence of a terminal flower, un-fused/fused anthers, free/partially fused filaments attached to the corolla (Asteraceae versus Calyceraceae, respectively), and variation in secondary pollen presentation mechanisms[15,17,18]. Additionally, the production of diverse phytochemicals in Asteraceae facilitated ecological adaptation, particularly through defense responses[9]. These traits, coupled with a diversity of capitulum forms derived from distinct morphological and organizational modifications (Fig. 1b)[16], likely contributed to the disparity in species richness between the two families, though the genomic mechanisms underlying their origin have yet to be elucidated.

Acting across multiple scales, from individual genes to entire genomes, duplications have been a major driver of plant diversification by providing the raw genetic material for functional innovation[19]. 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[20–22]. 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[23–26]. Although limited by available genomic data, evidence also suggests that the sister family Calyceraceae experienced the same paleohexaploidy event[4–6]. This finding raises further questions about the role of mechanisms beyond WGDs in plant diversification. For example, 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[27–29]. 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[30]. While both WGDs and TE-mediated small-scale duplications have profoundly shaped plant evolution, their relative contributions to diversification remain unresolved. Resolving the precise phylogenetic placement of these events within Asterales is key to understanding the enormous evolutionary success of Asteraceae.

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*[31] and *Lactuca*[26]). To date, assembled genomes are mostly concentrated to the highly derived clades of Asteraceae with no representation of the early South American lineages[32]. 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 fundamental 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[30], repeat content was high in Asteraceae (range = 80.9%–89.3%)[28], though similarly elevated in *Scaevola* (Goodeniaceae, 82.1%) and *Acicarpha* (Calyceraceae). Unexpectedly, *Acicarpha* exhibited the highest repeat proportion, reaching 90.1% (Fig. 1e, Supplementary Table 2), with *Ty1* being the most abundant repeat type, comprising 28.63% of the genome. *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* elements (45.4%), while *Chrysanthemum* had the lowest (11.1%). Except for *Chrysanthemum*, 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[30].

The proportions of duplicated gene types varied across Asterales genomes (Fig. 1d). Segmental duplicates, likely derived from WGDs[33], 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[33], 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

Asteraceae is hypothesized to have experienced one to multiple paleopolyploidization events during its evolution[4–6,25,26], though the phylogenetic position of WGDs within the family and related lineages remains uncertain. Syntenic gene depths corroborate the presence of ancient WGDs for all taxa assessed, ranging from two (*Platycodon, Nymphoides,* and *Scaevola*)

170 to five (*Helianthus*), as compared to *Vitis vinifera* (an ideal reference due to its lack of WGD
171 since the gamma [γ] event shared by all eudicots[34]; Fig. 2a, Supplementary Table 3). In
172 Calyceraceae, we observed syntenic duplications and, to a lesser extent, triplications between
173 *Acicarpha* and *Scaevola* (Calyceraceae versus Goodeniaceae), and between *Acicarpha* and
174 *Arctium* (Calyceraceae versus Asteraceae), indicating lineage-specific duplications within
175 Asteraceae and Calyceraceae (Fig. 2b, Supplementary Table 4).
176 Comparing synonymous substitution rate (K_S) distributions across Asterales, we inferred
177 multiple duplications: an ancient WGD around $K_S = 2$ (likely the γ event); a WGD shared by
178 Goodeniaceae, Calyceraecae, and Asteraceae; small-scale duplications in *Scaevola* and
179 *Acicarpha*; and separate WGDs in *Chrysanthemum* and *Helianthus* (Fig. 2c, Supplementary
180 Figs. 3 and 4). Alternatively, using K_S distributions adjusted to the rate of a focal taxon, we
181 found: three small-scale duplications in *Nymphoides*; a WGD shared by Goodeniaceae,
182 Calyceraceae, and Asteraceae; two WGDs in *Scaevola*; two WGDs in *Acicarpha*; one WGD at
183 the crown of Asteraceae; one WGD in *Lactuca*; two in *Chrysanthemum*; and one in *Helianthus.*
184 Prior to the divergence of Asteraceae, both analyses indicated a WGD shared by
185 Goodeniaceae, Calyceraceae, and Asteraceae, but only the rate-adjusted method identified the
186 Asteraceae-specific WGD (Supplementary Figs. 5 and 6).
187 To further investigate WGDs, we phased the genomes and examined relationships
188 among duplicated subgenomes. Using phased sub-genomes that retained the most collinear
189 genes homologous to *Vitis* (Supplementary Datasets 1 and 2; Supplementary Figs. 7-13),
190 phylogenetic analyses revealed that *Acicarpha* and Asteraceae generally share at least one
191 duplication event, while each lineage also experienced independent duplications (Fig. 2d).
192 These findings were further supported by gene fractionation results, which indicated that
193 *Acicarpha* sub-genomes differ in age from those of Asteraceae, and are generally more recent,
194 although this trend was not always consistent (e.g., patterns of gene fractionation are distinct in
195 *Lactuca* but overlap in *Chrysanthemum*; Fig. 2e).
196 Given the variation in WGD signals across analyses, we performed hypothesis testing to
197 better place WGDs under various models. Under the critical model[35], which allows duplication
198 and loss rates to vary among branches, we found support for WGDs before and after the
199 Calyceraceae-Asteraceae split and lineage-specific events within *Acicarpha*, *Chrysanthemum*,
200 and *Helianthus*. Under the relaxed model[35], which allows rate heterogeneity among and within
201 branches, we found support for events in all tested lineages (Supplementary Fig. 14). These
202 results propose multiple WGD scenarios (Figs. 2f and 3), underscoring the difficulty of placing
203 Asteraceae WGDs. Some findings support earlier work that inferred two events: one shared

204 between Calyceraceae and Asteraceae and a second, independent duplication in each lineage
205 (Fig. 2f)[36]. However, other analyses support an older event predating the Goodeniaceae-
206 Calyceraceae-Asteraceae split followed by lineage-specific duplications in Asteraceae and
207 Calyceraceae (Fig. 2f), a pattern not previously reported.

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

209 To assess how duplications contributed to Asteraceae diversification, we examined gene
210 orthogroup expansions across the family. Orthogroups that underwent substantial expansion at
211 the base of Asteraceae (n = 115; Supplementary Fig. 15) were significantly enriched in
212 processes related to stress response, signaling, and metabolic regulation, including cellular
213 response to amino acid stimulus, sesquiterpene biosynthesis, and cell surface receptor
214 signaling (Fig. 4a, Supplementary Fig. 16). Among the most expanded gene families were those
215 related to stress response and signaling (e.g., *GLP9*[37] and *WAKL2*[38]; Table 1), and to
216 biosynthesis of primary and secondary metabolites (e.g., *FAD2*, involved in fatty acid
217 synthesis[39]; and *CYP71B23*, involved in sesquiterpene synthesis and modification[40]; Table 1).
218 These were primarily tandem duplicates (30%), followed by dispersed (28%), proximal (23%),
219 and then segmental (18%) duplicates (Supplementary Fig. 17). Alternatively, *Acicarpha*
220 expansions were not enriched in macromolecule biosynthesis, but in pathways responding to
221 hormones like auxin and environmental stimuli such as wounding or organic substances (Fig.
222 4b). Of these, multiple orthogroups included genes related to *CRK8*, a cysteine-rich receptor-
223 like kinase typically involved abiotic and biotic stress responses[41] (Table 1). Over half (62%) of
224 *Acicarpha*'s expanded genes were dispersed duplicates, while 17% were tandem duplicates,
225 and only 4% were segmental duplicates (Supplementary Fig. 17).

226 The Asteraceae-specific expansion of *FAD2* homologs is hypothesized to have
227 catalyzed functional divergence[42,43] and supported the evolution of a unique N-C balance
228 system in conjunction with the loss of *PII,* a regulator of nitrogen-carbon assimilation[23]. The
229 absence of *PII* in *Acicarpha* and *Scaevola* suggests that this loss predates the Goodeniaceae
230 split, clarifying previous hypotheses about its timing (Fig. 4c, Supplementary Dataset 3).
231 Duplication patterns consisting of a few segmental genes, with much larger proportions of
232 proximal and tandem genes, mirrored previous research suggesting that this expansion in
233 Asteraceae arose via tandem duplications following genome duplications[23] (Fig. 4c,d,
234 Supplementary Fig. 18), contrasting with Calyceraceae which has fewer tandem and proximal
235 duplicates (Fig. 4d). Also involved in metabolism, the expansion of Asteraceae *CYP71B23*
236 homologs involved few segmental genes, but with higher proportions of tandem, proximal, and

237 dispersed genes (Fig. 4e,f, Supplementary Fig. 19). Overall, these patterns indicate that gene
238 family expansions, particularly those arising through tandem duplications, have shaped
239 metabolic and stress-response diversity in Asteraceae, linking genome duplication and
240 remodeling to the uneven diversification of Asterales.

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

242 Genes containing positively selected sites in all four Asteraceae taxa (n = 586
243 orthogroups) were significantly enriched for the regulation of biosynthetic and metabolic
244 processes (Supplementary Fig. 20). Of these genes, segmental and dispersed duplicates
245 comprised 45% and 37%, respectively (Supplementary Fig. 17). In addition to the expansions in
246 the CYP superfamily—a large and functionally diverse gene family involved in secondary
247 metabolite synthesis, hormone regulation, and stress responses[44,45]—we observed positively
248 selected sites in all Asteraceae taxa for orthogroups related to *CYP96A1*, *CYP72A7*,
249 *CYP706A4*, *CYP706A6*, *CYP72A13*, and *CYP72A15* (Supplementary Dataset 3). Previous
250 studies also revealed strong lineage-specific diversification of CYP genes in Asteraceae,
251 indicating rapid evolutionary shifts, likely reflecting a central role in metabolic innovation[46].
252 To test for lineage-specific positive selection, we used the resulting 586 orthogroups with sites
253 under positive selection in Asteraceae species as input for branch model analyses. Of these, 47
254 showed evidence for episodic positive selection in an Asteraceae lineage (Supplementary
255 Dataset 3). Genes involved in the predominant biological processes associated with this gene
256 set (stress response and regulation of metabolic processes) include: *TMAC2*, a regulator of
257 abscisic acid and salt stress[47]; *WRKY27,* a transcription factor (TF) involved in plant defense
258 and floral development[48]; and *TCP5*, a TF that controls thermomorphogenesis[49], regulates
259 ethylene-mediated petal development and growth[50], and impacts leaf margin patterning[51] (Table
260 1). Other class II TCP TFs, *TCP2* and *TCP24*, also exhibiting episodic positive selection in
261 Asteraceae, regulate key developmental transitions and have been linked to floral diversification
262 in sunflower[52].

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

264 To focus on genes and pathways involved in floral innovation in Asteraceae, we
265 performed branch model selection analyses on a curated set of 335 flowering-associated
266 orthogroups (Supplementary Table 5). We identified 27 orthogroups with episodic positive
267 selection specific to all Asteraceae taxa (Supplementary Dataset 3). Of these, several
268 orthogroups contained genes involved in related pathways (photoperiod, autonomous, and
269 gibberellin) that regulate flowering time (Fig. 5). These genes directly or indirectly suppress or

activate *SOC1*, a floral integrator[53], which activates floral meristem identity genes such as *LFY* (a floral meristem identity transcriptional regulator that promotes flowering transitions[54]) and *AP1* (a floral meristem identity MADS box gene[55]). Notable genes under selection in Asteraceae were: *GI*, which promotes flowering via *FT* through the photoperiod pathway[56,57]; *FVE*, which promotes flowering by suppressing *FLC* in the autonomous pathway[58,59]; *CGA1* and *GNC,* which delay flowering by suppressing *SOC1* in the gibberellin pathway[60,61]; *FD*, which interacts with *FT* to activate *SOC1*; and *AGL24,* which integrates with *SOC1* to link floral signals and promote flowering[62] (Fig. 5, Table 1). Of these, episodic positive selection was only identified in *Acicarpha* homologs of *FD* and *CGA1*. Overall, selection on multiple genes across interconnected pathways likely fine-tuned flowering time regulation in Asteraceae.

