# Gene duplications and genomic conflict underlie major pulses of phenotypic evolution in gymnosperms

Gregory W. Stull[1,2,†], Xiao-Jian Qu[3,†], Caroline Parins-Fukuchi[4], Ying-Ying Yang[1], Jun-Bo Yang[2], Zhi-Yun Yang[2], Yi Hu[5], Hong Ma[5], Pamela S. Soltis[6], Douglas E. Soltis[6,7], De-Zhu Li[1,2,*], Stephen A. Smith[8,*], Ting-Shuang Yi[1,2,*].

[1]Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan, China.
[2]CAS Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, China.
[3]Shandong Provincial Key Laboratory of Plant Stress Research, College of Life Sciences, Shandong Normal University, Jinan, Shandong, China.
[4]Department of Geophysical Sciences, University of Chicago, Chicago, IL, USA.
[5]Department of Biology, Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, PA, USA.
[6]Florida Museum of Natural History, University of Florida, Gainesville, FL, USA.
[7]Department of Biology, University of Florida, Gainesville, FL, USA.
[8]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA.
[†]Co-first author.
*Correspondence to: dzl@mail.kib.ac.cn; eebsmith@umich.edu; tingshuangyi@mail.kib.ac.cn.

## Abstract

Inferring the intrinsic and extrinsic drivers of species diversification and phenotypic disparity across the Tree of Life is a major challenge in evolutionary biology. In green plants, polyploidy (or whole-genome duplication, WGD) is known to play a major role in microevolution and speciation[1], but the extent to which WGD has shaped macroevolutionary patterns of diversification and phenotypic innovation across plant phylogeny remains an open question. Here we examine the relationship of various facets of genomic evolution—including gene and genome duplication, genome size, and chromosome number—with macroevolutionary patterns of phenotypic innovation, species diversification, and climatic occupancy in gymnosperms. We show that genomic changes, including WGD, underlie the origins of most major gymnosperm clades, and that spikes of gene duplication typically coincide with major spikes of phenotypic innovation. Increased rates of phenotypic evolution, however, are typically found at nodes with high gene-tree conflict, representing historic population-level dynamics during speciation. Most shifts in gymnosperm diversification since the rise of angiosperms are decoupled from putative WGDs and instead are associated with increased rates of climatic occupancy evolution, particularly in cooler and/or more arid climatic conditions, suggesting that ecological opportunity, especially in the latter Cenozoic, and environmental heterogeneity have driven a resurgence of gymnosperm diversification. Our study provides critical insight on the processes underlying diversification and phenotypic evolution in gymnosperms, with important broader implications for the major drivers of both micro- and macroevolution in plants.

## Main Text

Connecting microevolutionary processes with emergent macroevolutionary patterns of phenotypic disparity and species diversification is one of the grand challenges of evolutionary biology. In green plants, polyploidy is a common phenomenon in many lineages[1,2] that has long been considered a major evolutionary force[3,4], unique in its ability to generate manifold phenotypic novelties at the population level and in its potential to drive long-term evolutionary trends. However, despite its prevalence and clear microevolutionary importance, its macroevolutionary significance remains unclear and, in major clades such as gymnosperms, understudied. For example, analyses of the relationship between WGD and diversification have produced conflicting results[5,6,7,8], and no studies to our knowledge have examined whether spikes in gene duplication tend to directly coincide with bursts of phenotypic novelty on a macroevolutionary scale, although clearly gene duplications have facilitated the origins of numerous key traits in plants[9,10,11]. It is possible that other processes—e.g., major fluctuations in population size during rapid speciation events, in concert with extrinsic factors such as climate change or ecological opportunity—may contribute more fundamentally to major phases of morphological innovation and diversification[12,13]. More realistically, however, these various phenomena may interact in complex ways to shape macroevolutionary patterns in plants and, more broadly, across the Tree of Life, and detailed case studies are needed to disentangle their relative contributions to the origins of phenotypic and taxonomic diversity.

Here we examine potential drivers of phenotypic innovation and species diversification in seed plants (=*Spermatophyta*[14]), with a focus on *Acrogymnospermae*[14], the most ancient clade of living spermatophytes, including conifers, cycads, Gnetales, and *Ginkgo biloba*, as well now-extinct groups such as Cordaitales[15]. For the past ~300 million years, seed plants have occupied center stage in terrestrial ecosystems[16], constituting the most dominant plant clade in terms of both biomass and diversity and forming myriad ecological relationships with other organisms across the Tree of Life, including bacteria, fungi, insects, and vertebrates[17]. Reconstructing the diversification of seed plants is therefore essential for a better understanding of the history of terrestrial ecosystems across the latter half of the Phanerozoic. Many aspects of seed plant phylogeny remain poorly known or contentious[18], including the phylogenetic placements of numerous extinct lineages (e.g., Bennettitales and Caytoniales) with respect to living "gymnosperms" (*Acrogymnospermae*) and flowering plants (*Angiospermae*[14]). Nevertheless,

based on the extensive fossil record of seed plants[19] and phylogenetic studies of living representatives *Acrogymnospermae*[20,21] (which comprise ca. 1100 species), extant gymnosperm lineages clearly exhibit a complex history of ancient radiations, major extinctions, extraordinary stasis, and recent diversification, but the correlates and causes of major phases of gymnosperm evolution remain understudied. Given that recent studies[22–25] have challenged the traditional notion[26,27] that WGD is rare or insignificant in gymnosperms, a focused investigation on the prevalence and macroevolutionary significance of polyploidy in gymnosperms is timely.

We synthesize phylogenomic analyses, trait reconstructions, and examinations of extensive comparative data on a comprehensive phylogeny of gymnosperms to understand major processes that have shaped phenotypic innovation, species diversification, and climatic evolution in this clade. Specifically, we analyze a novel transcriptome dataset including representatives of nearly all extant genera of gymnosperms to identify hotspots of genomic conflict and gene duplication across gymnosperm phylogeny, which we relate to reconstructed bursts of phenotypic innovation and major shifts in species diversification and climatic evolution.

## Gymnosperm phylogenomics, polyploidy, and genome evolution

Based on phylogenomic analysis of our transcriptome dataset (including 790 inferred orthologs across 121 ingroup species, 77 of which were newly sequenced for this study; Supplementary Table 1), we recovered a well-resolved and highly supported species tree for gymnosperms (Fig. 1). We then performed a multifaceted approach for WGD inference to circumvent the limitations of some methods and datatypes[28]. Using the species-tree phylogeny and the inferred orthogroups (Methods), we identified hotspots of gene duplication indicative of WGD (Fig. 1), and compared these results against synonymous substitution (*Ks*) plots (Extended Data Figs. 1-3; Supplementary Figs. 1–14) as well as reconstructions of chromosome number (Supplementary Fig. 15) and genome size (Extended Data Fig. 4) across a broader sampling of gymnosperm species from our supermatrix phylogeny (Methods). Particular attention was paid to branches with previously inferred (and debated) instances of WGD in gymnosperms[22,29–32].

We inferred elevated levels of gene duplication in the branches subtending seed plants (56.2 %), gymnosperms (41.7 %), *Coniferae*[14] (including Gnetales, 18.8 %), Pinaceae (12.1 %), *Cupressophyta*[14] (i.e., Araucariaceae + Podocarpaceae + Sciadopityaceae + Taxaceae + Cupressaceae; 13.9 %), Podocarpaceae (11.3 %), and Araucariaceae (10.1 %). Most of these

branches also show shared *Ks* peaks, inferred changes in chromosome number, or extreme shifts in genome size (Fig. 1; Extended Data Figs. 1-4; Supplementary Figs. 1–15): seed plants (shared *Ks* peak), *Coniferae* (shared *Ks* peak), *Cupressophyta* ($n$ = 7 to $n$ = 11), Pinaceae (shared *Ks* peak; $n$ = 7 to $n$ = 12), Podocarpaceae ($n$ = 11 to $n$ = 10; major C-value decrease), and Araucariaceae ($n$ = 11 to $n$ = 13). We also observed shared *Ks* peaks in *Ginkgo* + cycads (hereafter called 'ginkads'), as well as possibly in Taxaceae + Cupressaceae (or in several subclades within), although neither of these instances corresponded with extreme levels of gene duplication.

Ancestral chromosomal changes were also observed in Cycadaceae ($n$ = 7 to $n$ = 11), Zamiaceae ($n$ = 7 to $n$ = 8), and Gnetaceae ($n$ = 7 to $n$ = 12), as well as within Cupressaceae, Podocarpaceae, and the genera of Gnetales, due to more recent paleopolyploidy or neopolyploidy[24,25,33] (Supplementary Fig. 15). Regarding genome size, rate shifts (inferred using BAMM) and major 'jumps' (i.e., extreme changes in reconstructed ancestor to descendant C-values) in genome size evolution (Methods) were detected at the origin of several major groups, including abrupt genome downsizing in Podocarpaceae (as noted above), Gnetales, and Cycadaceae and rate shifts in Zamiaceae (excluding *Dioon*), Taxaceae + Cupressaceae, and Ephedraceae; jumps or rate shifts were also detected in numerous more recent clades (Extended Data Fig. 4). See Supplementary Information for additional results and discussion of gymnosperm genome evolution.

## Genomic correlates of diversification and phenotypic evolution

Our diversification analysis (Supplementary Fig. 16), based our dated supermatrix phylogeny of extant gymnosperms (including 82% species representation), recovered 17 major shifts, only two of which correspond directly to branches with possible WGDs: *Cupressophyta* and Pinaceae (Fig. 1). Seven other diversification shifts corresponded to major jumps or rate shifts in genome size evolution or changes in chromosome number (Fig. 1, Extended Data Fig. 4, and Supplementary Information). Although most diversification shifts (15/17) were not associated with putative WGDs, it is possible that major extinctions have obscured instances of diversification shifts associated with WGD in deep time. On the other hand, extinction of depauperate branches separating a WGD event and a diversification shift could also lead to their erroneous association. With these considerations in mind, we argue at least that WGD has played

155 a minimal role in driving the more geologically recent diversification shifts detected here. Our
156 results also align with recent studies showing an inconsistent relationship between WGD and
157 diversification[5,6,7]. We suggest that in cases where WGD is associated with diversification, it is a
158 product of WGD mediating changes in traits (such as key innovations[9]) or life history that
159 influence population structuring, with diversification only being a secondary consequence[7]. This
160 is exemplified by Pinaceae, where a possible ancestral WGD is associated with not only a shift in
161 diversification (Fig. 1) but also a major spike in morphological innovation (Fig. 2) and a rate
162 shift in climatic evolution (see below).

163 To examine the general relationship of gene duplications and genomic (gene-tree)
164 conflict with levels of phenotypic innovation and rates of phenotypic evolution, we analyzed a
165 newly compiled trait dataset for gymnosperms comprising 148 phenotypic characters, with
166 sampling matched to the transcriptome phylogeny (Supplementary Information). Sub- and neo-
167 functionalization of duplicate genes are considered the primary means by which polyploidy can
168 lead to phenotypic novelty[34,35], and therefore an examination of the direct correspondence of gene
169 duplication numbers to levels of phenotypic novelty at nodes across gymnosperm phylogeny
170 serves as a means of assessing the potential contribution of large-scale (WGD) and small-scale
171 genomic events to macroevolutionary patterns of phenotypic evolution. Genomic conflict, on the
172 other hand, records information about population-level processes during speciation events—e.g.,
173 rapid population expansions or fragmentation, resulting in incomplete lineage sorting (ILS), or
174 introgression—that might also play a role in rapid acquisition of phenotypic novelties[36]. Thus,
175 these represent two different, albeit not mutually exclusive, processes that might shape
176 macroevolutionary patterns of morphological evolution in plants. We reconstructed levels of
177 phenotypic innovation or novelty (defined here simply as state changes along a branch) and rates
178 (state changes/time) for each branch in the phylogeny[13] and plotted these against corresponding
179 levels of gene duplication and gene-tree conflict (Figs. 2 and 3). Linear regression and
180 permutation tests were also performed to evaluate the statistical strength of these relationships
181 (Methods).