Discussion

While the role of WGDs in driving gene expansion, diversification, and selection is evident across our analyses, the number and timing of WGD events within Asterales remain difficult to resolve despite having high-quality genome assemblies. Contrasting results from different approaches (Figs. 2f, 3) highlight the inherent challenges in detecting ancient WGDs, particularly within large and diverse plant families. This complexity is compounded by processes such as biased gene retention and loss, genome rearrangements, chromosomal fission and fusion, incomplete lineage sorting, and hybridization, including allopolyploidy[63,64]. Synteny-based methods, while powerful, often yield incomplete reconstructions of ancient events, especially across distantly related taxa. Similarly, K_S-based approaches face limitations: signals from ancient WGDs may be obscured by subsequent small-scale duplications and losses, attenuated with time, or confounded with speciation signals, especially in cases of allopolyploidy where increased variation in K_S values blurs inference[65].

Our results support multiple hypotheses on WGD placement in Asterales. Triplicate syntenic gene depths echo previous studies indicating a shared paleohexaploidy event between Asteraceae and Calyceraceae[4–6] (Fig. 3a). However, this inference was refined through sub-genome phasing and hypothesis testing to indicate that Asteraceae and Calyceraceae shared a WGD, then each experienced their own (Fig. 3b). Some K_S distribution and gene family expansion results suggest that the core WGD is shared not only by Asteraceae and Calyceraceae, but also Goodeniaceae (Fig. 3c). However, this finding is unlikely given the comparative syntenic gene depths and is probably an artefact of unaccounted biological processes described above[65].

302 Nonetheless, it is evident that both large-scale and small-scale duplications played a
303 formative role in the divergence and diversification of Asterales. While WGDs account for most
304 positively selected genes in Asteraceae, duplications originating from TE activity (i.e., tandem,
305 proximal, and dispersed duplicates) explained most gene family expansions (Supplementary
306 Fig. 17). A biased retention of segmental duplicates from polyploidy, particularly those
307 supporting complex processes and protein interactions, has been observed in other systems,
308 and the high proportion of segmental duplicates in Asteraceae and *Acicarpha* may reflect
309 this[33,66,67]. Rather than being lost or pseudogenized, retained duplicate genes likely underwent
310 subfunctionalization, neofunctionalization, or dosage amplification to increase adaptive fitness.
311 These findings highlight the substantial roles tandem duplication and transposition, alongside
312 WGD, have in generating genomic novelty[68].
313 In Asteraceae, post-duplication outcomes may have enhanced ecologically beneficial
314 processes such as secondary metabolite biosynthesis and cellular responses to environmental
315 stimuli, which are enriched in expanded and positively selected gene families (Supplementary
316 Fig. 20). Given Asteraceae's rich array of secondary compounds used in defense, attraction,
317 and environmental interaction[69], the lineage-specific retention and expansion of genes involved
318 in metabolite production and regulation likely reflect specialization in metabolite biosynthesis.
319 Secondary chemistry contributes to key ecological interactions, including herbivore deterrence,
320 pathogen resistance, and pollinator attraction, and its diversification may have enhanced
321 environmental responsiveness and niche breadth[69]. Conversely, in *Acicarpha* the limited
322 expansion of metabolite biosynthesis genes may have limited their geographic range through
323 outcompetition. These patterns of gene family expansion, retention, and selection thus support
324 the broader view that chemical diversification, driven by underlying genomic innovation, has
325 been a major driver of Asteraceae's evolutionary success[5].
326 Advanced reproductive systems and strategies have also played a key role in
327 Asteraceae's evolutionary and ecological success. In a capitulum, flowering time and individual
328 floret development is precise and highly coordinated[15,16]. Our findings suggest that selective
329 pressures on duplicated genes and transposon activity may have fine-tuned flowering time gene
330 regulation , conferring a key reproductive advantage in Asteraceae. Among positively selected
331 flowering time genes, patterns of evolution varied. For example, *FD* homologs exhibited a two-
332 step episode of positive selection: first on Asteraceae/*Acicarpha* dispersed duplicates, then
333 again on Asteraceae and *Acicarpha* duplicates individually. The *FVE* homologs suggest that
334 Asteraceae and *Acicarpha* segmental duplicates were derived from a shared WGD, followed by
335 positive selection on only one Asteraceae lineage. These patterns reflect a shared history

336 between Asteraceae and *Acicarpha*. Conversely, *GI* and *CGA1* patterns reflect an Asteraceae-
337 specific WGD. *GI* homologs showed two episodes of diversifying selection in Asteraceae
338 segmental duplicates, likely indicating subfunctionalization or dosage compensation (Fig. 5).
339 Among *CGA1* homologs, however, episodic positive selection occurred only along one of the
340 segmentally duplicated Asteraceae lineages. While likely originating from transposon activity
341 rather than WGD, the dispersed duplicates of *Acicarpha CGA1* homologs exhibited a similar
342 pattern of selection on one duplicated lineage. These findings underscore distinct mechanisms
343 of functional divergence and the fate of duplicates between Asteraceae and *Acicarpha* that may
344 have conferred unique adaptive advantages.
345 Furthermore, the absence of positive selection on core integrators (e.g., *SOC1*) and
346 meristem identity genes (e.g., *LFY* and *AP1*) suggests that strong purifying selection preserves
347 central flowering functions, while upstream regulators offer an evolutionary interface for
348 adjusting developmental timing. This pattern suggests that adaptive evolution in Asteraceae has
349 primarily targeted regulatory modifications within the flowering gene network, enabling flexibility
350 without disrupting deeply conserved functions. These patterns imply lineage-specific fine-tuning
351 of flowering time, likely in response to environmental heterogeneity. Such adjustments may
352 have been especially advantageous during the late Paleocene-Eocene, when Asteraceae
353 experienced climatic optimums, increasing aridity, and concurrent pollinator radiations[8,70,71],
354 favoring genes capable of modulating reproductive timing and capitulum organization. Because
355 some genes in traditional plant model systems are lineage-specific, e.g., *FLC* in *Arabidopsis*
356 (Supplementary Fig. 21), these species may not serve as broadly representative models for
357 flowering time regulation. Instead, groups like Asteraceae and other diverse clades offer
358 additional insights into the evolutionary dynamics of these pathways.
359 Genome duplication has influenced not only speciation, but also the phenotypic and
360 ecological breadth of plants, shaping habitat use, life history strategies, competitive interactions,
361 responses to herbivores and pathogens, and pollinator relationships[72–77]. In both Asteraceae
362 and Calyceraceae, ancient large- and small-scale duplications left a strong genomic imprint, yet
363 distinct evolutionary pressures acting on duplicated genes, particularly those involved in
364 metabolite biosynthesis and flowering time regulation, drove the emergence of key ecological
365 traits and ultimately enabled the extraordinary diversification of Asteraceae.

366 **Online Methods**

367 **Taxon selection**

368 For this study, we generated de novo chromosome-level assemblies of *Acicarpha*
369 *procumbens* Less. (Calyceraceae), hereafter *Acicarpha*; *Scaevola aemula* R.Br.
370 (Goodeniaceae), hereafter *Scaevola*; and *Platycodon grandiflorus* (Jacq.) A.DC. 'Komachi'
371 (Campanulaceae), hereafter *Platycodon*, using the materials and methods described below. We
372 included additional publicly available Asterales species for comparison: *Arctium lappa* L.[78]
373 (Asteraceae; NCBI: PRJNA764011), hereafter *Arctium*; *Lactuca sativa* L.[26] (Asteraceae; NCBI:
374 PRJNA173551), hereafter *Lactuca*; *Chrysanthemum lavandulifolium* (Fisch. ex Trautv.)
375 Makino[79], hereafter *Chrysanthemum* (Asteraceae; NCBI: PRJNA681093); *Helianthus annuus*
376 L.[31,80,81] (Asteraceae; HA412-HO v2: https://sunflowergenome.org/), hereafter *Helianthus*; and
377 *Nymphoides indica* (L.) Kuntze[82] (Menyanthaceae; CNGB: CNP0002767), hereafter
378 *Nymphoides*. Additionally, for broader evolutionary and functional comparisons, we included the
379 well-studied genomes: *Vitis vinifera* L. cv. Cabernet Sauvignon cl. 08 v 1.1[83] (Vitaceae) ,
380 hereafter *Vitis,* and *Arabidopsis thaliana* (L.) Heynh. (Brassicaceae; TAIR10;
381 https://v2.arabidopsis.org/), hereafter *Arabidopsis* (Supplementary Table 6).
382 We collected seeds of *Acicarpha* in Uruguay, purchased seeds of *Platycodon* through
383 Park Seed (https://parkseed.com/), and acquired cuttings of *Scaevola* through North Carolina
384 Farms Inc. (https://www.ncfarmsinc.com/). We grew all plants in greenhouses at the University
385 of Memphis (Memphis, Tennessee, USA) and Auburn University (Auburn, Alabama, USA). A
386 single individual of *Platycodon* was taken through two successive rounds of inbreeding prior to
387 use. *Acicarpha* and *Scaevola* individuals were not inbred before sequencing. Vouchers for each
388 specimen of *Acicarpha*, *Scaevola*, and *Platycodon* are available at the University of Memphis
389 herbarium (MEM) and Auburn University Herbarium (AUA) (see Supplementary Table 7 for
390 details).

391 **Sequencing**

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

401 Flash-frozen tissue was sent to the HudsonAlpha Institute for Biotechnology (Huntsville,
402 Alabama, USA) for Omni-C library preparation using the Dovetail Omni-C kit (Dovetail
403 Genomics, Inc.). The gDNA, Omni-C libraries, and the RNA were sequenced using an Illumina
404 NovaSeq 6000 with PE150 chemistry, the HMW DNA was sequenced on a PacBio Sequel II at
405 HudsonAlpha Institute for Biotechnology.
406 For *Scaevola*, short- and long-read DNA, Hi-C, and RNA data were generated using
407 equivalent tissue amounts and extraction kits as for *Acicarpha*. Hi-C libraries were prepared with
408 the Proximo Hi-C Kit-Plant (Phase Genomics, cat #KT3040) and sequenced at the University of
409 North Carolina's High Throughput Sequencing Facility (UNC HTSF) on a NovaSeq 6000 (PE-50
410 and PE-100). The gDNA and HMW DNA were sequenced on NovaSeq PE150 and PacBio
411 Sequel II sequencers, respectively, at HudsonAlpha Institute for Biotechnology.
412 For *Platycodon*, we generated short- and long-read DNA, Hi-C, and RNA sequencing
413 data. Genomic DNA was extracted from ~100 mg of homogenized tissue using a modified
414 CTAB protocol. Short-read libraries were prepared at UNC's HTSF with the KAPA Hyper DNA
415 kit and sequenced at BGI-Genomics on the DNBseq platform (PE100). Total RNA from YL and
416 SAM tissues of five individuals was extracted using the E.Z.N.A. Plant RNA kit (Omega Bio-tek,
417 #R6827-01), treated with RNase-free DNase (#E1091), and used for library preparation (KAPA
418 stranded mRNA-seq kit, UNC HTSF). Hi-C libraries were prepared with the Proximo Hi-C Kit-
419 Plant (Phase Genomics, #KT3040) following the v4.0 protocol (October 2020) and sequenced at
420 UNC HTSF on a NovaSeq 6000 (PE100). For long reads, HMW DNA was extracted using the
421 ONT "High molecular weight gDNA extraction from plant leaves" (CTAB–Genomic-tip) protocol,
422 purified with Amicon Ultra filters (#UFC510008), and size-selected using the Circulomics Short
423 Read Eliminator Kit (#SS-100-101-01, v2). ONT libraries were prepared with the Ligation
424 Sequencing Kit (#SQK-LSK109) following previously described modifications[124] and sequenced
425 on a MinION Mk1B (MIN-101B) using FLO-MIN106D flow cells (R9.4.1 pore) controlled by
426 MinKNOW v3.6.5. Two flow cells were used; each washed three times with the Flow Cell Wash
427 Kit (#EXP-WSH003) and reloaded with the same library.
428 We checked raw long-read and short-read sequence data quality using FastQC[84] v.
429 0.11.9, a program designed to highlight potential issues with sequence data. Additionally, we
430 checked raw long-read data quality using PAUVRE[85] v. 0.2.3, which plots the length and quality
431 distribution of raw Oxford Nanopore (ONT) or PacBio reads (Supplementary Fig. 19). We then
432 trimmed raw reads to filter out inadequate reads and remove adapters using fastp[86] v. 0.23.2.

433 **Genome assembly and scaffolding**

434 We assembled long-read data for *Scaevola* and *Acicarpha* using HifiASM[87] v. 0.19.5 and
435 v. 0.18.5, respectively, with Hi-C data for *Scaevola* and Omni-C data for *Acicarpha* integrated
436 into the assemblies, resulting in two haplotypes per species. We base-called the *Platycodon*
437 MinION data using the python package guppy v. 3.1.6 (ONT) with model R9.4.1_450bps in
438 high-accuracy mode. We then assembled the genome with a hybrid approach utilizing ONT
439 long-read and DNBseq (BGI) short-read genomic DNA sequencing as inputs. For the hybrid
440 assembly, we used MaSuRCA[88] v. 3.4.2, and polished with short-read data using bowtie2[89] v.
441 2.1.4 and pilon[90] v. 1.22.
442 We aligned Omni-C and Hi-C reads to the assembly using BWA[91,92] v. 0.7.17 and used
443 Samtools[93] v1.11 to convert the output sam file into a bam file, sorted the bam file, and created
444 an index for the assembly fasta file to prepare for scaffolding. We conducted scaffolding,
445 followed by manual curation of the *Scaevola*, *Acicarpha*, and *Platycodon* assemblies, using Yet
446 another Hi-C Scaffolding tool (YaHS)[94] v1.1 for *Acicarpha* and v. 1.2a.2 for *Platycodon* and
447 *Scaevola*, with the Omni-C data for *Acicarpha*, Hi-C data for the *Scaevola*, and *Platycodon* used
448 as input. We prepared the resulting assemblies for Juicebox[95] (v. 2.17.00 for *Acicarpha* and v.
449 2.15.00 for *Scaevola* and *Platycodon*) visualization using the Juicer Pre tools v1.1 included in
450 YaHS before inputting them into the Juicebox GUI for manual scaffolding and correction to
451 produce pseudochromosomes. We assessed genome assembly quality and completeness
452 using Benchmarking Universal Single-Copy Orthologs (BUSCO)[96] v. 5.4.7 with embryophyta
453 from OrthoDB (www.orthodb.org) as the reference lineage. We then generated assembly
454 statistics of the finalized assemblies using the publicly available python script
455 (https://github.com/lexnederbragt/sequencetools/blob/master/assemblathon_stats.py).

456 **Genome annotation**

457 We screened the *Acicarpha*, *Scaevola*, and *Platycodon* genomes for bacterial
458 contaminants using StandardDB from Kraken2[97] v. 2.0.8 and removed all contigs with positive
459 hits. Using the resulting files as input, we then used the Extensive de-novo TE Annotator
460 (EDTA)[98] pipeline v. 1.9.3 to run RepeatModeler2[99] and identify remaining TEs utilizing the
461 default parameters, except --overwrite, --sensitive, --anno, and --evaluate were set to '1'. Using
462 the decontaminated genome file and the resulting TE library file from EDTA, we then ran
463 RepeatMasker[100] v. 4.1.5 to soft mask the genomes. We annotated the soft masked genomes
464 with RNA sequences using Braker3[101] v. 3.0.3 and labelled functional gene annotations using
465 InterProScan[102] v. 5.66-98.0. We then used OMArk v. 0.3.0 (https://omark.omabrowser.org/,

466 accessed June 3, 2024) to assess the proteome quality of *Acicarpha*, *Scaevola*, and *Platycodon*
467 with Asterids (NCBI ID: 71274) as the ancestral clade.
468 To ensure consistency between annotations and comparative genomic analyses, we re-
469 annotated five publicly available genomes from *Nymphoides, Arctium, Chrysanthemum,*
470 *Helianthus, Lactuca* following the steps detailed above. We assessed annotation quality across
471 all eight taxa using the gff3_stats python script
472 (https://github.com/HuffordLab/GenomeQC/blob/master/scripts/gff3_stats.py) from
473 GenomeQC[103].

474 **Plastid and mitochondrial genome assembly**

475 We extracted chloroplast sequences from all trimmed paired end short read data using
476 the GetOrganelle toolkit[104] following the recommended parameters for embryophyte plant
477 plastomes but increasing the number of rounds and the disentangle time limit (-R 75 --
478 disentangle-time-limit 7200). For long read data, PacBio and ONT minion, we used ptGAUL to
479 extract chloroplast sequences[105]. We then annotated complete circular sequences using
480 GeSeq[106] and assessed collinearity between plastomes using the progressive alignment in
481 Mauve[107]. In addition, we assembled mitochondrial genomes of *Platycodon, Scaevola,* and
482 *Acicarpha* using OatK v. 1.0 (https://github.com/c-zhou/oatk).

483 **Genome structure**

484 We characterized genomic sequences' size, heterozygosity, and repetitiveness by k-mer
485 frequency analyses (k = 21) using Jellyfish[108] and GenomeScope[109] 2.0. We assessed
486 heterozygous k-mer pairs, obtained with KMC3[110], using Smudgeplot[109] to estimate ploidy levels
487 and infer genomic complexity in terms of sub-genome composition. We identified lower (L) and
488 upper (U) end cut-off values, below and above which all k-mers were discarded as errors, using
489 k-mer coverage output from Genomescope 2.0, as recommended in the Smudgeplot
490 documentation (https://github.com/KamilSJaron/smudgeplot), where L = (kcov/2)-5.

491 **Gene duplications**

492 We investigated gene duplications of all eight target Asterales genomes, plus genomes
493 from *Vitis* and *Arabidopsis*, by assessing homology and synteny to identify duplicated genes
494 and genomic regions, then by analyzing K_S values to estimate the relative age of duplication
495 events. To quantify and classify types of gene duplications within each genome, we used the
496 "duplicate gene classifier" program in MCScanX[33]. Through this program, duplicated genes
497 were classified as: segmental, proximal, tandem, or dispersed, as defined by [33].

498 We utilized OrthoFinder v. 2.5.5[111] to identify conserved orthologs and gene duplications
499 across the ten genomes, using the -M msa parameter to generate a multiple species alignment.
500 We then identified the best evolutionary model and performed a maximum likelihood
501 phylogenetic analysis on the alignment using IQTree[112] v. 2.1.3 with 1,000 bootstraps to
502 generate a species tree. We used the orthologous gene groups identified by OrthoFinder for
503 downstream analyses to investigate evolutionary processes, however, we filtered the groups to
504 exclude multiple isoforms of the same gene by keeping only the longest isoform.

505 **Syntenic gene depths**

506 To investigate putative paleopolyploidy, we assessed synteny among the eight Asterales
507 genomes. We used the JCVI[113]v. 1.3.8 python function "compara catalog ortholog" with the
508 default c-score cutoff of 0.7 to identify syntenic regions, then generated genome-wide synteny
509 plots of the full eight genomes following phylogenomic placement. We evaluated pairwise
510 syntenic gene depth between all genomes using the "compara synteny depth" option. We then
511 used these inferred values as input for subsequent analyses to phase subgenomes and further
512 elucidate polyploid origins (detailed in "Phasing subgenomes" below).

513 **Phasing subgenomes**

514 We implemented the two most prominent software tools to identify and separate
515 subgenomes, SubPhaser[114] and Whole-Genome Duplication Integrated (WGDI[115]). However,
516 given that the tested taxa have not experienced recent polyploidization events and are currently
517 recognized as diploids, SubPhaser could not be used with our data as it is unable to accurately
518 phase genomes that experienced paleopolyploidy events. Instead, we used WGDI v. 0.6.5 to
519 further investigate polyploidization events across Asterales using *Vitis* as a reference. Steps to
520 investigate polyploidy followed the subgenome aware analyses available on GitHub
521 (https://github.com/SunPengChuan/wgdi-example/blob/main/phase_subgenomes.md). Multiple
522 chromosomes of *Vitis* were tested as a reference for phasing, though only the chromosome with
523 the largest gene retention and collinearity, chromosome 11, was used for further analyses. Age
524 estimates from K_S values (e.g., Supplementary Fig. 23) were used to help separate
525 subgenomes if they were one the same chromosome or were not distinct. We determined
526 phylogenomic relationships among the subgenomes using WGDI's '-at' parameter to generate
527 gene trees with IQTree and then ASTRAL-III[116] v. 5.7.8 to generate a final subgenome
528 coalescent tree. We also assessed gene fractionation within each subgenome to determine the
529 evolutionary age of each subgenome using the 'r' parameter of WGDI. We then reanalyzed the

530 resulting subgenomes within Asteraceae, Calyceraceae, and Goodeniaceae in JCVI to
531 investigate synteny across these phased regions using methods detailed above.

532 **Accelerated gene evolution in *Acicarpha* and Asteraceae**

533 To identify major targets of accelerated evolution within conserved genomic regions, we
534 used Cactus[117–119] v2.5.1, to generate a reference-free alignment of the Asterales subgenomes
535 that are syntenic to chromosome 11 of *Vitis*. Using the inferred phylogenomic relationship
536 between subgenomes as the guide tree to inform hierarchical alignment, we performed three
537 separate tests for accelerated evolution at different positions in the subgenome tree: across all
538 Asteraceae and *Acicarpha* subgenomes (clade 1; Supplementary Fig. 11), in the *Acicarpha*-
539 specific duplication event (clade 2; Supplementary Fig. 11), and in *Acicarpha* within the
540 Subgenome 1 clade (clade 3; Supplementary Fig. 11).

541 Specifically, we created a neutral model of sequence evolution based on ancestral
542 repetitive sequences[120]. Repetitive elements were identified within the most ancestral sequence
543 in the subgenome alignment using RepeatModeler and RepeatMasker[100]. A random sample of
544 100,000 sites from these ancestral repeats was then selected to generate a multiple alignment
545 file (MAF) of repetitive sites across all subgenomes. This MAF was used to construct the neutral
546 model with PHAST phyloFit[121,122], using default settings and the same phylogeny used to
547 generate the subgenome alignment (Supplementary Fig. 11).

548 With this neutral model, we identified conserved regions within the Asterales
549 subgenomes using PHAST phastCons[121,122], with *Vitis* as the reference. This was performed
550 separately for each of the three clade-specific analyses, with conserved regions detected in the
551 outgroup subgenomes. For each analysis, we obtained a list of the most-conserved elements
552 and converted their log-odds scores to length-normalized log-odds ratios. Elements with
553 normalized log-odds ratios below the 30th percentile were excluded from further analyses[123].
554 This filtering step produced three distinct lists of conserved regions in outgroup subgenomes,
555 allowing us to assess accelerated evolution within our clades of interest. We then tested for GC-
556 biased gene conversion (gBGC) within our ingroups using PHAST phastBias[121] and considered
557 any region with a score >0.5 to be impacted by gBGC. We filtered our lists of conserved regions
558 to remove any elements that were impacted by gBGC.

559 We used the likelihood ratio test of PHAST phyloP[121,124] to test the conserved elements
560 for accelerated evolution within our clades of interest. We corrected for multiple hypothesis
561 testing using Storey's correction method[125] (FDR = 0.05). This produced a separate list of
562 accelerated regions (ARs) for each of our clade-specific analyses. To assess the biological

563 impacts of the accelerated evolution detected in each of our analyses, we identified the genes
564 most likely to be impacted by each AR. We associated an AR with a gene if it overlapped an
565 exon. Exon identities and locations were obtained from annotations of the *Vitis* genome[83].
566 ARs can be distributed unevenly across the genome, often clustering into regions of high
567 density, known as AR hotspots. To detect these hotspots, we employed the permutation method
568 of Cahill et al.[126]. Specifically, we used a sliding window approach, counting ARs in 100 kb
569 windows along *Vitis* chromosome 11, with windows offset by 10 kb steps. We conducted 100
570 permutations, each time randomly designating N conserved elements as accelerated, where N
571 corresponds to the actual number of ARs on the chromosome. For each permutation, we
572 repeated the sliding window approach to determine the maximum simulated AR density in a 100
573 kb window. Regions were designated as AR hotspots if their true AR density exceeded the 95th
574 percentile of maximum AR densities observed across all permutations. This approach prevents
575 regions with high conserved element density from being falsely identified as hotspots. We
576 identified separate hotspots for each of our clade-specific analyses.