182 We find a positive relationship between levels of phenotypic innovation and levels of
183 gene duplication (linear regression: $r^2 = 0.2038$, $p = 0.000$; permutation test: $p = 0.038$;
184 Supplementary Figs. 17, 18), suggesting that gene duplications may have played a vital role in
185 the generation of phenotypic novelty in gymnosperms evident at a macroevolutionary scale (Fig.

186 2). Major examples include Araucariaceae, Cupressaceae, *Pinus*, Pinaceae, Podocarpaceae, and
187 Zamiaceae, which all show corresponding, elevated levels of gene duplication and phenotypic
188 innovation. There are also evident temporal patterns, with, for example, a noticeable increase in
189 duplications and phenotypic innovation near the present (Fig. 2). While causation between gene
190 duplication and phenotypic innovation cannot be demonstrated on the basis of these observed
191 patterns, this connection clearly deserves future study given that gene duplications have long
192 been considered to play a key role in the origin of evolutionary novelties[34]. It has been suggested
193 that small-scale and large-scale (e.g., WGD) duplications might tend to have different immediate
194 consequences and evolutionary trajectories[35]. Some of the putative WGD events examined here
195 seem most salient in terms of morphological innovation, but the general correspondence
196 (Supplementary Figs. 17, 18) suggests that perhaps both large- and small-scale duplication
197 events have played an important role in the generation of novel traits in gymnosperms.
198 Although we observed a significant relationship between gene duplication and
199 phenotypic novelty, there are notable discrepancies, the most salient being Gnetales (Fig. 2).
200 However, this is perhaps not surprising given that Gnetales has experienced significant life
201 history changes and stands out among gymnosperms in having relatively small genomes (despite
202 at least recent WGD events[22,25]) and remarkably fast ("angiosperm-like") rates of molecular
203 evolution[37,38]. *Coniferae* (including Gnetales) also represents a major discrepancy, and the
204 exclusion of Gnetales from the phenotypic reconstructions results in an increase in the number of
205 inferred apomorphies for conifers (Supplementary Fig. 19). However, another explanation for the
206 low levels of innovation inferred for the *Coniferae* node might be the prevalence of traits shared
207 between conifers and *Ginkgo*. Although our species tree places *Ginkgo* sister to cycads (Fig. 1),
208 morphological phylogenetic analyses of seed plants[18] have typically placed *Ginkgo* with extinct
209 and extant conifers (rather than cycads), and among our gene trees (Fig. 1) the dominant
210 alternative topology for *Ginkgo* is sister to *Coniferae* (with 81 gene trees showing this alternative
211 topology with strong support). This finding suggests that reticulate evolution in the early history
212 of *Acrogymnospermae* could be partially driving this discrepancy.
213 In contrast to cumulative levels of phenotypic innovation, rates of phenotypic evolution
214 show a strong relationship with levels of gene-tree conflict (Fig. 3; linear regression: $r^2 = 0.2308$,
215 $p = 0.000$; Supplementary Fig. 20), with extreme values of conflict and rate in particular tending
216 to co-occur (permutation test: $p = 0.001$; Supplementary Fig. 21). This finding is in line with a

recent study[13] documenting a similar correspondence in multiple clades across the Tree of Life. Figure 3 clearly shows that the branches with the highest rates of morphological evolution also have extreme levels of gene-tree conflict. A noteworthy example is the Podocarpaceae subclade including *Phyllocladus* and *Parasitaxus* (Fig. 3), which is characterized by extensive conflict as well as remarkable shifts in phenotype (e.g., the acquisition of phylloclades) and life history (e.g., parasitism) in several of the constituent lineages.

If gene-tree conflict reflects a signature of historical population-level processes such as ILS[39], the observed relationship between conflict and phenotypic rates suggests that, at macroevolutionary scales, many instances of rapid morphological evolution in gymnosperms are a product of episodic and dramatic population perturbations associated with rapid speciation events. Given the life history characteristics of (most) gymnosperms—including large genome sizes, long time to reproduction, and, in some species, exceptional longevity[40,41]—instances of population reduction or isolation should theoretically be essential for instances of rapid phenotypic evolution. Thus, areas of gymnosperm phylogeny with elevated conflict and morphological rates perhaps reflect the rapid fragmentation of broad, variable ancestral populations. As discussed below, this type of process might explain the prevalence of young conifer and cycad species with highly restricted distributions today. We also note that the processes outlined above are not mutually exclusive putative drivers of phenotypic evolution, given that gene duplications might serve as a source of novelties sorted by population-level processes and that WGD can theoretically result in the rapid establishment of small, reproductively isolated populations[9,42], which could result in the rapid fixation of polymorphisms via genetic drift.

## Temporal patterns of climatic occupancy evolution and species diversification

Reconstructions of climatic variables on the dated supermatrix phylogeny—focusing on mean annual temperature (Bio 1), mean annual precipitation (Bio 12), and principal components one and two of the 19 WorldClim bioclimatic layers—revealed that rate shifts and jumps in climatic occupancy are concentrated in the last ~70 Myr (and particularly the last ~25 Myr; Supplementary Figs. 22–27). Additionally, many of the branches (10/17) with a diversification shift also included a shift in the rate of climatic evolution (Fig. 4), with four additional diversification shifts separated from a climatic shift by only one or two nodes (i.e., 14/17

diversification shifts are directly or strongly associated with climatic shifts; Fig. 4). Except for two of the concerted shifts (in the branches subtending the cycad genera *Ceratozamia* and *Zamia*), all others involved shifts associated with cooler and/or drier climatic conditions (Fig. 4 and Supplementary Figs. 22–27). Undoubtedly, there have been many deeper bursts of diversification and shifts in climatic occupancy in gymnosperms that are not evident in our phylogeny due to extinction[19]; this renders the inference of diversification dynamics across the depth of gymnosperm history a major, if not insurmountable, challenge without extensive data from the fossil record[43]. Nevertheless, considering an extensive body of research concerning gymnosperm vs. angiosperm functional morphology and competitive dynamics today[44] and patterns in the fossil record[45], we might predict many gymnosperms (and especially conifers) to thrive with the expansion of cooler and more arid conditions after late Eocene climatic deterioration and particularly after the mid-Miocene[46].

The global ecological ascendance of angiosperms, from the early Cretaceous to early Cenozoic, was a complex and relatively protracted process, with gymnosperms showing heterogeneous patterns of decline, in terms of abundance and diversity, across different lineages, ecological guilds, geographic areas, and points in time[45,47]. It was not until the early Cenozoic, perhaps in part facilitated by major ecological upheaval at the K-Pg boundary, that angiosperms came to dominate (or co-dominate) nearly all geographic areas and vegetation types[47]. Numerous attributes of angiosperms may underpin their macroevolutionary success, including their reproductive innovations and ecological associations[48], proclivity for polyploidy[32,49,50], and faster rates of molecular evolution[37]. However, their functional morphology and correspondingly elevated growth rates have been highlighted as particularly important in the context of competition with gymnosperms[44]. The early Cenozoic was a time of warm, equable climate—conditions well suited to rapid growth strategies of many angiosperms, facilitated by broad leaves and a more efficient vascular system including vessels. Gymnosperms, with narrow tracheids and (generally) smaller, more poorly vascularized evergreen leaves, may have struggled to compete with angiosperms across much of the landscape of the early Cenozoic given the growth-rate limitations imposed by their less efficient vascular systems[44]. Exceptions that prove the rule include *Gnetum* and some members of Podocarpaceae that independently evolved angiosperm-like vasculature and/or broader leaves, allowing them to compete in wet tropical conditions where gymnosperms (especially conifers) are otherwise largely excluded[51].

279 However, the same basic traits (and consequent slow-growth strategies) that exclude most
280 gymnosperms from habitats that are more favorable to angiosperm plant growth allow many to
281 persist in more extreme, stressful habitats[52]. The expansion of cool and arid conditions after the
282 Eocene and especially after the mid-Miocene therefore may have created many new
283 opportunities for gymnosperm expansion and diversification. This hypothesis is consistent with
284 our results showing many gymnosperm lineages shifting into cooler and/or more arid conditions
285 after the late Eocene (Supplementary Figs. 22–27). Diversification shifts associated with
286 cool/arid transitions that occurred prior to the late Eocene (e.g., *Pinus* subsection *Ponderosae*,
287 ~56 Ma; the *Juniperus-Cupressus* clade, ~70 Ma; the *Podocarpus-Retrophyllum* clade, ~68 Ma)
288 might represent competitive displacement by angiosperms and/or expansion into upland or
289 montane habitats[53]. Future biogeographic analyses integrating fossil evidence might shed light on
290 the geographic context of these coordinated climatic-diversification shifts. In general, our results
291 challenge suggestions that recent climatic cooling precipitated macroevolutionary declines in
292 conifers[54] (Supplementary Information). The Northern and Southern Hemispheres, however,
293 clearly do show distinct patterns with respect to the extinction and turnover of conifer diversity
294 during the Cenozoic[21], and numerous extant gymnosperm species (especially cycads) are at risk
295 of extinction due to pressures including habitat destruction[55,56,57].
296 While climatic shifts have played a significant role in the Cenozoic history of
297 gymnosperms, many major climatic shifts are not associated with changes in diversification (Fig.
298 4 and Supplementary Figs. 22–27), suggesting that climatic evolution might be an important but
299 insufficient vehicle for diversification. Ecological factors (e.g., environmental heterogeneity
300 related to elevation or soil type) and/or traits (e.g., dispersal strategies) that more directly control
301 population size and spatial structuring are likely necessary concomitant factors for driving
302 diversification patterns in gymnosperms, and in green plants more broadly. This finding is
303 consistent with the observation that many recent gymnosperm radiations (e.g., multiples lineages
304 of cycads, Pinaceae, and Cupressaceae) include numerous species with insular/restricted
305 distributions[58]. The small distributions of many gymnosperm species have typically been
306 interpreted as relictual[40], but clearly many of these species stem from recent, Neogene radiations
307 given their remarkably short molecular branch lengths[20,21]. Aside from these numerous recent
308 radiations, however, extensive data from the fossil record show that many extant
309 gymnosperms—e.g., *Ginkgo biloba*, members of Taxaceae (including *Cephalotaxus*), and the

310 "taxodioid" Cupressaceae—are clearly relics of historically more diverse and broadly distributed
311 groups.
312

313 **Macroevolutionary synthesis**

314 Our analyses provide an unprecedented and multifaceted view of genome evolution in
315 gymnosperms, with broad implications for how different attributes of genome evolution might
316 both drive and signal major phases of phenotypic evolution in plants. Using multiple lines of
317 evidence, we show that polyploidy and dynamic shifts in chromosome number and genome size
318 are common occurrences across the phylogenetic breadth of gymnosperms. We further provide
319 one of the first direct documentations of a general relationship between levels of gene
320 duplication and phenotypic innovation at a macroevolutionary scale. Gene duplications have
321 long been emphasized as a fundamental source of novelty, but their evolutionary outcomes have
322 typically been studied at a microevolutionary scale or with respect to particular traits or nodes of
323 interest[9,10,11] (but see Walden et al.[8]). The broader patterns documented here provide a framework
324 of promising areas of gymnosperm phylogeny in which to investigate connections between
325 duplications or gene-family expansions and phenotypic novelties.

326 Contrary to the association of duplications and innovation, we do not observe a consistent
327 relationship between polyploidy and diversification—consistent with other recent studies also
328 showing a sporadic association[5,6,7]. This implies that, across the euphyllophyte radiation, WGD is
329 more fundamentally tied to phenotypic innovation, and that subsequent diversification is
330 contingent on the nature of the traits acquired and the presence of ecological opportunity[9]. In
331 gymnosperms, recent phases of diversification appear closely tied with shifts in climatic
332 occupancy, suggesting that the expansion of cool and arid climates in the latter Cenozoic created
333 new ecological opportunities suited to the basic functional morphology of this clade. Finally, we
334 show that elevated rates of phenotypic evolution appear most closely tied to phylogenetic regions
335 with extensive gene-tree conflict—a signature of dynamic population processes during rapid
336 speciation events. Collectively these threads create a nuanced picture of how different
337 processes—including gene and genome duplication, population dynamics such as ILS and
338 introgression, and ecological opportunity in the face of climate change—may interact to drive the
339 emergence of novel morphologies and major radiations in plants.