577 **K_S distributions**

578 To investigate the influence of major genome duplication events and positive selection
579 on Asterales genomes, we assessed the non-synonymous (K_a) and synonymous (K_S)
580 substitution rates between gene pairs. Peaks in K_S value distributions, indicating the presence
581 of duplicated genes, were used to infer histories of WGD. For homologous gene pairs, we
582 estimated K_S values using an established pipeline[127] in which orthologs were aligned using
583 MUSCLE[128], then converted into nucleotides using PAL2NAL[129], then K_S values were calculated
584 between gene pairs using the F3X4 model in codeml within PAML[130]. We plotted K_S values of
585 paralogs and pairwise orthologous gene pairs in R[131,132] with kernel density estimations using
586 geom_density() from ggplot2[133]. Additionally, we used output from these analyses to identify
587 signatures of positive selection ($K_a/K_S > 1$), neutral selection ($K_a/K_S \approx 1$), and purifying selection
588 ($K_a/K_S < 1$) of paralogous gene pairs. To more accurately compare estimations of divergence
589 derived from orthologous K_S distributions with estimations of WGD derived from paralogous K_S
590 distributions, we then re-scaled ortholog distributions to the paralog scale of each focal species
591 using *ksrates*[134]. For these analyses, *Platycodon* was used as the outgroup, therefore, WGD
592 events were not assessed for this lineage.

593 **WGD hypothesis testing**

594 Directly following previously published methods and scripts[135,136]
595 (https://gitlab.com/barker-lab/ajb_special_issue_2024.git), we tested predicted WGD events,

596 hypothesized by previous literature and identified by *ksrates*, using WHALE[35]. WHALE is a tool
597 designed to test WGD hypotheses by analyzing the relative support for different evolutionary
598 scenarios. It uses gene tree topologies and considers the expected patterns of gene retention
599 and loss following WGD events. By applying a likelihood-based framework, WHALE evaluates
600 whether observed gene tree discordance is best explained by WGD or alternative models,
601 making it a valuable approach for investigating genome evolution. Given results from *ksrates*
602 and primary literature, we tested nine total WGD events split into three separate analyses: one
603 based solely off findings from previous studies, the second analysis used only the results of
604 *ksrates*, and the third analysis combined results of the first two analyses (Supplementary Fig.
605 15).

606 **Evolution of gene family size**

607 To investigate the evolution of gene family sizes in Asterales, we reconstructed ancestral
608 states of the orthologous groups identified by OrthoFinder by employing the fast algorithm[137]
609 through the 'ace' function in the *phytools*[138] R package. From the reconstructed ancestral states,
610 we applied targeted filters to classify orthogroups as: 1) specific to *Acicarpha*, 2) lost in
611 *Acicarpha*, 3) specific to Asteraceae (evolved in the ancestral branch of Asteraceae and absent
612 from other lineages), 4) lost in the last common ancestor of Asteraceae, 5) specific to both
613 Asteraceae and *Acicarpha*, and 6) lost in the last common ancestor of both Asteraceae and
614 *Acicarpha*. Filtering was based on a Bayes factor threshold of 10, indicating that the likelihood of
615 an orthogroup being present or absent at a given node was ten times higher than the alternative
616 scenario.

617 Next, we used CAFE[139] v. 5.0 to further investigate the dynamics of orthogroup
618 expansion and contraction. Since CAFE requires an ultrametric tree, we used TreePL[140,141] to
619 calibrate the phylogeny reconstructed by OrthoFinder. We selected four calibration points based
620 on previous work[8], constraining the minimum ages of ancestral nodes for the following groups:
621 Asteraceae and Calyceraceae (*Tubulifloridites lillei* Type A[142]; 72.1 Mya), all Asteraceae
622 included in the study (*Raiguenrayun cura*[143]; 47.5 Mya), *Helianthus* and *Lactuca* (*Cichorium*
623 *intybus* type pollen[144]; 22 Mya), and *Helianthus* and *Chrysanthemum* (*Artemisia* type pollen[145];
624 31 Mya). The root age was constrained between 121 and 126 Mya using a secondary
625 calibration point for Pentapetalae[146] (Supplementary Fig. 24). The optimal smoothing parameter
626 was determined through cross-validation, testing 20 values across a range of magnitudes (from
627 1e-10 to 1e+09) to ensure the best fit.

628 The tested CAFE models varied in the number of gamma parameters (ranging from 1 to
629 3), reflecting assumed differences in expansion and contraction rates across orthogroups.
630 Orthogroups identified by the best-fitting model as significantly expanded or contracted (Viterbi
631 ≤ 0.05) in Asteraceae, Acicarpha, and both Asteraceae and Acicarpha were retained for
632 downstream analyses, in addition to those identified as lost or specific to these groups. Genes
633 from significantly expanded orthogroups were then blasted to the *Arabidopsis thaliana*
634 database, TAIR10, using BLASTP[147] to determine closest gene similarity and trees were then
635 visualized in R.

636 **Positive selection**

637 We further investigated positive selection across Asterales genomes, focusing on
638 Asteraceae, using site-specific and branch-specific methods. For all taxa, we assessed
639 orthology and subsequent positive selection of single-copy orthologous groups using site-
640 specific methods within the AlexandrusPS pipeline[148]. For the orthogroups which were found to
641 exhibit positive selection for genes pertaining to all four Asteraceae taxa and no others, we
642 analyzed further by using branch-specific methods to determine if the found signatures of
643 positive selection originated from taxon-specific or family-specific lineages. Additionally, we
644 assessed a subset of orthogroups relating to floral development for lineage-specific positive
645 selection using branch-specific methods. These floral-related orthogroups were chosen based
646 on GO annotations for *Arabidopsis thaliana* genome (TAIR10: available at
647 https://v2.arabidopsis.org/) to include 18 terms related to meristem maintenance and
648 organization (GO:0010582, GO:0009933, GO:0009934, GO:0048507, GO:0010074,
649 GO:0019827), flower development (GO:0009908, GO:0010076, GO:0009910, GO:009911,
650 GO:0048438, GO:0010093, GO:2000032), vegetative shoot system/inflorescence development
651 (GO:0048367, GO:0010077, GO:0010492, GO:0090506,), and vegetative-reproductive phase
652 transition (GO:0048510). For both subsets of orthogroups, we generated codon alignments from
653 coding sequences by adding gaps for frameshifts and replacing stop codons with gaps using
654 MACSE[149]. Then we used the aBSREL (adaptive Branch-Site Random Effects Likelihood)
655 function in HyPhy[150] to test if positive selection has occurred on a proportion of branches. We
656 inspected results using HyPhy Vision (http://vision.hyphy.org/aBSREL), then visualized resulting
657 gene trees in R using Ape 5.0[151] and ggtree[152].

658 **GO enrichment**

659 For inferences on the functional role of genes identified in these analyses, we used the
660 basic local alignment search tool (BLAST[147]) to compare target proteins to a TAIR10 protein

661 database and identify matches based on an e-value threshold of 1e-10. For groups of gene
662 groups identified to be expanded or contracted or under positive selection, we performed gene
663 ontology (GO) enrichment analyses to summarize the significant GO terms for biological
664 processes[153]. We used the clusterProfiler[154] package in R, with annotations based on BLASTP
665 best-hit matches to the TAIR10 database. To account for multiple testing, p-values were
666 adjusted using the Benjamini–Hochberg false discovery rate (FDR) correction. Enrichment was
667 considered significant at an adjusted p-value (P-adjust) threshold of < 0.01.

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

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

680 **Data availability**
681 Raw data and genome assembly data for *Acicarpha*, *Scaevola*, and *Platycodon* are uploaded to
682 NCBI BioProject PRJNA1253301. Accession numbers of these newly developed and publicly
683 available data can be found in Supplementary Table 6. Hardmasked, annotated genome
684 assemblies for *Acicarpha, Scaevola,* and *Platycodon,* as well as the re-annotated assemblies
685 for *Helianthus, Chrysanthemum, Lactuca, Arctium,* and *Nymphoides* were uploaded and made
686 publicly accessible on the Comparative Genomics Platform[152] (id67898, id67904, id67903,
687 id67900, id67899, id67901, id67867, id67902, respectively).

688 **Funding**
689 This work was supported by the National Science Foundation under grants IOS-2214472, IOS-
690 2214473, IOS-2214474, DBI-2319905, and DEB-745197, and by the National Institutes of
691 Health through an R35 MIRA award (R35GM119614).

692 **References**

693 1. Vamosi, J. C., Magallón, S., Mayrose, I., Otto, S. P. & Sauquet, H. Macroevolutionary
694 Patterns of Flowering Plant Speciation and Extinction. *Annu Rev Plant Biol* **69**, 685–706
695 (2018).
696 2. Benítez-Villaseñor, A. *et al.* The use of Anchored Hybrid Enrichment data to resolve
697 higher-level phylogenetic relationships: A proof-of-concept applied to Asterales
698 (Eudicotyledoneae; Angiosperms). *Mol Phylogenet Evol* **181**, 107714 (2023).
699 3. Li, H.-T. *et al.* Origin of angiosperms and the puzzle of the Jurassic gap. *Nat Plants* **5**,
700 461–470 (2019).
701 4. Barker, M. S. *et al.* Most Compositae (Asteraceae) are descendants of a paleohexaploid
702 and all share a paleotetraploid ancestor with the Calyceraceae. *Am J Bot* **103**, 1203–
703 1211 (2016).
704 5. Huang, C.-H. *et al.* Multiple Polyploidization Events across Asteraceae with Two Nested
705 Events in the Early History Revealed by Nuclear Phylogenomics. *Mol Biol Evol* **33**, 2820–
706 2835 (2016).
707 6. Zhang, C. *et al.* Phylotranscriptomic insights into Asteraceae diversity, polyploidy, and
708 morphological innovation. *J Integr Plant Biol* **63**, 1273–1293 (2021).
709 7. Palazzesi, L. *et al.* Asteraceae as a model system for evolutionary studies: from fossils to
710 genomes. *Bot. J. Linn. Soc.* **200**, 143–164 (2022).
711 8. Mandel, J. R. *et al.* A fully resolved backbone phylogeny reveals numerous dispersals
712 and explosive diversifications throughout the history of Asteraceae. *Proc Natl Acad Sci U*
713 *S A* **116**, 14083–14088 (2019).
714 9. Panero, J. L. & Crozier, B. S. Macroevolutionary dynamics in the early diversification of
715 Asteraceae. *Mol Phylogenet Evol* **99**, 116–132 (2016).
716 10. Susanna, A. *et al.* The classification of the Compositae: A tribute to Vicki Ann Funk
717 (1947–2019). *Taxon* **69**, 807–814 (2020).
718 11. Diazgranados, M. *et al.* World Checklist of Useful Plant Species. *KNB Data Repository*
719 https://knb.ecoinformatics.org/view/doi:10.5063/F1CV4G34 (2020).
720 12. Lundberg, J. G. Asteraceae and relationships within Asterales. in (2009).
721 13. Brignone, N. F., Mazet, N., Pozner, R. & Denham, S. S. Calyceraceae: Unexpected
722 diversification pattern in the Southern Andes. *Perspect Plant Ecol Evol Syst* **60**, 125744
723 (2023).
724 14. Denham, S. S., Zavala-Gallo, L., Johnson, L. A. & Pozner, R. E. Insights into the
725 phylogeny and evolutionary history of Calyceraceae. *Taxon* **65**, 1328–1344 (2016).
726 15. Zhang, T. & Elomaa, P. Development and evolution of the Asteraceae capitulum. *New*
727 *Phytol* **242**, 33–48 (2024).
728 16. Gurung, V., Muñoz-Gómez, S. & Jones, D. S. Putting heads together: Developmental
729 genetics of the Asteraceae capitulum. *Curr Opin Plant Biol* **81**, 102589 (2024).
730 17. Pozner, R., Zanotti, C. & Johnson, L. A. Evolutionary origin of the Asteraceae capitulum:
731 Insights from Calyceraceae. *Am J Bot* **99**, 1–13 (2012).
732 18. Teeri, T. H. *et al.* Reproductive meristem fates in *Gerbera*. *J Exp Bot* **57**, 3445–3455
733 (2006).
734 19. Flagel, L. E. & Wendel, J. F. Gene duplication and evolutionary novelty in plants. *New*
735 *Phytol* **183**, 557–564 (2009).