340 1. Otto, S. P. & Whitton, J. Polyploid incidence and evolution. *Ann. Rev. Genet.* **34**, 401–437
341 (2000).
342 2. Wood, T. E. et al. The frequency of polyploid speciation in vascular plants. *Proc. Natl. Acad.*
343 *Sci. U.S.A.* **106**, 13875–13879 (2009).
344 3. Stebbins, G. L. *Variation and evolution in plants* (Columbia Univ. Press, 1950).
345 4. Levin, D. A. Polyploidy and novelty in flowering plants. *Am. Nat.* **122**, 1–25 (1983).
346 5. Tank, D. C. et al. Nested radiations and the pulse of angiosperm diversification: increased
347 diversification rates often follow whole genome duplications. *New Phytol.* **207**, 454–467
348 (2015).
349 6. Landis, J. B. et al. Impact of whole-genome duplication events on diversification rates in
350 angiosperms. *Am. J. Bot.* **105**, 348–363 (2018).
351 7. Smith, S. A. et al. Disparity, diversity, and duplications in the Caryophyllales. *New Phytol.*
352 **217**, 836–854 (2017).
353 8. Walden, N. et al. Nested whole-genome duplications coincide with diversification and high
354 morphological disparity in Brassicaceae. *Nature Comm.* **11**, 3795.
355 9. Soltis, P.S. & Soltis, D. E. Ancient WGD events as drivers of key innovations in angiosperms.
356 *Curr. Opin. Genet. Dev.* **30**, 159–165 (2016).
357 10. Guo, J., et al. Phylotranscriptomics in Cucurbitaceae reveal multiple whole-genome
358 duplications and key morphological and molecular innovations. *Mol. Plant* **13**, 1117–1133
359 (2020).
360 11. Sheehan, H. et al. Evolution of L-DOPA 4,5-dioxygenase activity allows for recurrent
361 specialization to betalain pigmentation in Caryophyllales. *New Phytol.* **227**, 914–929 (2020).
362 12. Stroud, J. T. & Losos, J. B. Ecological opportunity and adaptive radiation. *Annu. Rev. Ecol.*
363 *Evol. Syst.* **47**, 507–532 (2016).
364 13. Parins-Fukuchi, Stull, G. W. & Smith, S. A. Phylogenomic conflict coincides with rapid
365 morphological innovation. *bioRxiv* doi: https://doi.org/10.1101/2020.11.04.368902 (2020).
366 14. Cantino, et al. Towards a phylogenetic nomenclature of *Tracheophyta*. *Taxon* **56**, E1–E44.
367 15. Doyle, J. A. Phylogenetic analyses and morphological innovations in land plants. *Ann. Plant*
368 *Rev.* **45**, 1–50 (2013).

16. DiMichele, W. A., Pfefferkorn, H. W. & Gastaldo, R. A. Response of Late Carboniferous and Early Permian plant communities to climate change. *Annu. Rev. Earth Pl. Sci.* **29**, 461–487 (2001).
17. Behrensmeyer, A. K., et al. (eds) *Terrestrial ecosystems through time: evolutionary paleoecology of terrestrial plants and animals* (Univ. Chicago Press, 1992).
18. Crane, P. R. Phylogenetic analysis of seed plants and the origin of angiosperms. *Ann. Mo. Bot. Gard.* **72**, 716–793 (1985).
19. Taylor, E. L., Taylor, T. N. & Krings, M. *Paleobotany: the biology and evolution of fossil plants* (Academic Press, 2009).
20. Nagalingum, N. S. et al. Recent synchronous radiation of a living fossil. *Science* **334**, 796–799 (2011).
21. Leslie, A. B. et al. Hemisphere-scale differences in conifer evolutionary dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 16217–16221 (2012).
22. Li, Z. et al. Early genome duplications in conifers and other seed plants. *Sci. Adv.* **1**, e1501084 (2015).
23. Guan, R. et al. Draft genome of the living fossil *Ginkgo biloba*. *GigaScience* **5**, 49 (2016).
24. Farhat, P. et al. Polyploidy in the conifer genus *Juniperus*: an unexpectedly high rate. *Front. Plant Sci.* **10**, 676 (2019).
25. Ickert-Bond, S. M. et al. Polyploidy in gymnosperms—Insights into the genomic and evolutionary consequences of polyploidy in *Ephedra*. *Mol. Phylogenet. Evol.* **147**, 106786 (2020).
26. Khoshoo, T. N. Polyploidy in gymnosperms. *Evolution* **13**, 24–39 (1959).
27. Lewis, W. H. (ed) *Polyploidy—Biological Relevance* (Plenum Press, 1980).
28. Yang, Y. et al. Improved transcriptome sampling pinpoints 26 ancient and more recent polyploidy events in Caryophyllales, including two allopolyploidy events. *New Phytol.*, **217**, 855–870 (2018).
29. Jiao, Y. et al. Ancestral polyploidy in seed plants and angiosperms. *Nature* **473**, 97–100 (2011).
30. Roodt, D. et al. Evidence for an ancient whole genome duplication in the cycad lineage. *PLoS One* **12**, e0184454 (2017).

31. Ruprecht, C. et al. Revisiting ancestral polyploidy in plants. *Science Adv*. **3**, e1603195 (2017).
32. Leebens-Mack, J. H. et al. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **574**, 679–685 (2019).
33. Scott, A. D., Stenz, N. W. M., Ingvarsson, P. K. & Baum, D. A. Whole genome duplication in coast redwood (Sequoia sempervirens) and its implications for explaining the rarity of polyploidy in conifers. *New Phytol*. **211**, 186–193 (2016).
34. Ohno, S. *Evolution by Gene Duplication* (Springer Verlag, 1970).
35. Rensing, S. A. Gene duplication as a driver of plant morphogenetic evolution. *Curr. Opin. Plant Biol*. **14**, 43–48 (2014).
36. Marques, D. A., Meier, J. I. & Seehausen, O. A combinatorial view on speciation and adaptive radiation. *Trends Ecol. Evol*. **34**, 531–544 (2019).
37. De La Torre, A. R., Li, Z., Van de Peer, Y. & Ingvarsson, P. K. Contrasting rates of molecular evolution and patterns of selection among gymnosperms and flowering plants. *Mol. Biol. Evol*. **34**, 1363–1377 (2017).
38. Ran, J.-H., Shen, T.-T., Wang, M.-M. & Wang, X.-Q. Phylogenomics resolves the deep phylogeny of seed plants and indicates partial convergent or homoplastic evolution between Gnetales and angiosperms. *Proc. R. Soc. B* **285**, 20181012 (2018).
39. Oliver, J. C. Microevolutionary processes generate phylogenomic discordance at ancient divergences. *Evolution*, 67, 1823–1839 (2013).
40. Farjon, A. *A natural history of conifers* (Timber Press, 2008).
41. Pellicer, J., Hidalgo, O., Dodsworth, S. & Leitch, I. J. 2018. Genome size diversity and its impact on land plants. *Genes* **9**, 88 (2018).
42. Otto, S. P. 2007. The evolutionary consequences of polyploidy. *Cell* **131**, 452–462.
43. Louca, S. & Pennell, M. W. Extant timetrees are consistent with a myriad of diversification histories. *Nature* **580**, 502–505 (2020).
44. Bond, W. J. The tortoise and the hare: ecology of angiosperm dominance and gymnosperm persistence. *Biol. J. Linn. Soc*. **36**, 227–249 (1989).
45. Lupia, R., Ligard, S. & Crane, P. R. Comparing palynological abundance and diversity: implications for biotic replacement during the Cretaceous angiosperm radiation. *Paleobiology* ***25***, 305–340. (1999).

46. Zachos, J., Pagani, M., Sloan, L., Thomas, E. & Billups, K. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* ***292***, 686–693 (2001).

47. Wing, S. L., & Boucher, L. D. Ecological aspects of the Cretaceous flowering plant radiation. *Annu Rev. Earth Plant. Sci.* **26**, 379–421 (1998).

48. Tiffney, B. H. Seed size, dispersal syndromes, and the rise of the angiosperms: Evidence and hypothesis. *Ann. Missouri Bot.* **71**, 551–576 (1984).

49. Rice, A., et al. The Chromosome Counts Database (CCBD)—a community resource of plant chromosome numbers. *New Phyt.* **206**, 19–26.

50. Pellicer, J., & Leitch, I. J. The Plant DNA C-values database (release 7.1): an updated online repository of plant genome size data for comparative studies. *New Phyt.* **226**, 301–305 (2020).

51. Biffin, E. et al. Leaf evolution in Southern Hemisphere conifers tracks the angiosperm ecological radiation. *Proc. R. Soc. B* **279**, 341–348 (2012).

52. Brodribb, T. J., McAdam, S. A., Jordan, G. J. & Martins, S. C. Conifer species adapt to low-rainfall climates by following one of two divergent pathways. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 14489–14493 (2014).

53. English, J. M. & Johnson, S. T. The Laramide orogeny: what were the driving forces? *Int. Geol. Rev.* **46**, 83–838 (2004)

54. Condamine, F. L., Silvestro, D., Koppelhus, E. B. & Antonelli, A. The rise of angiosperms pushed conifers to decline during global cooling. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 28867–28875 (2020).

55. Fragnière, Y., Bétrisey, S., Cardinaux, L., Stoffel, M. & Kozlowski, G. Fighting their last stand? A global analysis of the distribution and conservation status of gymnosperms. *J. Biogeogr.* **42**, 809–820 (2015).