736 20. Doyle, J. J. Polyploidy in Legumes. in *Polyploidy and Genome Evolution* 147–180
737 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2012). doi:10.1007/978-3-642-31442-
738 1_9.
739 21. Eric Schranz, M., Mohammadin, S. & Edger, P. P. Ancient whole genome duplications,
740 novelty and diversification: the WGD Radiation Lag-Time Model. *Curr Opin Plant Biol* **15**,
741 147–153 (2012).
742 22. Soltis, D. E. *et al.* Polyploidy and angiosperm diversification. *Am J Bot* **96**, 336–348
743 (2009).
744 23. Shen, F. *et al.* Comparative genomics reveals a unique nitrogen-carbon balance system
745 in Asteraceae. *Nat Commun* **14**, (2023).
746 24. Kong, X. *et al.* Two-step model of paleohexaploidy, ancestral genome reshuffling and
747 plasticity of heat shock response in Asteraceae. *Hortic Res* **10**, (2023).
748 25. Barker, M. S. *et al.* Multiple Paleopolyploidizations during the Evolution of the
749 Compositae Reveal Parallel Patterns of Duplicate Gene Retention after Millions of Years.
750 *Mol Biol Evol* **25**, 2445–2455 (2008).
751 26. Reyes-Chin-Wo, S. *et al.* Genome assembly with in vitro proximity ligation data and
752 whole-genome triplication in lettuce. *Nat Commun* **8**, 14953 (2017).
753 27. Oliver, K. R., McComb, J. A. & Greene, W. K. Transposable Elements: Powerful
754 Contributors to Angiosperm Evolution and Diversity. *Genome Biol Evol* **5**, 1886–1901
755 (2013).
756 28. Fedoroff, N. Transposons and genome evolution in plants. *Proc Natl Acad Sci USA* **97**,
757 7002–7007 (2000).
758 29. Lippman, Z. *et al.* Role of transposable elements in heterochromatin and epigenetic
759 control. *Nature* **430**, 471–476 (2004).
760 30. Staton, S. E. & Burke, J. M. Evolutionary transitions in the Asteraceae coincide with
761 marked shifts in transposable element abundance. *BMC Genomics* **16**, 623 (2015).
762 31. Badouin, H. *et al.* The sunflower genome provides insights into oil metabolism, flowering
763 and Asterid evolution. *Nature* **546**, 148–152 (2017).
764 32. Schwacke, R., Bolger, M. E. & Usadel, B. PubPlant: a continuously updated online
765 resource for sequenced and published plant genomes. Preprint at
766 https://doi.org/10.1101/2025.03.12.642823 (2025).
767 33. Wang, Y. *et al.* MCScanX: a toolkit for detection and evolutionary analysis of gene
768 synteny and collinearity. *Nucleic Acids Res* **40**, e49–e49 (2012).
769 34. Jiao, Y. *et al.* A genome triplication associated with early diversification of the core
770 eudicots. *Genome Biol* **13**, R3 (2012).
771 35. Zwaenepoel, A. & Van de Peer, Y. Inference of Ancient Whole-Genome Duplications and
772 the Evolution of Gene Duplication and Loss Rates. *Mol Biol Evol* **36**, 1384–1404 (2019).
773 36. Zhang, C. *et al.* Asterid Phylogenomics/Phylotranscriptomics Uncover Morphological
774 Evolutionary Histories and Support Phylogenetic Placement for Numerous Whole-
775 Genome Duplications. *Mol Biol Evol* **37**, 3188–3210 (2020).
776 37. Jin, Y. *et al.* Functional divergence of *GLP* genes between *G. barbadense* and *G.*
777 *hirsutum* in response to *Verticillium dahliae* infection. *Genomics* **114**, 110470 (2022).

778 38. Wang, Z. *et al.* Comparative genomics analysis of WAK/WAKL family in Rosaceae
779 identify candidate WAKs involved in the resistance to *Botrytis cinerea*. *BMC Genomics*
780 **24**, 337 (2023).
781 39. Dar, A. A., Choudhury, A. R., Kancharla, P. K. & Arumugam, N. The *FAD2* Gene in
782 Plants: Occurrence, Regulation, and Role. *Front Plant Sci* **8**, (2017).
783 40. Li, Y. & Wei, K. Comparative functional genomics analysis of cytochrome P450 gene
784 superfamily in wheat and maize. *BMC Plant Biol* **20**, 93 (2020).
785 41. Quezada, E.-H. *et al.* Cysteine-Rich Receptor-Like Kinase Gene Family Identification in
786 the Phaseolus Genome and Comparative Analysis of Their Expression Profiles Specific
787 to Mycorrhizal and Rhizobial Symbiosis. *Genes (Basel)* **10**, 59 (2019).
788 42. Feng, T. *et al. FAD2* Gene Radiation and Positive Selection Contributed to Polyacetylene
789 Metabolism Evolution in Campanulids. *Plant Physiol* **181**, 714–728 (2019).
790 43. Dong, Y. *et al.* The *Carthamus tinctorius* L. genome sequence provides insights into
791 synthesis of unsaturated fatty acids. *BMC Genomics* **25**, 510 (2024).
792 44. Pandian, B. A., Sathishraj, R., Djanaguiraman, M., Prasad, P. V. V. & Jugulam, M. Role
793 of Cytochrome P450 Enzymes in Plant Stress Response. *Antioxidants* **9**, 454 (2020).
794 45. Chakraborty, P., Biswas, A., Dey, S., Bhattacharjee, T. & Chakrabarty, S. Cytochrome
795 P450 Gene Families: Role in Plant Secondary Metabolites Production and Plant Defense.
796 *J Xenobiot* **13**, 402–423 (2023).
797 46. Chen, Y., Klinkhamer, P. G. L., Memelink, J. & Vrieling, K. Diversity and evolution of
798 cytochrome P450s of *Jacobaea vulgaris* and *Jacobaea aquatica*. *BMC Plant Biol* **20**, 342
799 (2020).
800 47. Alshareef, N. O. *et al.* Root remodeling mechanisms and salt tolerance trade-offs: The
801 roles of HKT1, TMAC2, and TIP2;2 in Arabidopsis. *PLoS Genet* **21**, e1011713 (2025).
802 48. Mukhtar, M. S., Deslandes, L., Auriac, M., Marco, Y. & Somssich, I. E. The Arabidopsis
803 transcription factor WRKY27 influences wilt disease symptom development caused by
804 *Ralstonia solanacearum*. *The Plant Journal* **56**, 935–947 (2008).
805 49. Han, X. *et al. Arabidopsis* Transcription Factor TCP5 Controls Plant
806 Thermomorphogenesis by Positively Regulating PIF4 Activity. *iScience* **15**, 611–622
807 (2019).
808 50. van Es, S. W. *et al.* Novel functions of the Arabidopsis transcription factor TCP5 in petal
809 development and ethylene biosynthesis. *The Plant Journal* **94**, 867–879 (2018).
810 51. Yu, H. *et al.* TCP5 controls leaf margin development by regulating KNOX and BEL-like
811 transcription factors in Arabidopsis. *J Exp Bot* **72**, 1809–1821 (2021).
812 52. Chapman, M. A., Leebens-Mack, J. H. & Burke, J. M. Positive Selection and Expression
813 Divergence Following Gene Duplication in the Sunflower *CYCLOIDEA* Gene Family. *Mol*
814 *Biol Evol* **25**, 1260–1273 (2008).
815 53. Lee, J. & Lee, I. Regulation and function of SOC1, a flowering pathway integrator. *J Exp*
816 *Bot* **61**, 2247–2254 (2010).
817 54. Blackman, B. K. *et al.* Contributions of Flowering Time Genes to Sunflower Domestication
818 and Improvement. *Genetics* **187**, 271–287 (2011).
819 55. Li, D. *et al.* Arabidopsis Class II TCP Transcription Factors Integrate with the FT–FD
820 Module to Control Flowering. *Plant Physiol* **181**, 97–111 (2019).

821 56. Sawa, M. & Kay, S. A. GIGANTEA directly activates *Flowering Locus T* in *Arabidopsis*
822 *thaliana*. *Proc Natl Acad Sci USA* **108**, 11698–11703 (2011).
823 57. Sawa, M., Nusinow, D. A., Kay, S. A. & Imaizumi, T. FKF1 and GIGANTEA Complex
824 Formation Is Required for Day-Length Measurement in *Arabidopsis*. *Science (1979)* **318**,
825 261–265 (2007).
826 58. Xie, H. *et al.* DNA Methylation of the Autonomous Pathway Is Associated with Flowering
827 Time Variations in *Arabidopsis thaliana*. *Int J Mol Sci* **25**, 7478 (2024).
828 59. Simpson, G. G. The autonomous pathway: epigenetic and post-transcriptional gene
829 regulation in the control of Arabidopsis flowering time. *Curr Opin Plant Biol* **7**, 570–574
830 (2004).
831 60. Richter, R., Bastakis, E. & Schwechheimer, C. Cross-Repressive Interactions between
832 SOC1 and the GATAs GNC and GNL/CGA1 in the Control of Greening, Cold Tolerance,
833 and Flowering Time in Arabidopsis. *Plant Physiol* **162**, 1992–2004 (2013).
834 61. Bellucci, E. *et al.* Selection and adaptive introgression guided the complex evolutionary
835 history of the European common bean. *Nat Commun* **14**, 1908 (2023).
836 62. Huang, N., Tien, H. & Yu, T. Arabidopsis leaf-expressed *AGAMOUS-LIKE 24* mRNA
837 systemically specifies floral meristem differentiation. *New Phytol* **241**, 504–515 (2024).
838 63. Van de Peer, Y., Mizrachi, E. & Marchal, K. The evolutionary significance of polyploidy.
839 *Nat Rev Genet* **18**, 411–424 (2017).
840 64. Soltis, P. S., Marchant, D. B., Van de Peer, Y. & Soltis, D. E. Polyploidy and genome
841 evolution in plants. *Curr Opin Genet Dev* **35**, 119–125 (2015).
842 65. Dunn, T. & Sethuraman, A. Accurate Inference of the Polyploid Continuum Using
843 Forward-Time Simulations. *Mol Biol Evol* **41**, (2024).
844 66. Freeling, M. Bias in Plant Gene Content Following Different Sorts of Duplication: Tandem,
845 Whole-Genome, Segmental, or by Transposition. *Annu Rev Plant Biol* **60**, 433–453
846 (2009).
847 67. Freeling, M. & Thomas, B. C. Gene-balanced duplications, like tetraploidy, provide
848 predictable drive to increase morphological complexity. *Genome Res* **16**, 805–814
849 (2006).
850 68. Qiao, X. *et al.* Different Modes of Gene Duplication Show Divergent Evolutionary Patterns
851 and Contribute Differently to the Expansion of Gene Families Involved in Important Fruit
852 Traits in Pear (Pyrus bretschneideri). *Front Plant Sci* **9**, (2018).
853 69. *Systematics, Evolution, and Biogeography of Compositae*. (International Association for
854 Plant Taxonomy, Vienna, 2009).
855 70. Zachos, J. C., Dickens, G. R. & Zeebe, R. E. An early Cenozoic perspective on
856 greenhouse warming and carbon-cycle dynamics. *Nature* **451**, 279–283 (2008).
857 71. Korasidis, V. A., Wing, S. L., Morse, P. E., Vitek, N. S. & Bloch, J. I. Evidence for
858 increased animal pollination during the Paleocene–Eocene thermal maximum.
859 *Paleobiology* **51**, 574–587 (2025).
860 72. Ramsey, J. & Ramsey, T. S. Ecological studies of polyploidy in the 100 years following its
861 discovery. *Phil. Trans. R. Soc. B* **369**, 20130352 (2014).
862 73. Strong, D. R. & Ayres, D. R. Ecological and Evolutionary Misadventures of *Spartina*.
863 *Annu Rev Ecol Evol Syst* **44**, 389–410 (2013).

74. Martin, S. L. & Husband, B. C. Adaptation of diploid and tetraploid *Chamerion angustifolium* to elevation but not local environment. *Evolution* **67**, 1780–1791 (2013).
75. Boalt, E., Arvanitis, L., Lehtilä, K. & Ehrlén, J. The association among herbivory tolerance, ploidy level, and herbivory pressure in *cardamine pratensis*. *Evol Ecol* **24**, 1101–1113 (2010).
76. Thompson, J. N. & Merg, K. F. Evolution of polyploidy and the diversification of plant-pollinator interactions. *Ecology* **89**, 2197–2206 (2008).
77. Oswald, B. P. & Nuismer, S. L. Neopolyploidy and pathogen resistance. *Proc. R. Soc. B* **274**, 2393–2397 (2007).
78. Fan, W. *et al.* The genomes of chicory, endive, great burdock and yacon provide insights into Asteraceae palaeo-polyploidization history and plant inulin production. *Mol Ecol Resour* **22**, 3124–3140 (2022).
79. Wen, X. *et al.* The *Chrysanthemum lavandulifolium* genome and the molecular mechanism underlying diverse capitulum types. *Hortic Res* **9**, (2022).
80. Todesco, M. *et al.* Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature* **584**, 602–607 (2020).
81. Huang, K. *et al.* The genomics of linkage drag in inbred lines of sunflower. *Proc Natl Acad Sci USA* **120**, (2023).
82. Yang, J.-S., Qian, Z.-H., Shi, T., Li, Z.-Z. & Chen, J.-M. Chromosome-level genome assembly of the aquatic plant *Nymphoides indica* reveals transposable element bursts and NBS-LRR gene family expansion shedding light on its invasiveness. *DNA Res.* **29**, (2022).
83. Minio, A., Cochetel, N., Vondras, A. M., Massonnet, M. & Cantu, D. Assembly of complete diploid-phased chromosomes from draft genome sequences. *G3: Genes, Genomes, Genet.* **12**, (2022).
84. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. Available at https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (2010).
85. Schultz, D., Ebbert, M. & De Coster, W. PAUVRE. Available at https://github.com/conchoecia/pauvre/ (2019).
86. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
87. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods* **18**, 170–175 (2021).
88. Zimin, A. V. *et al.* Hybrid assembly of the large and highly repetitive genome of *Aegilops tauschii*, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. *Genome Res* **27**, 787–792 (2017).
89. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359 (2012).
90. Walker, B. J. *et al.* Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS One* **9**, e112963 (2014).
91. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at https://doi.org/10.48550/arXiv.1303.3997 (2013).
92. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

93. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, (2021).
94. Zhou, C., McCarthy, S. A. & Durbin, R. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics* **39**, (2023).
95. Durand, N. C. *et al.* Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Syst* **3**, 99–101 (2016).
96. Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A. & Zdobnov, E. M. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Mol Biol Evol* **38**, 4647–4654 (2021).
97. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol* **20**, 257 (2019).
98. Ou, S. *et al.* Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol* **20**, 275 (2019).
99. Flynn, J. M. *et al.* RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci USA* **117**, 9451–9457 (2020).
100. Smit, A. F. A., Hubley, R. & Green, P. RepeatMasker Open-4.0. Preprint at http://www.repeatmasker.org (2013).
101. Gabriel, L. *et al.* BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS and TSEBRA. Preprint at https://doi.org/10.1101/2023.06.10.544449 (2023).
102. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
103. Manchanda, N. *et al.* GenomeQC: a quality assessment tool for genome assemblies and gene structure annotations. *BMC Genomics* **21**, 193 (2020).
104. Jin, J. J. *et al.* GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol* **21**, 1–31 (2020).
105. Zhou, W. *et al.* Plastid Genome Assembly Using Long-read data. *Mol Ecol Resour* **23**, 1442–1457 (2023).
106. Tillich, M. *et al.* GeSeq- versatile and accurate annotation of organelle genomes. *Nucleic Acids Res* **45**, (2017).
107. Darling, A. C. E., Mau, B., Blattner, F. R. & Perna, N. T. Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. *Genome Res* **14**, 1394–1403 (2004).
108. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).
109. Ranallo-Benavidez, T. R., Jaron, K. S. & Schatz, M. C. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun* **11**, (2020).
110. Kokot, M., Dlugosz, M. & Deorowicz, S. KMC 3: counting and manipulating k-mer statistics. *Bioinformatics* **33**, 2759–2761 (2017).
111. Emms, D. M. & Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol* **20**, (2019).
112. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* **37**, 1530–1534 (2020).

113. Tang, H. *et al.* JCVI: A versatile toolkit for comparative genomics analysis. *iMeta* **3**, (2024).
114. Jia, K. *et al.* Sub Phaser: a robust allopolyploid subgenome phasing method based on subgenome-specific k-mers. *New Phytol* **235**, 801–809 (2022).
115. Sun, P. *et al.* WGDI: A user-friendly toolkit for evolutionary analyses of whole-genome duplications and ancestral karyotypes. *Mol Plant* **15**, 1841–1851 (2022).
116. Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinf* **19**, 15–30 (2018).
117. Zhang, R.-G., Shang, H.-Y., Jia, K.-H. & Ma, Y.-P. Subgenome phasing for complex allopolyploidy: case-based benchmarking and recommendations. *Briefings Bioinf* **25**, (2023).
118. Paten, B. *et al.* Cactus: Algorithms for genome multiple sequence alignment. *Genome Res* **21**, 1512–1528 (2011).
119. Armstrong, J. *et al.* Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature* **587**, 246–251 (2020).
120. Sullivan, P. F. *et al.* Leveraging base-pair mammalian constraint to understand genetic variation and human disease. *Science* **380**, (2023).
121. Hubisz, M. J., Pollard, K. S. & Siepel, A. PHAST and RPHAST: phylogenetic analysis with space/time models. *Briefings Bioinf* **12**, 41–51 (2011).
122. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* **15**, 1034–1050 (2005).
123. Keough, K. C. *et al.* Three-dimensional genome rewiring in loci with human accelerated regions. *Science* **380**, (2023).
124. 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).
125. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* **100**, 9440–9445 (2003).
126. 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).
127. Yocca, A. *et al.* A chromosome-scale assembly for ‘d’Anjou’ pear. *G3: Genes, Genomes, Genet.* **14**, (2024).
128. Edgar, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792–1797 (2004).
129. Suyama, M., Torrents, D. & Bork, P. PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res* **34**, (2006).
130. 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).
131. 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. Available at https://stat.ethz.ch/pipermail/r-help/2008-May/161481.html.

132. R Core Team. R: A language and environment for statistical computing. Preprint at https://www.r-project.org/ (2017).
133. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag, New York, 2016).
134. Sensalari, C., Maere, S. & Lohaus, R. ksrates: positioning whole-genome duplications relative to speciation events in Ks distributions. *Bioinformatics* **38**, 530–532 (2022).
135. McKibben, M. T. W., Finch, G. & Barker, M. S. Species-tree topology impacts the inference of ancient whole-genome duplications across the angiosperm phylogeny. *Am J Bot* **111**, (2024).
136. Chen, H. *et al.* Revisiting ancient polyploidy in leptosporangiate ferns. *New Phytol* **237**, 1405–1417 (2023).
137. Pupko, T., Pe, I., Shamir, R. & Graur, D. A Fast Algorithm for Joint Reconstruction of Ancestral Amino Acid Sequences. *Mol Biol Evol* **17**, 890–896 (2000).
138. Revell, L. J. phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol* **3**, 217–223 (2012).
139. De Bie, T., Cristianini, N., Demuth, J. P. & Hahn, M. W. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* **22**, 1269–1271 (2006).
140. Maurin, K. J. L. An empirical guide for producing a dated phylogeny with treePL in a maximum likelihood framework. (2020).
141. Smith, S. A. & O'Meara, B. C. treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* **28**, 2689–2690 (2012).
142. Barreda, V. D. *et al.* Early evolution of the angiosperm clade Asteraceae in the Cretaceous of Antarctica. *Proc Natl Acad Sci USA* **112**, 10989–10994 (2015).
143. Barreda, V. D. *et al.* An extinct Eocene taxon of the daisy family (Asteraceae): evolutionary, ecological and biogeographical implications. *Ann Bot* **109**, 127–134 (2012).
144. Hochuli, P. A. *Palynologische Untersuchungen Im Oligozän Und Untermiozän Der Zentralen Und Westlichen Paratethys*. (Geologischen Institut der Eidg. Technischen Hochschule und der Universtät Zürich, Zurich, 1978).
145. Beentje, H. J. & Caligari, P. D. S. *Compositae: Proceedings of the International Compositae Conference, Kew, 1994*. (Royal Botanic Gardens, 1996).
146. Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L. L. & Hernández-Hernández, T. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytol* **207**, 437–453 (2015).
147. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J Mol Biol* **215**, 403–410 (1990).
148. Ceron-Noriega, A., Schoonenberg, V. A. C., Butter, F. & Levin, M. AlexandrusPS: A User-Friendly Pipeline for the Automated Detection of Orthologous Gene Clusters and Subsequent Positive Selection Analysis. *Genome Biol Evol* **15**, (2023).
149. Ranwez, V., Harispe, S., Delsuc, F. & Douzery, E. J. P. MACSE: Multiple Alignment of Coding SEquences Accounting for Frameshifts and Stop Codons. *PLoS One* **6**, e22594 (2011).
150. Kosakovsky Pond, S. L. *et al.* HyPhy 2.5—A Customizable Platform for Evolutionary Hypothesis Testing Using Phylogenies. *Mol Biol Evol* **37**, 295–299 (2020).

1037 151. Paradis, E. & Schliep, K. Ape 5.0: An environment for modern phylogenetics and
1038 evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).
1039 152. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T. ggtree: an r package for visualization
1040 and annotation of phylogenetic trees with their covariates and other associated data.
1041 *Methods Ecol Evol* **8**, 28–36 (2017).
1042 153. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat Genet* **25**, 25–
1043 29 (2000).
1044 154. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R Package for Comparing
1045 Biological Themes Among Gene Clusters. *OMICS* **16**, 284–287 (2012).
1046 155. Yuan, B. *et al.* Identification and functional characterization of the *RPP13* gene family in
1047 potato (*Solanum tuberosum* L.) for disease resistance. *Front Plant Sci* **15**, (2025).
1048 156. Diaz-Ramirez, D. *et al.* Expression and Functional Analyses of the *WIP* Gene Family in
1049 Arabidopsis. *Plants* **11**, 2010 (2022).
1050 157. Tan, Y. *et al.* Functional Characterization of UDP-Glycosyltransferases Involved in Anti-
1051 viral Lignan Glycosides Biosynthesis in *Isatis indigotica*. *Front Plant Sci* **13**, (2022).
1052 158. Zhou, F.-Y. *et al.* Cytochrome *CYP72A15* may play a role in metabolic resistance to
1053 mesotrione in wild radish. *Pestic Biochem Physiol* **210**, 106380 (2025).
1054 159. Liu, Y., Koornneef, M. & Soppe, W. J. J. The Absence of Histone H2B Monoubiquitination
1055 in the Arabidopsis hub1 (rdo4) Mutant Reveals a Role for Chromatin Remodeling in Seed
1056 Dormancy. *Plant Cell* **19**, 433–444 (2007).
1057 160. Nath, K., O'Donnell, J. P. & Lu, Y. Chloroplastic iron-sulfur scaffold protein NFU3 is
1058 essential to overall plant fitness. *Plant Signal Behav* **12**, e1282023 (2017).
1059 161. Cao, Y. *et al.* Functional characterization of *NBS-LRR* genes reveals an *NBS-LRR* gene
1060 that mediates resistance against *Fusarium* wilt. *BMC Biol* **22**, 45 (2024).
1061 162. Li, C. *et al. OsSAUR2*, a small auxin-up RNA gene, is crucial for arsenic tolerance and
1062 accumulation in rice. *Environ Exp Bot* **226**, 105894 (2024).
1063 163. Prall, W., Hendy, O. & Thornton, L. E. Utility of a Phylogenetic Perspective in Structural
1064 Analysis of CYP72A Enzymes from Flowering Plants. *PLoS One* **11**, e0163024 (2016).
1065 164. Bakshi, M. & Oelmüller, R. WRKY transcription factors. *Plant Signal Behav* **9**, e27700
1066 (2014).
1067 165. Baek, I.-S., Park, H.-Y., You, M. K., Lee, J. H. & Kim, J.-K. Functional Conservation and
1068 Divergence of FVE Genes that Control Flowering Time and Cold Response in Rice and
1069 Arabidopsis. *Mol Cells* **26**, 368–372 (2008).
1070 166. Chiang, Y.-H. *et al.* Functional Characterization of the GATA Transcription Factors GNC
1071 and CGA1 Reveals Their Key Role in Chloroplast Development, Growth, and Division in
1072 Arabidopsis. *Plant Physiol* **160**, 332–348 (2012).
1073 167. Yang, H. *et al.* Multifaceted roles of *FD* gene family in flowering, plant architecture, and
1074 adaptation. *Front Plant Sci* **16**, (2025).
1075 168. GBIF.org User. Occurrence Download. https://doi.org/10.15468/DL.RJ2PNP (2025).
1076 169. GBIF.org User. Occurrence Download. https://doi.org/10.15468/DL.BRQ5TF (2025).
1077 170. GBIF.org User. Occurrence Download. https://doi.org/10.15468/DL.66G87F (2025).
1078

1079 **Table 1. Representative genes from analyses aiming to understand the genomic mechanisms underpinning diversification**
1080 **in Asterales.** Gene symbols were obtained from BLASTP search results to a TAIR10 protein database. Functions were summarized
1081 from representative publications. Asterisk indicates that this gene was the best match for six different orthogroups. Complete results
1082 are listed in Supplementary Dataset 3.