56. Forest, F. et al. Gymnosperms on the EDGE. *Sci. Rep.* **8,** 6053 (2018).

57. Mankga, L. T., Yessoufou, K. Factors driving the global decline of cycad diversity. *AoB Plants* **9**, plx022.

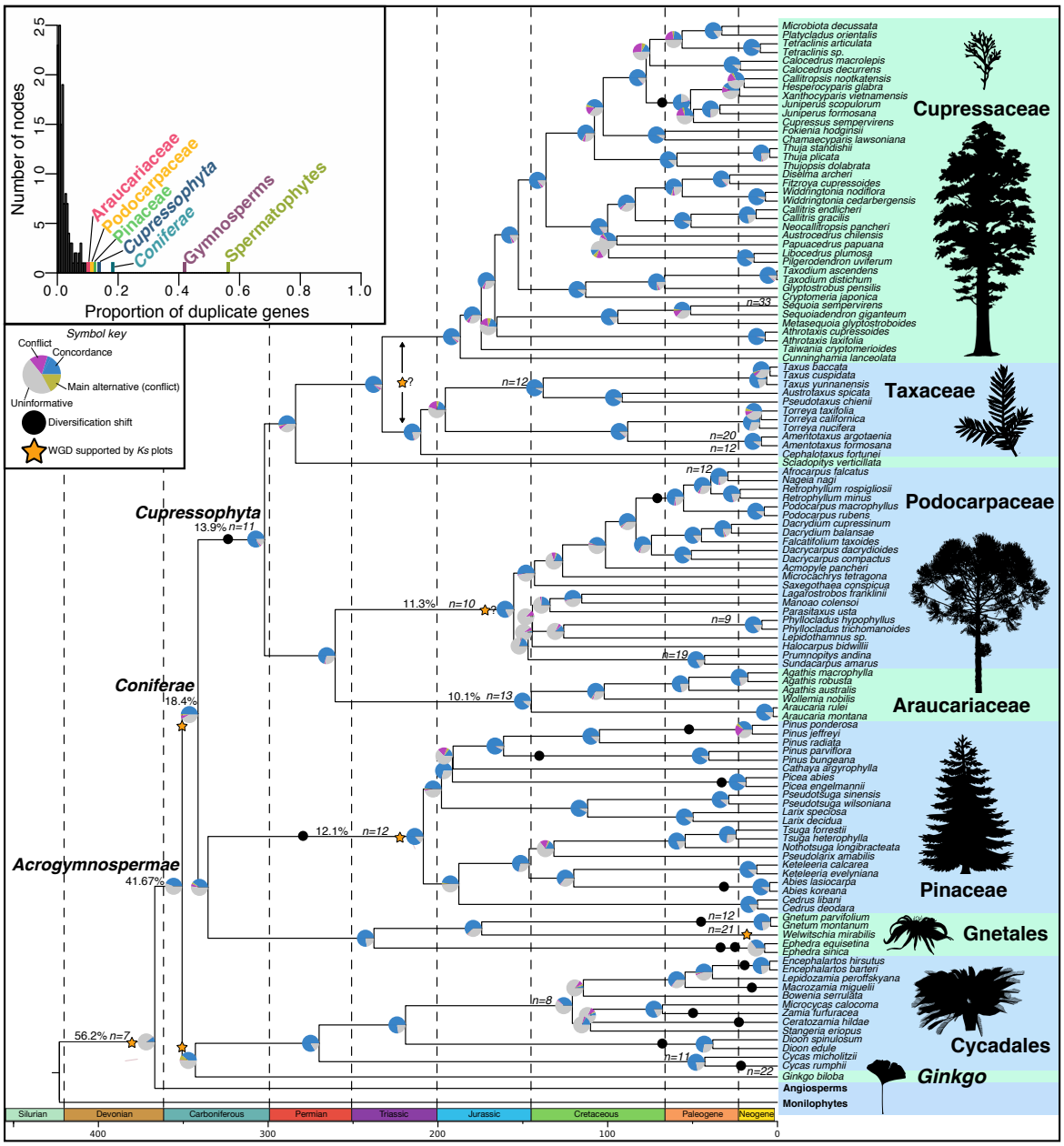
58. Rundel, P. W. 2019. A Neogene Heritage: Conifer Distributions and Endemism in Mediterranean-Climate Ecosystems. *Front. Ecol. Evol.* **7**, 364.

**Acknowledgements**. We thank the Germplasm Bank of Wild Species at the Kunming Institute of Botany (KIB) for facilitating this study. We also thank the curators and staff of the Kunming

Botanical Garden of the Kunming Institute of Botany, the University of California Botanical Garden at Berkeley, the Arnold Arboretum of Harvard University, the Missouri Botanical Garden, the Royal Botanic Garden Edinburgh, and the Royal Botanical Garden Kew for providing fresh and silica-dried leaves and DNA samples. This work was funded by the Strategic Priority Research Program of the Chinese Academy of Sciences (CAS) (grant No. XDB31000000), CAS' large-scale scientific facilities (grant No. 2017-LSF-GBOWS-02), the National Natural Science Foundation of China [key international (regional) cooperative research project No. 31720103903]. GWS acknowledges support from the Chinese Academy of Sciences (CAS) President's International Fellowship Initiative (No. 2020PB0009) and the China Postdoctoral Science Foundation (CPSF) International Postdoctoral Exchange Program.

**Author contributions**. G.W.S., D.-Z.L., S.A.S., and T.-S.Y. conceived the study; X.-J.Q., Y.-Y.Y., Y.H., J.-B.Y., Z.-Y.Y., and H.M. and collected and prepared samples for transcriptome and plastome sequencing; G.W.S. generated the trait dataset, and compiled publically available data for the supermatrix and comparative analyses; G.W.S. conducted analyses with help from C.P.-F., S.A.S., and X.-J.Q; G.W.S, C.P.-F., P.S.S., D.E.S., S.A.S., and T.-S.Y. interpreted the results; G.W.S. wrote the manuscript, with contributions from C.P.-F., P.S.S., D.E.S., S.A.S., and T.-S.Y. All authors approved the manuscript.

492 **Figures**




493

494 **Figure 1. Transcriptome species tree showing major genomic events**. Species tree inferred
495 from the 790 gene trees using ASTRAL, scaled to time by fixing the nodes to the ages inferred in
496 the supermatrix dating analysis. Pie charts at each node show the proportion of gene trees in
497 concordance (blue), in conflict (chartreuse = dominant alternative; magenta = other conflicting
498 topologies), or without information for that particular bipartition (grey). Nodes with the highest
499 percentages of duplicate genes (indicative of WGD) are highlighted in the inset plot in the top

left corner; the percentages of duplicate genes inferred for these outlier nodes are also displayed on the tree. Orange stars indicated WGD events supported by *Ks* plots; uncertainty regarding the phylogenetic placement(s) of such events is indicated with a question mark. Changes in the haploid chromosome number (based on the ChromEvol analyses; Methods) are indicated on the phylogeny. Black circles indicate diversification rate shifts (inferred using BAMM; Methods). Plant silhouettes were obtained from PhyloPic (http://phylopic.org); credit is acknowledged for the silhouettes of *Araucaria angustifolia* (LeonardoG and T. Michael Keesey); *Zamites* (Robert Gay), which closely resembles the architecture of cycads; and *Ginkgo* (Joe Schneid and T. Michael Keesey). The *Taxus* silhouette was generated from a line drawing of *Taxus brevifolia* obtained from the website of Natural Resources Canada, Canadian Forest Service (https://tidcf.nrcan.gc.ca/en/trees/factsheet/288).

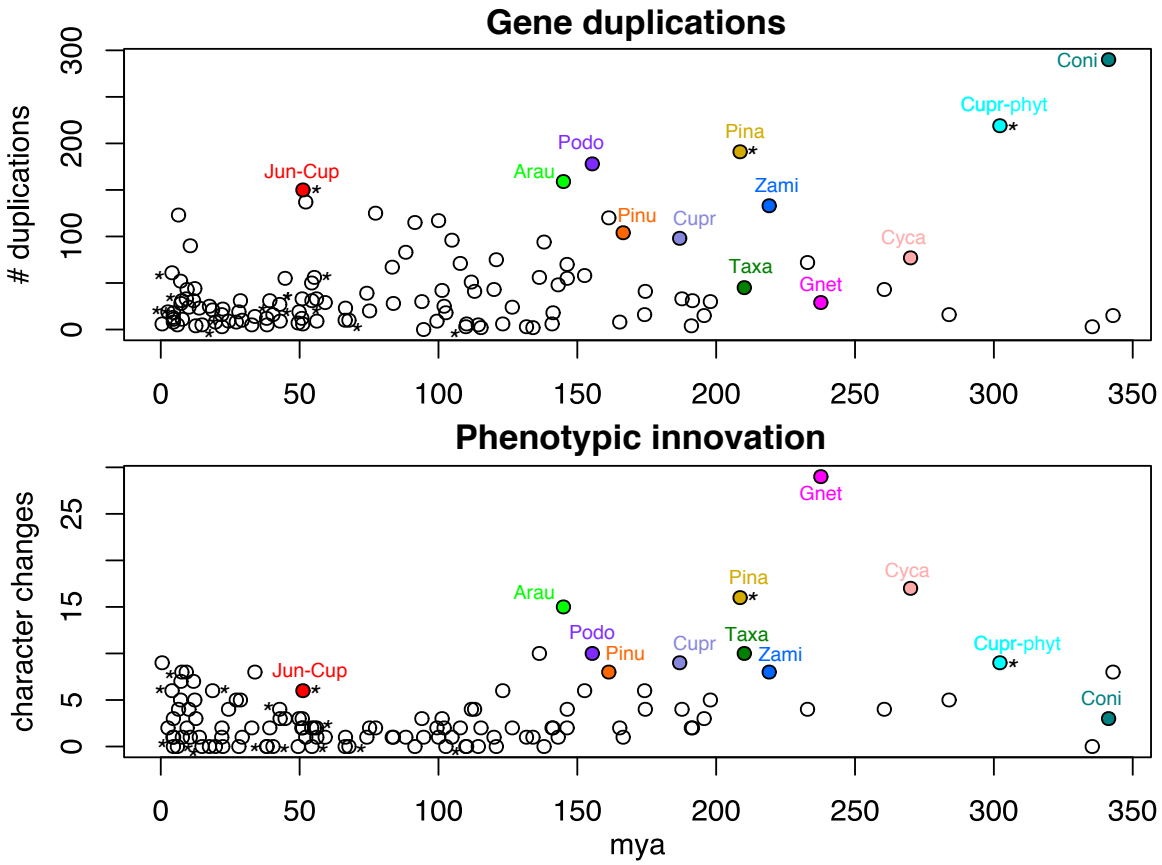



**Figure 2**. **Gene duplications vs. phenotypic innovation**. Plots juxtaposing levels of gene duplication and phenotypic innovation (i.e., the number of state changes) at corresponding

branches across gymnosperm phylogeny ($n$ = 119). The x-axis represents time in millions of years. Nodes with an inferred diversification shift are marked with an asterisk.

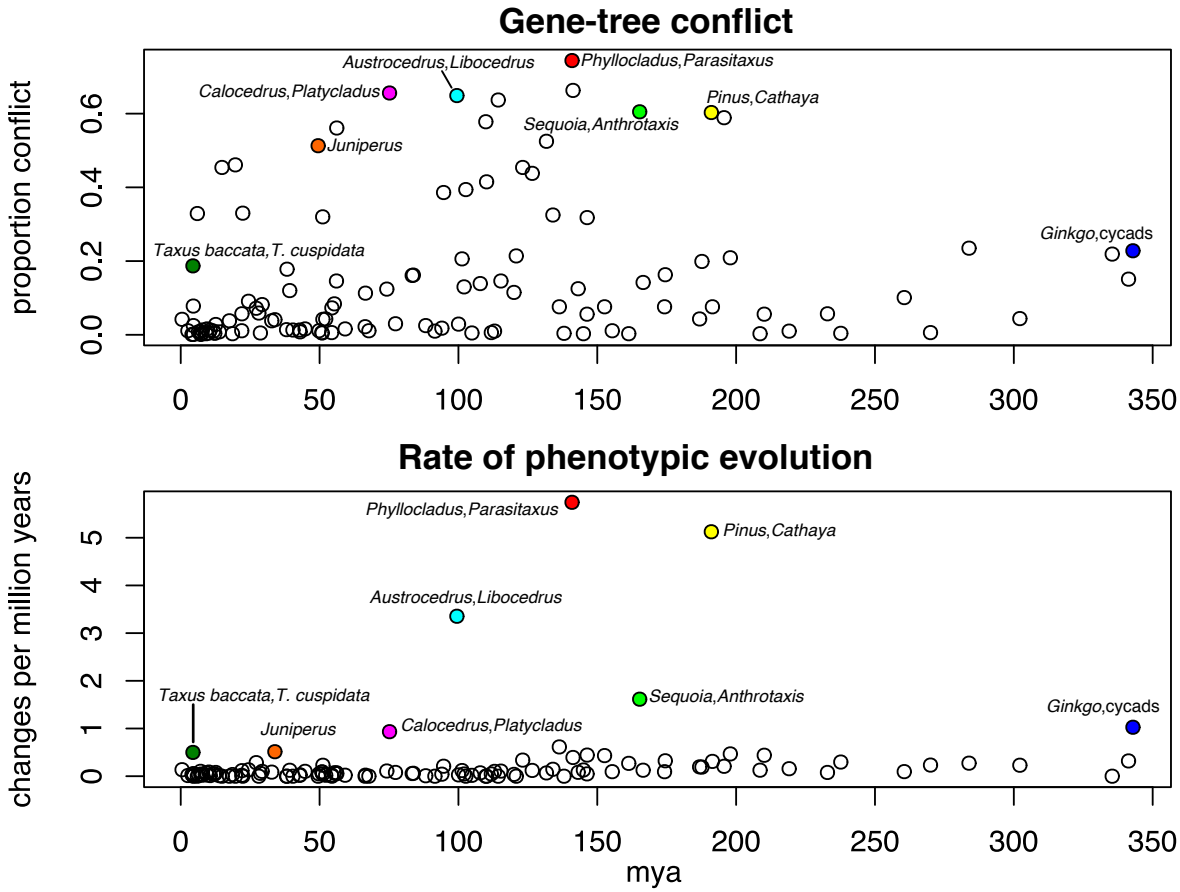
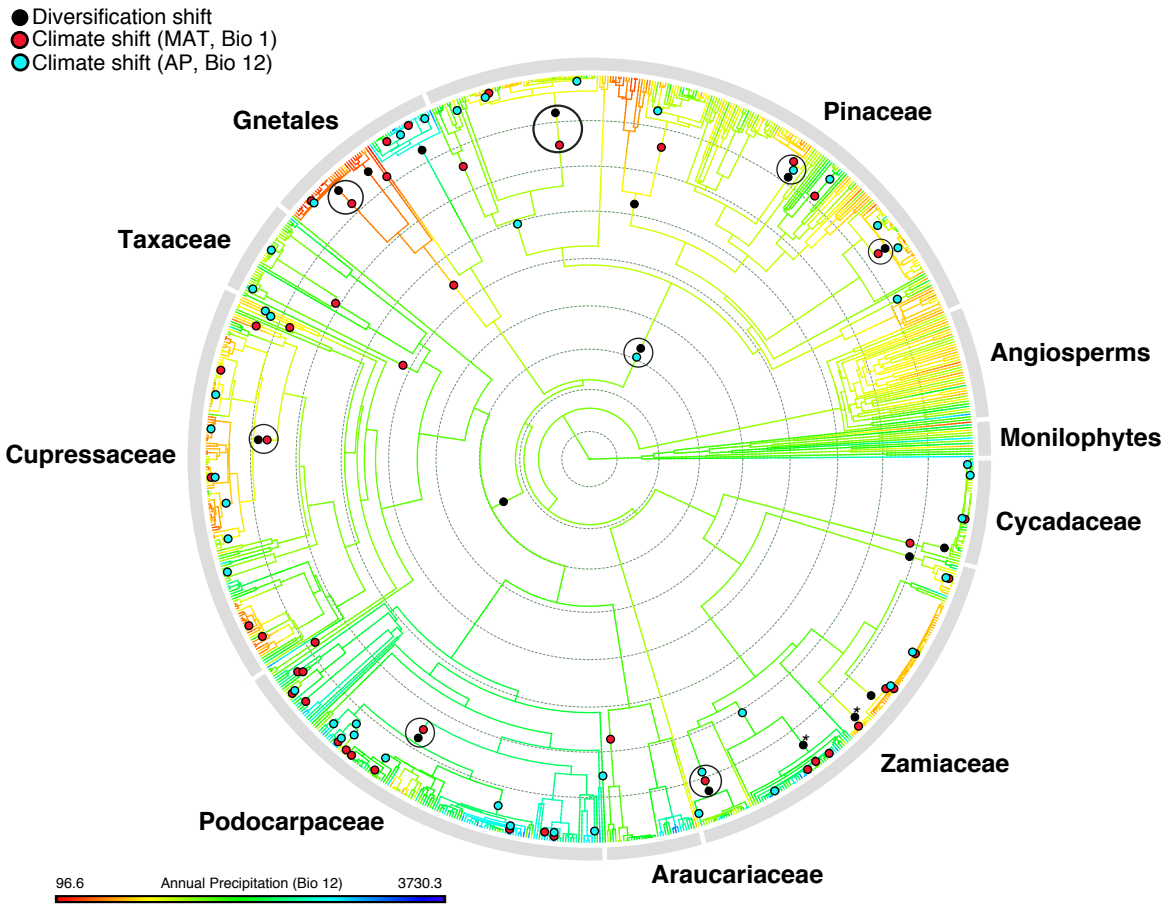



**Figure 3**. **Gene-tree conflict vs. phenotypic rates**. Plots juxtaposing rates of phenotypic evolution and levels of gene-tree conflict at corresponding branches across gymnosperm phylogeny ($n$ = 119). Note that the conflict node highlighted as corresponding with the *Juniperus* rate node is its immediate parent, comprising *Cupressus sempervirens*, *Juniperus*. The x-axis represents time in millions of years.




528

529 **Figure 4**. **Climatic and diversification shifts across gymnosperms**. Ancestral reconstruction of
530 annual precipitation (Bio 12) on the supermatrix phylogeny, showing diversification rate shifts
531 (black circles) and shifts in rates of climatic evolution (Bio 1, mean annual temperature = red
532 circle; Bio 12, annual precipitation = cyan circle) inferred using BAMM. The diversification
533 shifts marked with asterisks (corresponding to *Macrozamia* and *Zamia*) show concerted rate
534 shifts with PC2 of the climatic variables, bringing the number of directly concerted
535 diversification and climate shifts to ten. See Supplementary Fig. 22 for species tip labels.

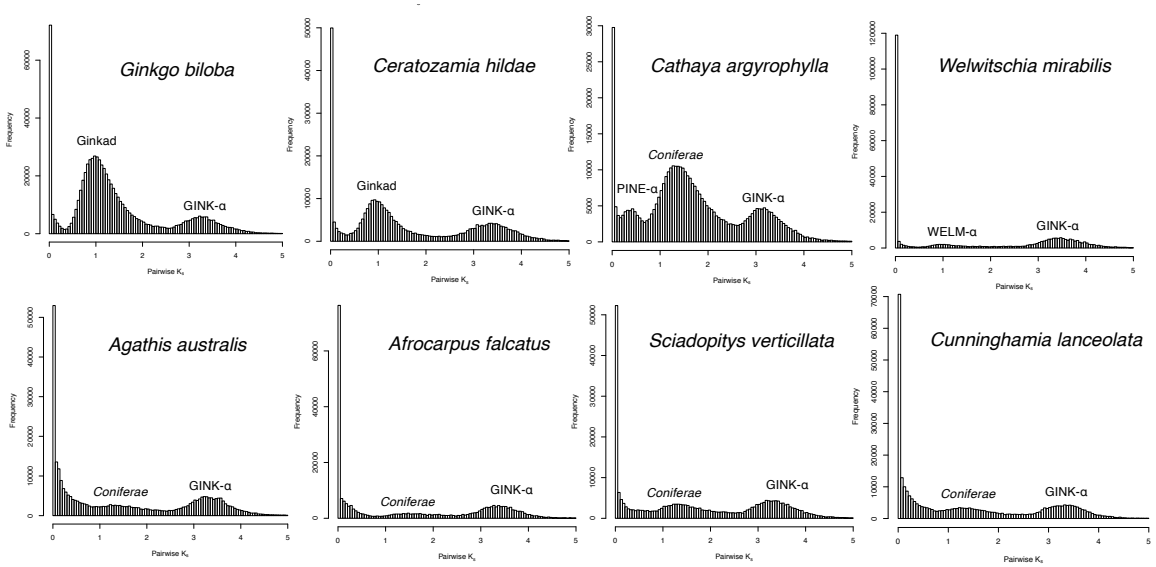
536

537

538

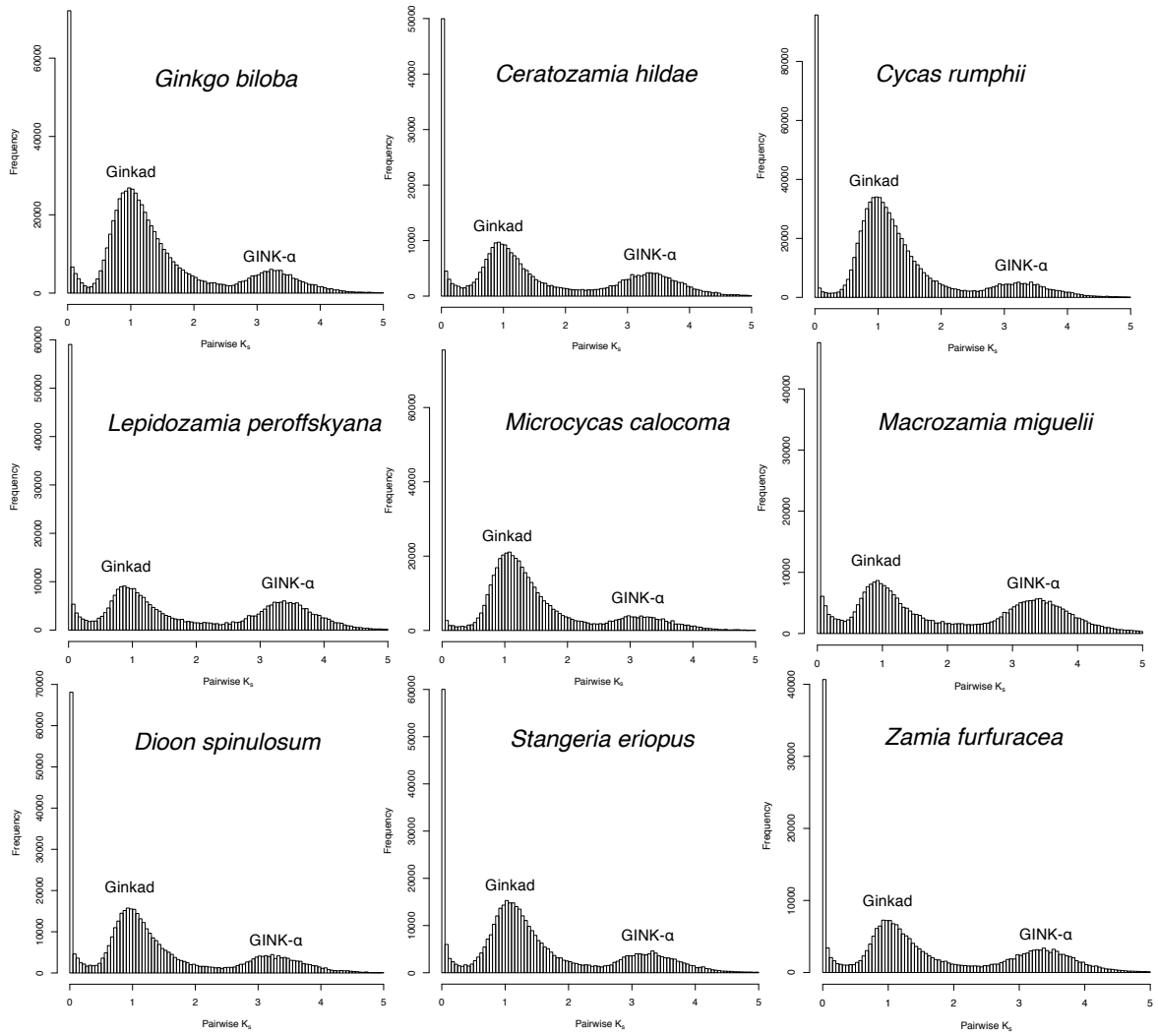
539

540




541

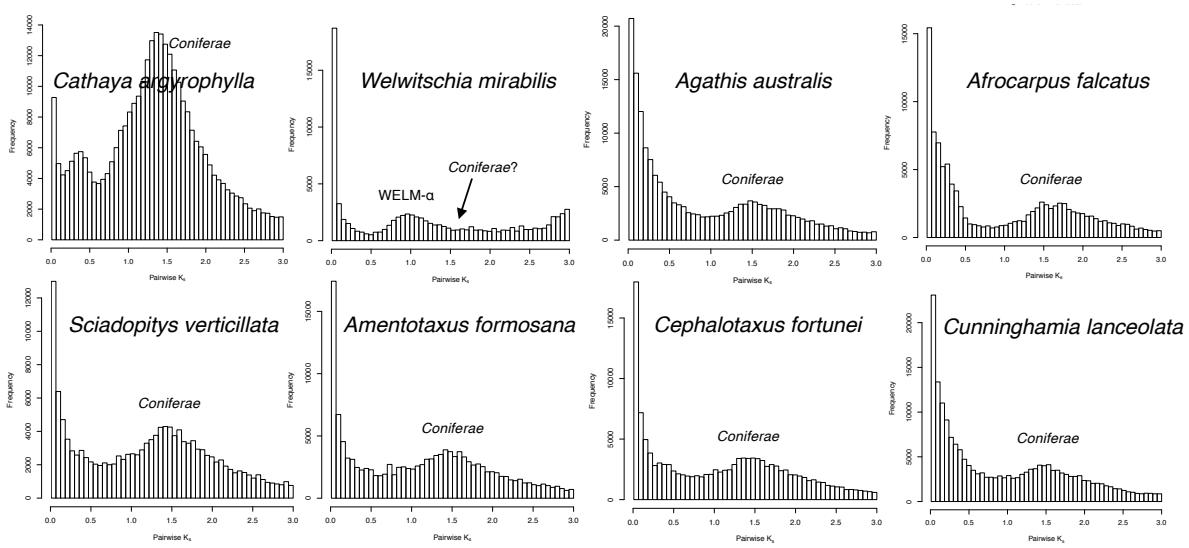
542 **Extended Data Figure 1**. **Exemplar *Ks* plots of major gymnosperm clades**. 'GINK-α'
543 represents the whole-genome duplication (WGD) event shared by seed plants. 'Ginkad'
544 represents a putative WGD event shared by cycads and *Ginkgo*; the putative WGD event for
545 *Coniferae* is also labeled. 'PINE-α' represents the Pinaceae WGD. 'WELM-α' represents the
546 *Welwitschia* WGD.