Category	Gene	Description	Involved in
Top ten expanded orthogroups in Asteraceae	*GLP9*	germin-like protein 9	disease resistance[37]
	WAKL2	wall associated kinase-like 2	cellular communication, disease resistance[38]
	FAD2	fatty acid desaturase 2	fatty acid synthesis[39,42]
	CYP71B23	cytochrome P450, family 71, subfamily B, polypeptide 37	sisquiterpene synthesis and modification[40]
	AT1G43760	DNAse I-like superfamily protein	NA
	RPP13	NB-ARC domain-containing disease resistance protein	disease resistance[155]
	WIP4	Wound-responsive family protein	root development[156]
	HYR1	UDP-Glycosyltransferase superfamily protein	growth regulation[157]
	CYP72A15	cytochrome P450, family 72, subfamily A, polypeptide 15	merabolic resistance[158]
	HTB9	Histone superfamily protein	chromatin remodeling[159]
Top ten expanded orthogroups in *Acicarpha*	*NFU3*	NFU domain protein 3	plant growth, development, and flowering[160]
	*CRK8**	cysteine-rich RECEPTOR-like kinase	abiotic and biotic stress response[41]
	LRR4	NB-ARC domain-containing disease resistance protein	disease resistance[161]
	AT2G01050	zinc ion binding; nucleic acid binding	NA
	SAUR2	SAUR-like auxin-responsive protein family	stress response[162]
Lost in Goodeniaceae, Calyceraceae, and Asteraceae	*GLB1*(*PII*)	Nitrogen regulatory protein P-II homolog	nitrogen and metabolic regulation[23]
Positive Selection in Asteraceae	*CYP72A15*	cytochrome P450, family 72, subfamily A, polypeptide 15	secondary metabolism[163]
	TMAC2	AFP2 (ABI five-binding protein 2) family protein	stress response[47]
	TCP2,5,24	TEOSINTE BRANCHED 1, cycloidea and PCF transcription factor 2,5,24)	plant growth and development[50,51,55]
	WRKY6,21,22,27, 31,39,42,57,74,75	WRKY DNA-binding proteins	stress response[164]
	GI	gigantea protein	flowering time regulation[56,57]
	FVE	Transducin family protein / WD-40 repeat family protein	regulation of flowering time and cold response[165]
	CGA1	cytokinin-responsive gata factor 1	regulation of chloroplast development, stress response, and flowering time[60,166]
	FD	Basic-leucine zipper (bZIP) transcription factor family protein	regulation of floral transition, plant development, and environmental response[167]
	AGL24	AGAMOUS-like 24	regulation of flowering time[62]

1083

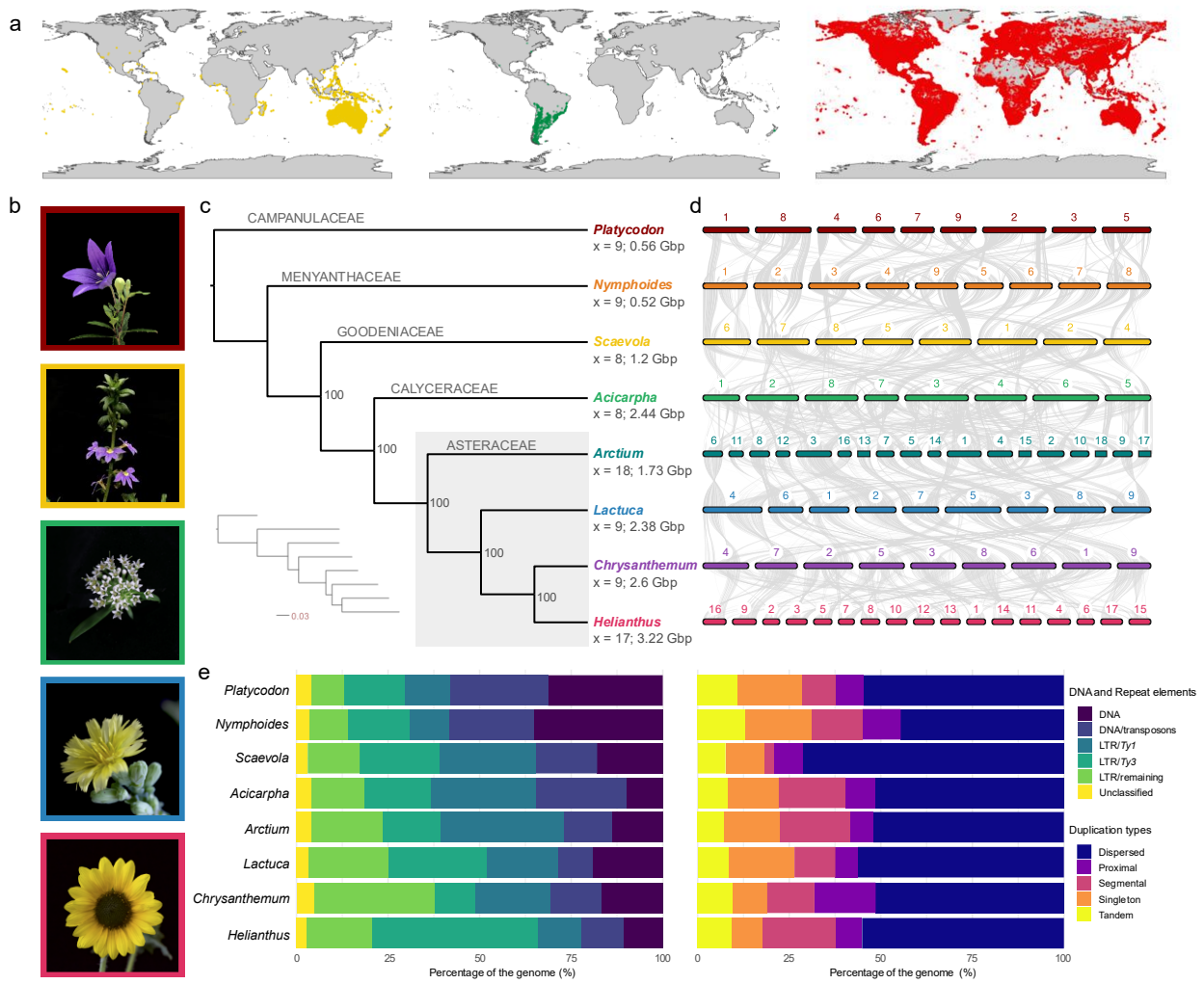

Fig. 1. Genomic characteristics of the eight focal Asterales species. **a)** from left to right, geographic distributions of Goodeniaceae (yellow), Calyceraceae (green), and Asteraceae (red), based on occurrence points of preserved specimens accessed from the Global Biodiversity Information Facility (accessed October 8, 2025)[168–170]. **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 giga-base pairs

1098 (Gbp), respectively. **d)** Synteny plots generated with JCVI ordered by phylogenetic placement of
1099 the eight target taxa. Each bar is representative of a single chromosome, with circled numbers
1100 above the bar indicating chromosome number. Gray lines between species' chromosomes
1101 indicate a single syntenic gene block. **e)** Barplots indicating the percentage of DNA and repeat
1102 elements (left) and gene duplication type (right) across the genomes with percentage of the
1103 genome on the x-axis and species on the y-axis. Colors correspond to the legends to the right.
1104 Taxon order follows phylogenetic placement from **c.**

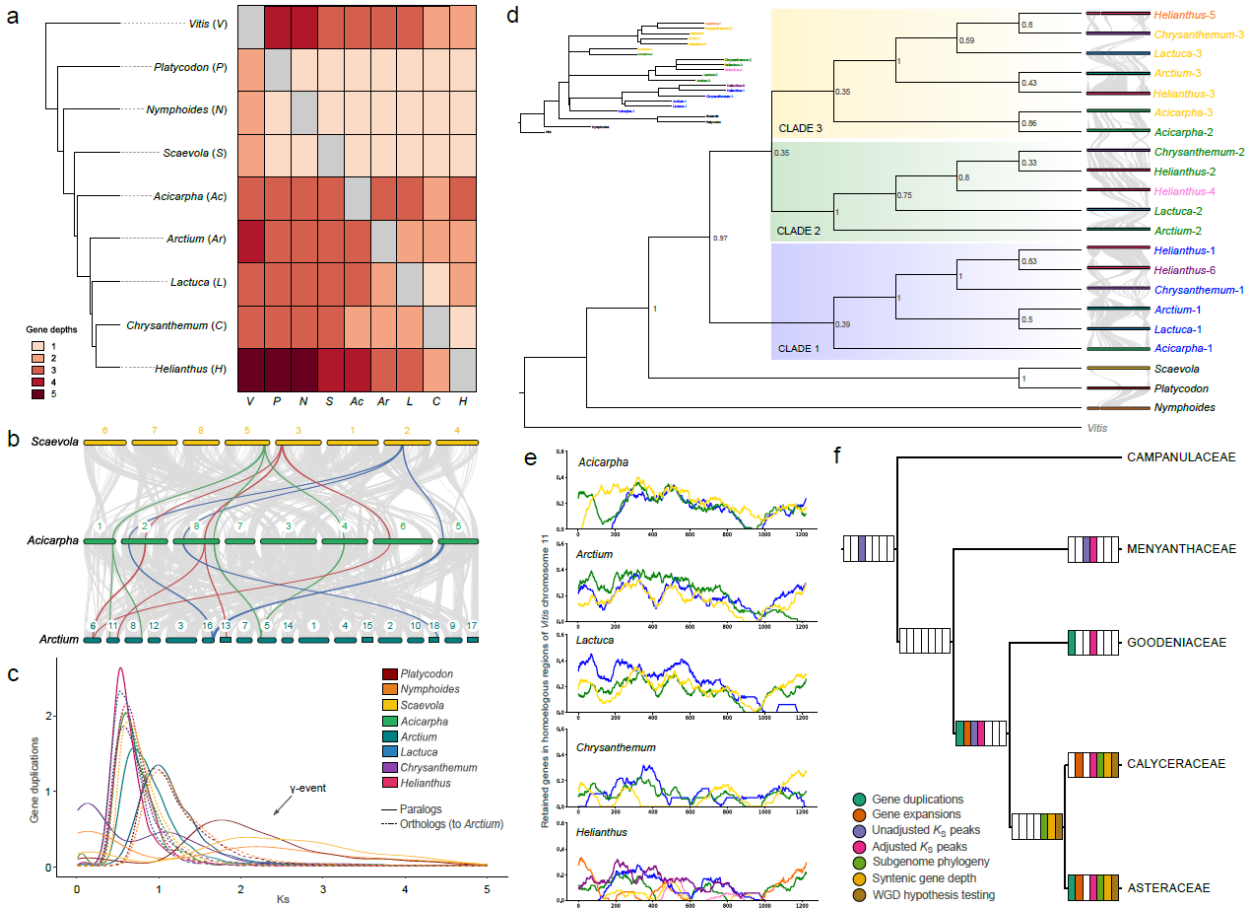

Fig. 2. Evidence of WGD events across the eight tested Asterales taxa. a) Pairwise gene depth ratios from JCVI between the eight Asterales species plus *Vitis*. The top right portion of the matrix indicates the gene depth ratio of the y-axis, while the bottom left portion is for the x-axis (e.g., *Scaevola:Acicarpha* 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 (green, red, and blue) 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. The predicted gamma event is indicated with the arrow. **d)** *Vitis* 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 1d. Local posterior probability (LPP) support values are at each

1122 node. **e)** Genome fractionation results from the subgenome analysis of *Vitis* chromosome 11
1123 from WGDI for *Acicarpha* and four Asteraceae subgenomes. Gene retention is on the y-axis and
1124 base pair number on the x-axis. Colors in the graph correspond to the subgenome numbers
1125 which are also represented as the tip label color in **d**. **f)** Summary of WGD results from across
1126 this study. Colored boxes indicate the findings, as indicated by the legend, that provide
1127 evidence for WGDs on that branch. White boxes indicate no evidence was found for that
1128 analysis. Campanulaceae has no boxes as WGDs were not directly tested for that family.