547

548 **Extended Data Figure 2**. **Exemplar *Ks* plots of the *Ginkgo* + cycad clade**. 'GINK-α'
549 represents the seed plant WGD. 'Ginkad' represents a putative WGD event shared by cycads and
550 *Ginkgo*.

551




552

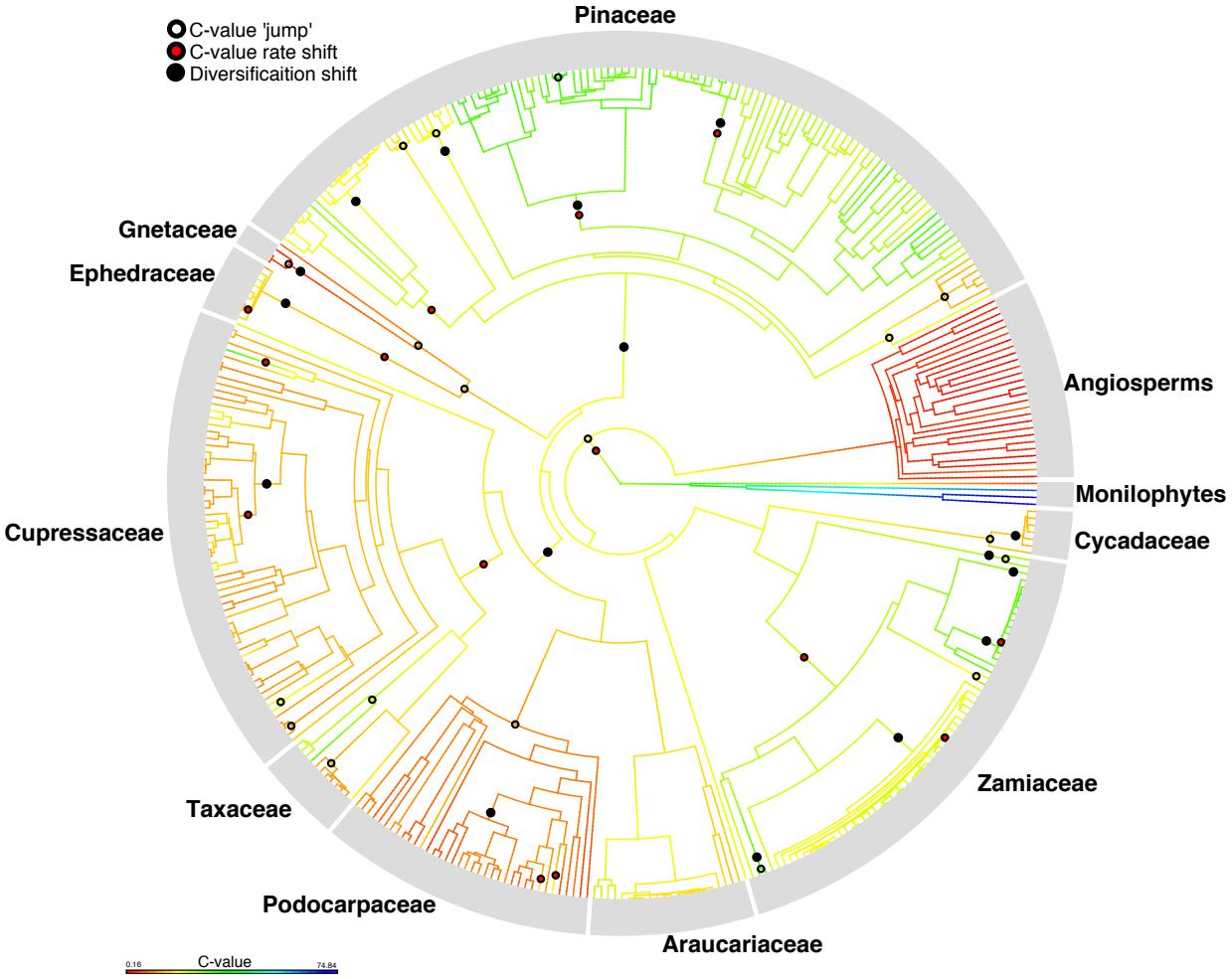
553 **Extended Data Figure 3**. **Exemplar *Ks* plots of *Coniferae***. The putative WGD event for

554 *Coniferae* is highlighted. The *Welwitschia* WGD is also labeled ('WELM-α'). *Welwitschia* lacks

555 a clear signal of the *Coniferae* peak (at ~1.5 *Ks*), as do other members of Gnetales (see

556 Supplementary Fig. 1).

557




558

559 **Extended Data Figure 4**. **Genome size evolution in gymnosperms**. Ancestral reconstruction of
560 genome size (C-value) on the pruned supermatrix phylogeny, showing BAMM rate shifts (red
561 circles) and jumps (i.e., extreme differences in ancestor-descendent values; white circles) in
562 genome size evolution, as well as BAMM diversification shifts (larger black circles).
563 Supplementary Fig. 28 shows this figure with species tip labels.

564

565

566

567

568

569

570

571 # Methods

572 **Sampling overview**

573 We generated two major phylogenetic datasets. The first is a transcriptomic and genomic dataset
574 comprising 144 species (including 121 ingroup species, representing 83 of the 85 gymnosperm
575 genera recognized by the Kew checklist: https://wcsp.science.kew.org/home.do); this was used to
576 resolve broad-scale phylogenetic relationships, characterize patterns of phylogenomic conflict,
577 and infer hotspots of gene duplication. Representatives of angiosperms, monilophytes, and
578 lycophytes were used as outgroups. In all, 77 transcriptomes were newly generated for this study;
579 the remaining transcriptomes were obtained from the One Thousand Plants (1KP) Consortium
580 (https://sites.google.com/a/ualberta.ca/onekp) and several genomes were obtained from the
581 Phytozome database v9.1[58]. See Supplementary Table 1 for transcriptomic and genomic
582 sampling details. The two missing gymnosperm genera, *Pectinopitys* and *Pherosphaera*,
583 represent genera nested within Podocarpaceae and thus their absence should not impact the
584 resolution of major relationships. The newly generated transcriptomes were deposited in the
585 Sequence Read Archive (provided upon acceptance). We also generated a phenotypic dataset of
586 148 traits for the seed plant species in the transcriptomic dataset to permit examination of the
587 correspondence of different aspects of genomic and phenotypic evolution; the trait dataset is
588 described below.

589 The other phylogenetic dataset is a plastid supermatrix, based on publicly available
590 plastomes and plastid genes in NCBI (data deposit), as well as 32 plastid genomes newly
591 generated through genome skimming (Supplementary Table 2). In total, this dataset includes 78
592 plastid genes across 1055 accessions, including 995 ingroup accessions. The ingroup names were
593 checked against the Kew gymnosperm checklist (https://wcsp.science.kew.org/home.do) to
594 resolve issues with synonymy. In total, 890 gymnosperm species are represented, comprising ca.
595 82% of the 1090 recognized gymnosperm species[56]; our sampling also includes numerous
596 infraspecific taxa (i.e., varieties and subspecies; https://wcsp.science.kew.org/home.do).

597 The supermatrix dataset was used to generate a comprehensive dated phylogeny for
598 gymnosperms for reconstructions of diversification shifts, climatic evolution, genome size
599 evolution, and chromosome evolution. We used exclusively plastid data for two reasons: (1)
600 plastid genes are more widely sequenced across gymnosperms, allowing for greater species
601 representation; and (2) conflict within phylogenetic datasets can greatly (and negatively) impact

branch length estimation and downstream analyses such as divergence-time estimation[59]. While studies show that the plastome is not without conflict[60], it clearly shows more minimal levels of conflict compared to nuclear genes. Plastid relationships of gymnosperms, as shown by previous studies[56,61] and our results here, largely agree with those of the nuclear genome, and thus a plastid supermatrix tree should accurately capture major macroevolutionary patterns.

**Transcriptome and plastome sequencing and assembly**

To generate the new transcriptomes, total RNAs were isolated from young buds or leaves using the Spectrum™ Plant Total RNA Kit (SIGMA-ALDRICH). Approximately 5 ug of total RNAs were used to construct cDNA libraries (NEBNext Ultra Directional RNA Library Prep Kit for Illumina, Illumina, San Diego, CA, USA). Transcriptome sequencing was performed using the Illumina Hi-Seq 2000 platform. Paired-end reads of 150 bp were *de novo* assembled with Trinity v2.4.0 (Grabherr et al., 2011) using default parameters. Transcriptomes from the 1KP project were previously assembled (Wickett 2014; Leebens-Mack et al. 2019) using SOAPdenovo v1 (Li et al. 2010), and amino acid sequences for three gymnosperm species derived from genome annotations were downloaded from the Phytozome database v9.1[58]. All transcripts were translated using TransDecoder v5.3.0[62], and then cd-hit-est[63] (-c 0.99 –n 5) was used to reduce sequence redundancy in all CDS.

Plastid genomes were sequenced and assembled as follows. Total genomic DNA was extracted from fresh leaves using the CTAB method[64]. Approximately one microgram of total genomic DNA from each sample was used for ~400-bp library construction with the NEBNext DNA Library Prep Kit (New England Biolabs, Ipswich, MA, USA) and Covaris-based fragmentation. The Illumina MiSeq instrument (Illumina, San Diego, CA, USA) was used to carry out 150-bp, paired-end sequencing. We assembled the plastomes using both the Organelle Genome Assembler (OGA) pipeline[65] and SPAdes v3.13.0[66], the latter with the “careful”-option and kmers of 61, 81, 101, and 121. In order to validate the plastome assembly, we mapped all paired-end reads to the consensus sequence of the assembled plastomes with the local-sensitive option of Bowtie v2.3.2[67]. Plastid Genome Annotator[68] was used to perform plastome annotation, coupled with manual correction in Geneious v8.0.2 (https://www.geneious.com/).

**Phylogenomic workflow**

633 **Homology and orthology inference.** We followed two major procedures for homology and
634 orthology inference in the transcriptomic dataset. First, we employed the hierarchical homology
635 inference approach of Walker et al.[69]. Given the size of our transcriptomic dataset (including 144
636 transcriptomes), performing an all-by-all BLAST search across all samples for initial homology
637 assessment would have been computationally prohibitive. The Walker et al.[69] approach
638 overcomes this problem by conducting all-by-all BLAST searches on subsets of the data (*tip*
639 *clustering*), which are then combined using a post-order (tip to root) tree traversal method (*node*
640 *clustering*). We divided our samples into 11 subgroups for tip clustering: **(1)** monilophytes and
641 lycophytes (7 samples), **(2)** angiosperms (16 samples), **(3)** cycads and *Ginkgo* (14 samples), **(4)**
642 Gnetales (5 samples), **(5)** Pinaceae (22 samples), **(6)** Podocarpaceae (23 samples), **(7)**
643 Araucariaceae (6 samples), **(8)** *Sciadopitys* and Taxaceae (12 samples), **(9)** 'basal' Cupressaceae
644 (11 samples), **(10)** cupressoid Cupressaceae (17 samples), and **(11)** callitroid Cupressaceae (11
645 samples); see Supplementary Table 1 for lists of species included in each grouping. For tip
646 clustering, for each subset, we conducted an all-by-all BLASTP search[70] with an e-value cutoff
647 of 10. The tip clusters were then sequentially combined in a post-order fashion (node-clustering)
648 using the scripts developed by Walker et al.[69] (https://github.com/jfwalker/Clustering).
649 For homolog tree and orthology inference from the combined clusters, we followed the
650 Yang and Smith[71] pipeline (scripts available at
651 https://bitbucket.org/yangya/phylogenomic_dataset_construction/src/master/), which we briefly
652 outline here. Fasta files were written from each cluster. The clusters were then aligned using
653 MAFFT v7[72] and cleaned using the phyx[73] program 'pxclsq' (retaining alignment columns with
654 at least 10% occupancy), after which homolog trees were built using FastTree v2 with the WAG
655 model[74]. The trees were then pruned of extreme long branches, monophyletic tips were masked
656 (i.e., only one sequence was retained of species/sample-specific alleles/paralogs/duplications),
657 and then long internal branches (representing deep paralogs) were pruned. Fasta files were then
658 written again from the resulting subtrees, and two more iterations of alignment, homolog tree
659 construction, and homolog tree pruning were conducted. A final stage of homolog tree
660 construction was then conducted using RAxML v8.2.12[75] to generate more robust homolog trees.
661 The resulting 1,992 homolog trees were then used for orthology inference and gene-duplication
662 mapping. For orthology inference, we used the 'RT' method of Yang and Smith[71], which extracts
663 ingroup clades and then cuts paralogs from root to tip. For this procedure, the three lycophyte

samples were designated as outgroups (and thus these were pruned and excluded from subsequent analyses), and the minimum ingroup sampling was set at 100. This approach resulted in a set of 790 orthologs. Nucleotide fasta files were then written from each extracted ortholog, re-aligned using MAFFT, and then cleaned by removing alignment columns with under 25% occupancy using 'pxclsq' in phyx.

**Coalescent species-tree inference.** Gene trees (from the 790 orthologs inferred above) were constructed using RAxML, using the GTRGAMMA model and 100 bootstrap replicates. The gene trees were then used for species-tree analysis in ASTRAL-III[76]. Prior to the ASTRAL analysis, low supported branches (<10 BS) in the gene trees were collapsed using Newick utilities[77], as this has been shown to increase accuracy[76]. ASTRAL was then run under default settings, producing a 'species tree' tree with local posterior probability (LPP) support values for the bipartitions and branch lengths in coalescent units (Supplementary Fig. 29). The resulting phylogeny was used as a framework for various subsequent analyses, including conflict analysis, analysis of gene duplications, inference of whole-genome duplications, morphological trait reconstructions, and comparisons of phenotypic innovation and rates of phenotypic evolution with levels of genomic conflict and gene duplications.

**Concatenation species-tree inference.** The individual gene/ortholog alignments were concatenated using the phyx program 'pxcat'. We then inferred phylogenies from the concatenated nucleotide alignment (comprising 1,444,037 bp) using IQ-TREE[78] with the GTRGAMMA model and ultrafast bootstrapping (1000 replicates). Two analyses were conducted: one unpartitioned, and the other with separate model parameters estimated for each gene region[79] (Supplementary Fig. 30).

**Conflict analysis.** To examine patterns of gene-tree conflict across gymnosperm phylogeny, we used the program PhyParts[80], which maps gene trees onto an input species tree and summarizes the number of gene trees in concordance or conflict with each bipartition. When gene trees have inadequate species sampling or bootstrap support (if a support threshold is specified) to speak to a particular bipartition, those are considered uninformative. We use the python script 'phypartspiecharts.py'

695 (https://github.com/mossmatters/phyloscripts/tree/master/phypartspiecharts) to visualize the
696 PhyParts results, which are shown in Fig. 1. The levels of conflict calculated for each bipartition
697 were also plotted against corresponding levels of morphological innovation (i.e., the number of
698 new states appearing at each node; see below). We also used the program 'minority_report.py'
699 (https://github.com/mossmatters/phyloscripts/tree/master/minorityreport) to examine the number
700 of gene trees supporting alternative placements for particular lineages (namely, *Ginkgo*).
701

702 **Inference of whole-genome duplication**

703 We used a multi-faceted approach to examine genomic changes and infer putative WGDs across
704 gymnosperm phylogeny, employing gene-duplication mapping, plots of synonymous
705 substitutions per site (*Ks*), and reconstructions of chromosome number and genome size (C-
706 value). Gene duplications were mapped onto the species tree topology using the scripts
707 (https://bitbucket.org/blackrim/clustering/src/master/) and method of Yang et al.[28] This involved
708 extracting rooted orthogroups from the homolog trees (with lycophytes designated as the
709 outgroup), which were then mapped onto the species tree topology. We used two methods of
710 mapping: one maps duplications to the most recent common ancestor (MRCA) in the species
711 tree, with a bootstrap filter of 50% (Supplementary Fig. 31); the other imposes a topology filter
712 such that duplications are only mapped when the sister clade of the gene duplication node
713 contains at least a subset of the taxa present in that clade in the species tree[28] (Supplementary
714 Fig. 32).

715 *Ks* plots of within-taxon paralog pairs were generated for the newly sequenced
716 transcriptomes. First, CD-HIT (-c 0.99 –n 5) was used to reduce highly similar sequences, after
717 which an all-by-all BLASTP search was carried out within each taxon (e-value cutoff = 10). The
718 results were then filtered such that sequences with ten or more hits were removed (to avoid
719 overrepresentation of large gene families), as were hits with pident/nident less than 20/50%. *Ks*
720 values were then calculated from the CDS of the remaining paralog pairs, using the pipeline
721 https://github.com/tanghaibao/bio-pipeline/tree/master/synonymous_calculation, and plotted
722 using scripts from Yang et al.[28] The *Ks* plots (Supplementary Figs. 1–14; data deposit) were
723 visually inspected to identify clade-specific shared peaks, which were compared against the
724 results from gene-duplication mapping, as well as the reconstructions of chromosome number
725 and genome size (outlined below).

726

727 **Supermatrix tree construction and dating**

728 To assemble a comprehensive species-level phylogeny of gymnosperms, we mined plastid gene
729 and genome sequences available on NCBI and combined these with genes extracted from the 32
730 newly sequenced and assembled plastomes. While the plastomes from NCBI were directly
731 downloaded, individual plastid genes were mined from NCBI using the program PyPhlawd[81].
732 Large-scale plastid gene clusters obtained using PyPhlawd were then combined with the
733 corresponding individual gene sequences derived from the plastomes.

734 The plastid genes were aligned individually using MAFFT v7[72], after which alignment
735 columns with under 25% occupancy were removed using the phyx function 'pxclsq'. The
736 alignments were then concatenated using the phyx function 'pxcat'. The final concatenated
737 matrix included 20.35% matrix occupancy (i.e., 79.65% missing data). Prior to phylogenetic
738 analysis, PartitionFinder v2[82] was used to determine the optimal model partitioning strategy for
739 the concatenated alignment. Phylogenetic analyses were conducted using RAxML-ng[83] with a
740 topological constraint enforcing major relationships to match the transcriptomic species tree
741 topology: ((*Ginkgo*,cycads),((Gnetales,Pinaceae),((Araucariaceae,Podocarpaceae),
742 (Sciadopityaceae, (Taxaceae,Cupressaceae))))). When unconstrained, the plastome tree only
743 differed from the constraint in the placement of Gnetales. Using RAxML-ng, we conducted 25
744 independent searches for the maximum likelihood tree ("raxml-ng --search1"), from which the
745 tree with the best likelihood score was selected for downstream analyses of diversification,
746 climate evolution, and genome evolution (chromosome number and cval), although all 25 trees
747 were dated in treePL[84] to obtain confidence intervals on node ages (see below). We also inferred
748 100 bootstrap trees to evaluate topological support for the supermatrix phylogeny.

749

750 **Divergence dating.** Given the scale of the supermatrix tree (1055 terminals), typical methods of
751 Bayesian divergence dating were unfeasible. Therefore, we used treePL[84] to generate a dated
752 phylogeny. We used extensive fossil evidence to inform our calibration scheme (see
753 Supplementary Information for detailed discussion of the fossils used, their justification, and
754 references). We conducted independent analyses on the 25 inferred ML trees using the same set
755 of calibrations as a means of obtaining confidence intervals for inferred node ages
756 (Supplementary Fig. 33). Cross-validation was performed for each analysis to determine the

757 optimal smoothing parameter. Node ages from the dated supermatrix tree were then applied as
758 fixed ages (also using treePL) to the corresponding nodes in the transcriptomic species tree. This
759 allowed us to place subsequent analyses involving the transcriptomic phylogeny (i.e.,
760 examinations of conflict and rates of morphological evolution) in a temporal context.
761
762
763 **Chromosome number and genome size reconstructions**
764 Chromosome counts were obtained from the Chromosome Count Database[49]
765 (http://ccdb.tau.ac.il; accessed 9 December 2019), and the parsed (haploid) counts were used for
766 subsequent analysis. In cases where several different counts were recorded for a given species,
767 the most frequently occurring count was used. Measurements of genome size (i.e., C-values, the
768 amount of DNA in the haploid genome) were obtained from the Plant DNA C-values Database[50]
769 (https://cvalues.science.kew.org; accessed 9 December 2019). The dated supermatrix phylogeny
770 was pruned to match the sampling in each dataset (chromosome counts and C-values), and these
771 two trees were used in the analyses outlined below. The chromosome count tree/dataset included
772 478 species, and the cval tree/dataset include 373 species. The chromosome count and C-value
773 datasets examined are available in the data deposit.
774 Reconstructions of ancestral chromosome number were implemented in ChromEvol
775 v2.0[81]. We examined the data under eight different available models, each including different
776 combinations of the rate parameters *gain* (ascending dysploidy), *loss* (descending dysploidy),
777 *duplication* (polyploidization), and *demi-duplication* (multiplication of the chromosome number
778 by a factor of 1.5), with the rates for each treated either as constant or linear. The model with the
779 lowest Akaike information criterion (AIC) score (i.e., 'LINEAR RATE DEMI EST') was used in
780 a more thorough analysis involving default settings and 100,000 simulations. We then visualized
781 changes in chromosome number across the tree using FigTree (Supplementary Fig. 15) and
782 compared changes against reconstructions of genome size and gene duplication.
783 Ancestral reconstructions of genome size (using C-values) were implemented in R[86] (R
784 Core Team 2020) using the 'fastAnc' function in phytools[87]. The C-value data were not
785 transformed prior to analysis as they conform to a more-or-less normal distribution. To identify
786 the most extreme changes in genome size (which we call C-value 'jumps'), we then calculated
787 the difference between each node's reconstructed value and that of its immediate parent node,