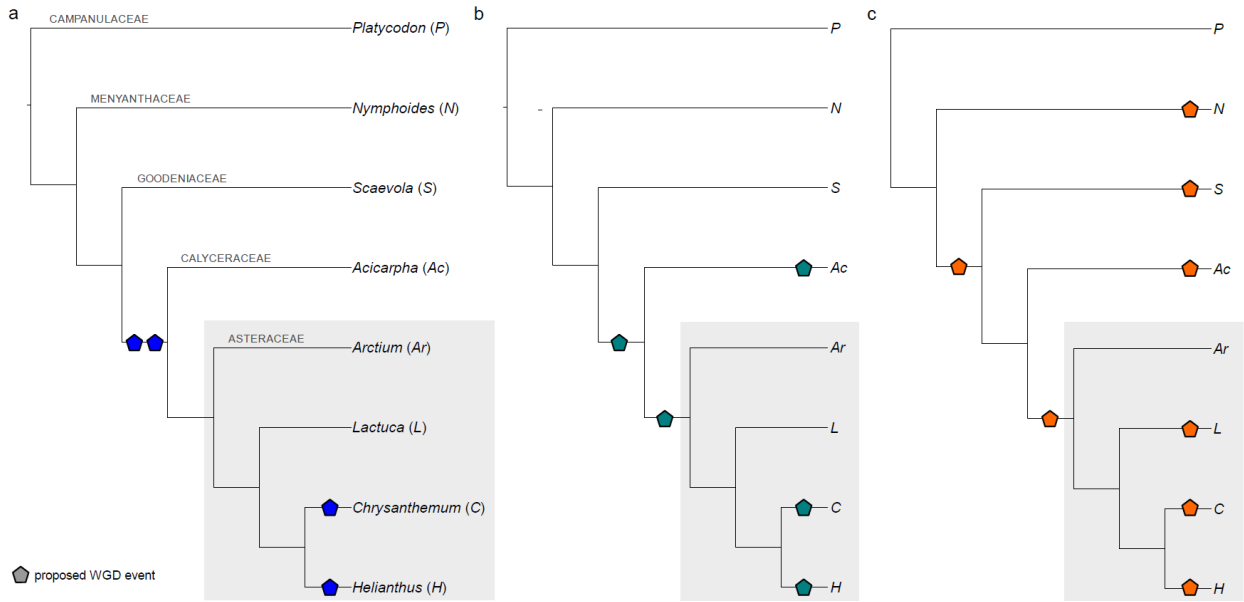

Fig. 3. Possible WGD scenarios in Asterales. A colored pentagon indicates a predicted WGD placement. The scenarios are as follows: **a)** hypothesis for WGT at the crown of Asteraceae and Calyceraceae, followed by unique WGDs in *Chrysanthemum* and *Helianthus* lineages, as supported by previous literature, **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.

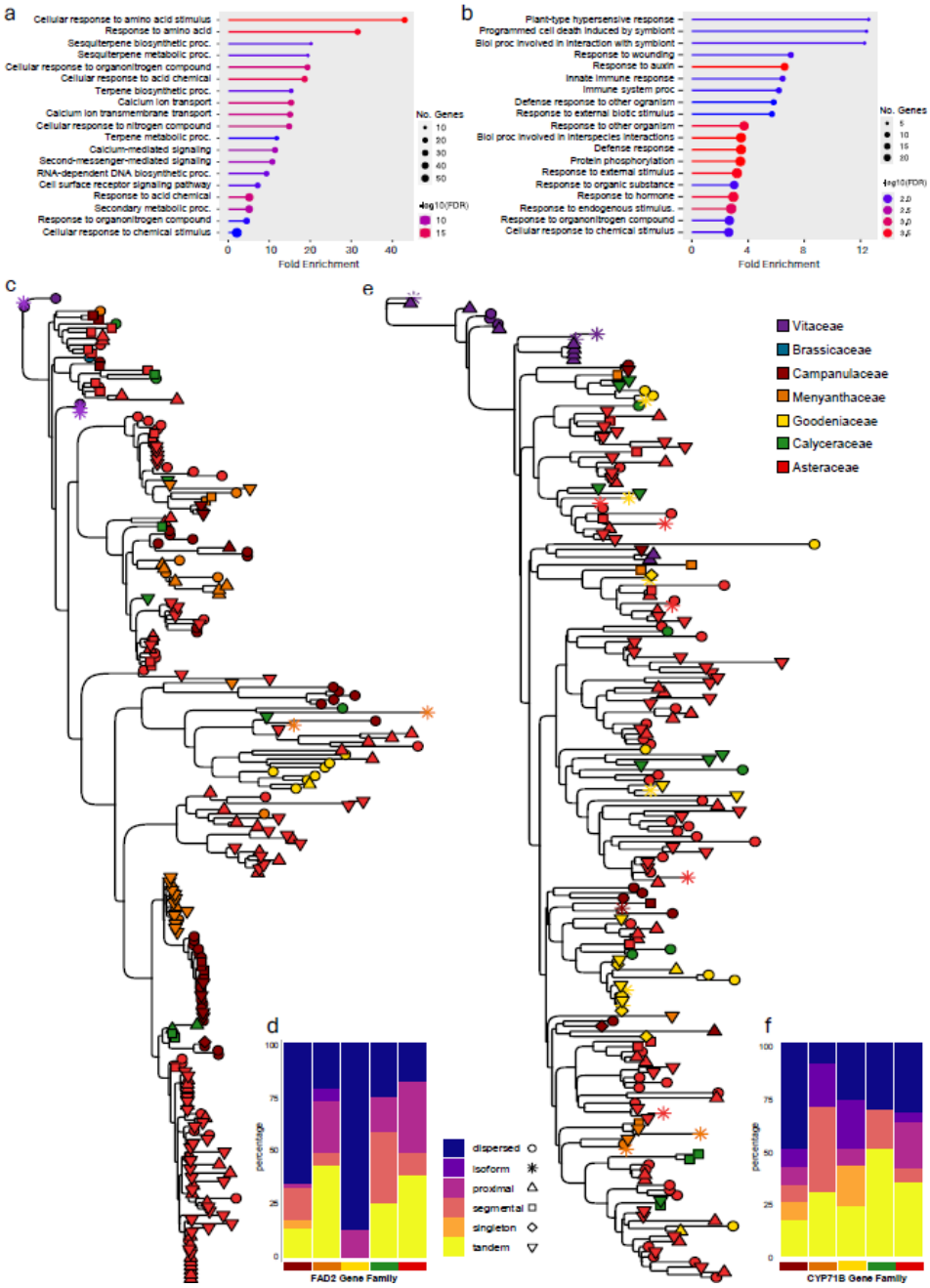

Fig. 4. Genes underpinning Asteraceae diversification. GO enrichment of biological processes associated with expanded orthogroups in **a)** Asteraceae and **b)** *Acicarpha;* **c)** Expanded *FAD2* gene family tree where tip colors represent the taxonomic family of each gene as indicated by the legend 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**.**

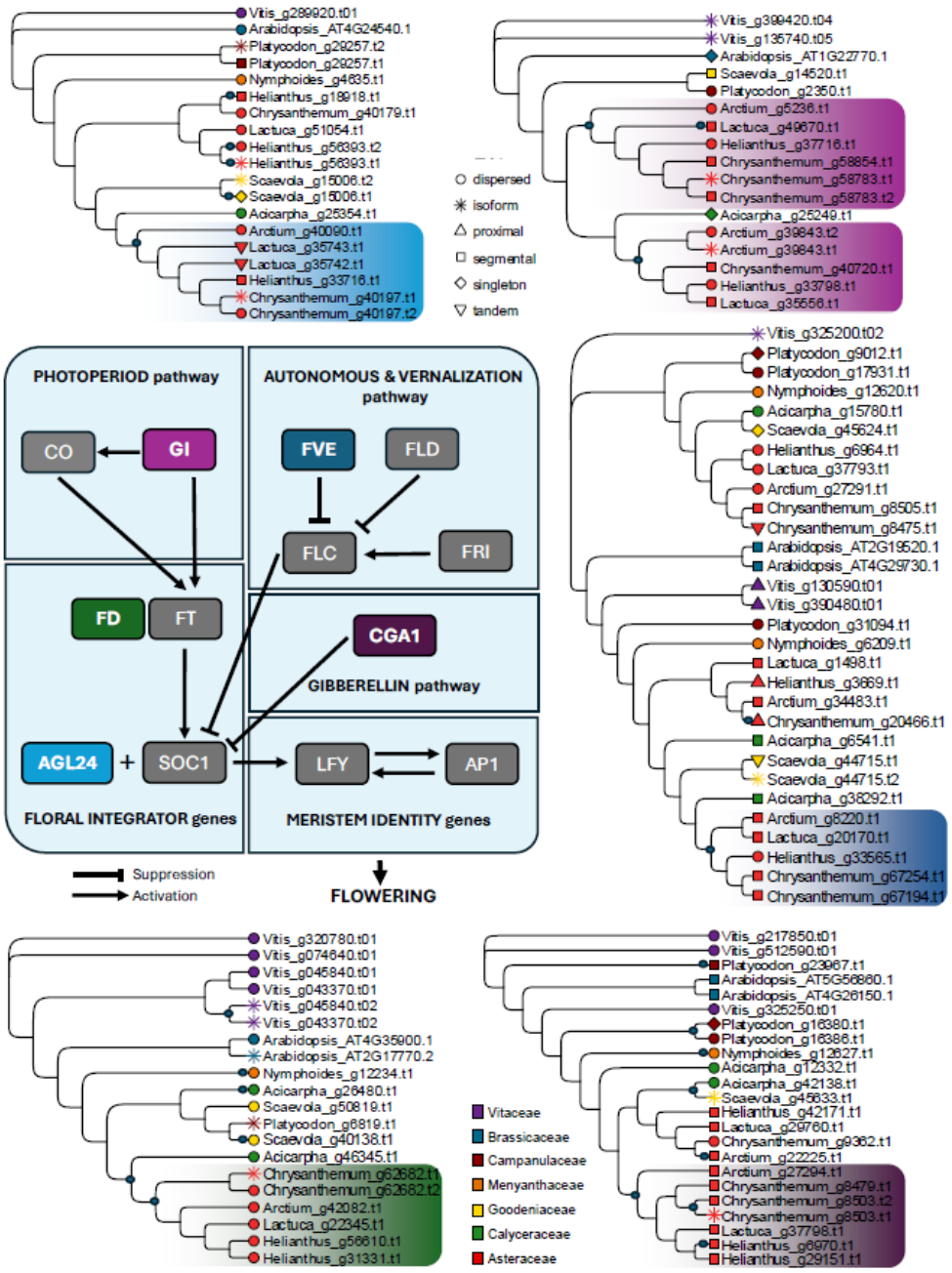

Fig. 5. Positively selected flowering time regulatory genes. Flowering time gene network involving the photoperiod, autonomous, vernalization, and gibberellin pathways. Colored gene boxes indicate orthogroups which experienced episodic positive selection in a lineage of Asteraceae homologs. Episodes 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.