788 and selected branches in the 95th percentile of the distribution of parent-child C-value
789 differences. These are shown in Extended Data Fig. 4. We also examined rates of genome size
790 evolution using BAMM[88,89], using the same tree as above. The priors for the BAMM analysis
791 were determined using the function 'setBAMMpriors' in the BAMMtools package[90]. The
792 analysis was run for one billion generations using the phenotypic evolution model, with
793 convergence assessed using BAMMtools. Rate shifts from the maximum a posteriori probability
794 (MAP) shift configuration are shown in Extended Data Fig. 4.
795
796 **Species diversification analysis**
797 We used BAMM to detect diversification rate shifts across gymnosperm phylogeny, using the
798 dated supermatrix tree as input. The analysis employed the speciation-extinction model, with the
799 priors determined using the function 'setBAMMpriors' in BAMMtools. We also specified
800 sampling fractions specified for each family of *Acrogymnospermae*. This analysis was initially
801 run for 1 billion generations, but then was re-run for 2 billion generations in order to achieve
802 convergence. Rate shifts from the MAP shift configuration are shown in Fig. 1, 4, and
803 Supplementary Fig. 16.
804
805 **Reconstructions of climatic occupancy evolution**
806 Geographic occurrences of gymnosperms were pulled from GBIF (https://www.gbif.org;
807 accessed 15 May 2020), and then compared against occurrences from iDigBio
808 (https://www.idigbio.org/portal/search) to check for available data of the species missing from
809 GBIF. Occurrences were then cleaned using the R package CoordinateCleaner[91], removing
810 duplicate records as well as those with problematic coordinates (e.g., non-terrestrial, identical
811 lat/lon, located within major cities or biodiversity institutions). Synonymy was reconciled using
812 data from the Kew checklist (https://wcsp.science.kew.org/home.do). The occurrences were then
813 used to extract climatic values from the 19 WorldClim bioclimatic variables, and means were
814 calculated for each species for each climatic variable. We dropped tips from the dated
815 supermatrix tree that were absent from the climatic dataset sampling, resulting in a matched
816 sampling of 908 species for subsequent climatic analyses. We then conducted a principal
817 component analysis in R, and principal components (PC) 1 and 2 were used in subsequent

818 analyses of climatic evolution. We also conducted analyses of mean annual temperature (Bio1)
819 and annual precipitation (Bio12), using the mean values for each species.
820 As with the genome size analyses, for the climate analysis we reconstructed major jumps
821 and rate shifts in climatic evolution, involving separate analyses for PC1, PC2, Bio1, and Bio12.
822 Jumps were determined by reconstructing ancestral climatic values (using the 'fastAnc' function
823 in phytools), calculating the difference between each node and its immediate parent, and
824 identifying those branches in the 95th percentile. Extreme climatic jumps for PC1, PC2, Bio1,
825 and Bio12 are shown in Supplementary Figs. 26 and 27. BAMM analysis of rates of climatic
826 occupancy evolution were also conducted on PC1, PC2, Bio1, and Bio12, with each run for 1
827 billion generations, employing the phenotypic evolution model, with priors determined using
828 BAMMtools. Rate shifts from the MAP shift configuration for Bio1 and Bio 12 are shown in
829 Fig. 4 and Supplementary Figs. 22 and 23; those of PC1 and PC2 are shown in Supplementary
830 Figs. 24 and 25.
831 Although analysis of historical climatic information would be ideal for understanding the
832 climatic context of previous diversification shifts in greater detail, this absence of reliable
833 estimates spanning the depth of gymnosperm evolutionary history makes this, at the present, not
834 feasible given the scope of our study. Thus, we only used current global climate data for our
835 reconstructions. However, the use of historical climate data to investigate particular Cenozoic
836 radiations of gymnosperms would be a promising avenue for future research.
837
838 **Morphological reconstructions and genomic comparisons**
839 A phenotypic dataset comprising 148 traits across the 137 seed plant species included in the
840 transcriptomic dataset was assembled to examine the correspondence of phenotypic and genomic
841 evolution (matrix available in the data deposit). Many of the traits examined were also included
842 in previous morphological datasets. However, the taxonomic sampling in some of these previous
843 datasets was limited; furthermore, many traits from clade-specific datasets needed to be modified
844 and reconciled with similar traits in other datasets in order to establish characters applicable
845 across all sampled taxa. The traits examined (and their corresponding states) are listed in
846 Supplementary Information, and we indicate how these relate to traits in previously published
847 datasets. The primary source datasets were those of Hart[92], Hilton and Bateman[93], Mao et al.[94],
848 Escapa and Catalano[95], Coiro and Pott[96], Herrera et al.[97,98], Gernandt et al.[99], and Andruchow-

849 Colombo et al.[100] See Supplementary Information for a complete list of references from which
850 phenotypic data were obtained. We note that while determining homology for character
851 delimitation in seed plants is a major challenge[101], we do not believe that instances of
852 misidentified homology (i.e., treating independently evolved structures as a single trait or
853 expression of a trait) should necessarily undermine our goals here, given that we are not using
854 this trait set to reconstruct phylogeny. Our goal is to identify state changes along branches, and
855 even in instances of conflated homology, e.g., with two distantly related lineages scored with the
856 same state for a particular trait despite independent origins of the relevant structures, a state
857 change should nevertheless be 'correctly' reconstructed and recorded for each branch.
858 Morphological reconstructions on the species-tree were performed using Maximum
859 Parsimony, following the procedure of Parins-Fukuchi et al.[13], briefly outlined here. Each trait
860 was reconstructed on the phylogeny, and then the number of state changes along each branch
861 was tallied—this is considered the degree or level of phenotypic 'innovation' or 'novelty' along
862 each branch. To determine rates of phenotypic evolution for each branch, we divided the number
863 of state changes by the duration of branch (in million years). We then plotted levels of
864 morphological innovation and rates of phenotypic evolution for each branch/node (with the x-
865 axis representing time or node age), and compared these against plots of the same nodes showing
866 their levels of gene duplication and gene-tree conflict (Figs. 2 and 3), allowing for visual
867 inspection of these patterns and their correspondence through time. Because Gnetales was
868 detected as a major outlier, with extreme levels of morphological innovation perhaps driven by
869 life history shifts, we also conducted morphological reconstructions excluding Gnetales to ensure
870 that the inclusion/placement of Gnetales was not significantly biasing reconstructions for other
871 near-by nodes (Supplementary Fig. 19).
872 On the basis of the visual correspondences observed from the plots noted above, we then
873 explicitly tested the relationships of (1) phenotypic innovation vs. gene duplication and (2)
874 phenotypic rates vs. gene tree conflict by conducting linear regression and permutation analyses
875 of each pair of variables. Linear regression was performed using the 'lm' function in R, treating
876 phenotypic innovation and rate as the dependent variables in the respective analyses. For the
877 examination of innovation vs. duplication, we conducted an analysis including all nodes
878 (Supplementary Fig. 34), as well as one excluding two major outlier nodes: Gnetales and
879 conifers (Supplementary Fig. 17). As noted in the main text, we believe exclusion of these

outliers is warranted given that (1) Gnetales has experienced major changes in genome size (i.e., reduction) and life history with generally elevated rates of molecular evolution that might erase signatures of a relationship between gene duplication and morphological innovation, and (2) trait reconstruction for conifers is complicated by its inclusion of Gnetales (a highly derived lineage) and the prevalence of features shared among conifers and *Ginkgo*, which may have been involved in a reticulation event early in the evolution of the conifer lineage based on patterns of gene-tree conflict.

The permutation test, developed by Parins-Fukuchi et al.[13], identifies nodes with significantly elevated values for a particular variable by resampling (with replacement) observed values 1000 times and using the means from each replicate to create an empirical null distribution. We executed this procedure for each of the four variables considered above (duplication levels, conflict levels, innovation levels, and phenotypic rate), with nodes/values falling in the 95th percentile deemed significant. Then, to examine the strength of the association between (1) innovations and duplications and (2) rates and conflict, we determined whether the number of nodes with significantly high values for both of the variables in question (e.g., innovations and duplications) departed from the expectation if the same number of nodes was randomly sampled (Supplementary Figs. 18, 21). This permutation test was designed specifically to examine whether extreme/significant values (in, for example, conflict and rates) tend to coincide, even if there is not a strict positive monotonic relationship between the two variables across the spectrum of values. Finally, we also used this test to determine if diversification shifts tend to co-occur with significant/extreme levels of phenotypic innovation (Supplementary Fig. 35).

## Data availability

The newly generated raw sequence data is available at the NCBI Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra) under BioProject accessions <number provided upon acceptance> (transcriptomic samples) and <number provided upon acceptance> (genome skimming samples). The newly assembled plastid genomes are also available at NCBI (https://www.ncbi.nlm.nih.gov); see Supplementary Table 2 for sample accession numbers. Sequence alignments, phylogenies, *Ks* plots, phenotypic trait data, and other data analyzed (chromosome counts, c-values) are available on figshare

(https://figshare.com/s/8431009039a19d1dc8c8).

58. Goodstein, D. M. et al. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res*. **40**, 1178–1186 (2012).
59. Mendes, F. K., Hahn, M. W. Gene tree discordance causes apparent substitution rate variation. *Syst*. *Biol*. **65**, 711–721 (2016).
60. Walker, J. F., Walker-Hale, N., Vargas, O. M., Larson, D. A. & Stull, G. W. Characterizing gene tree conflict in plastome-inferred phylogenies. *PeerJ* **7**, e7747 (2019).
61. Zhong, B., Yonezawa, T., Zhong, Y., Hasegawa, M. The position of Gnetales among seed plants: overcoming pitfalls of chloroplast phylogenomics. *Mol*. *Biol*. *Evol*. **27**, 2855–2863 (2010).
62. Haas, B.J. et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat*. *Protoc*. **8**, 1494–1512 (2013).
63. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W., 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150-3152 (2012).
64. Doyle, J. J. & Doyle, J. L. A rapid DNA isolation procedure for small quantities for fresh leaf tissue. *Phytochemical Bulletin* **19**, 11–15 (1987).
65. Qu, X. J., Fan, S. J., Wicke, S. & Yi, T. S. 2019. Plastome reduction in the only parasitic gymnosperm *Parasitaxus* is due to losses of photosynthesis but not housekeeping genes and apparently involves the secondary gain of a large inverted repeat. *Genome Biol*. *Evol*. **11**, 2789–2796 (2019).
66. Bankevich, A. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J*. *Comput*. *Biol*. **19**, 455–477 (2012).
67. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat*. *Methods* **9**, 357 (2012).
68. Qu, X. J., Moore, M. J., Li, D. Z. & Yi, T. S. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods* **15**, 50 (2019).
69. Walker, J. F. et al. From cacti to carnivores: improved phylotranscriptomic sampling and hierarchical homology inference provide further insights into the evolution of Caryophyllales. *Am*. *J*. *Bot*. **105**, 446–462 (2018).

941 70. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database
942 search programs. *Nucleic Acids Res*. **25**, 3389–3402 (1997).
943 71. Yang, Y. & Smith, S. A. Orthology inference in nonmodel organisms using transcriptomes
944 and low-coverage genomes: improving accuracy and matrix occupancy for
945 phylogenomics. *Mol. Biol. Evol*. **31**, 3081–3092 (2014).
946 72. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7:
947 Improvements in Performance and Usability. *Mol. Biol. Evol*. **30**, 772–780 (2013).
948 73. Brown, J. W., Walker, J. F. & Smith, S. A. Phyx: phylogenetic tools for
949 unix. *Bioinformatics* **33**, 1886–1888 (2017).
950 74. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 – Approximately Maximum-Likelihood
951 Trees for Large Alignments. *PLoS One* **5**, e9490 (2010).
952 75. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
953 phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
954 76. Zhang, C. et al. ASTRAL-III: polynomial time species tree reconstruction from partially
955 resolved gene trees. *BMC Bioinformatics* **19**, 153 (2018).
956 77. Junier, T. & Zdobnov, E. M. The Newick utilities: high-throughput phylogenetic tree
957 processing in the UNIX shell. *Bioinformatics* **26**, 1669–1670 (2010).
958 78. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and
959 effective stochastic algorithm for estimating maximum likelihood phylogenies. *Mol. Biol.*
960 *Evol*. **32**, 268–274 (2015).
961 79. Chernomor, O., von Haeseler, A. & Minh, B. Q. Terrace aware data structure for
962 phylogenomic inference from supermatrices. *Syst. Biol*. **65**, 997–1008 (2016).
963 80. Smith, S. A., Moore, M. J., Brown, J. W. & Yang, Y. Analysis of phylogenomic datasets
964 reveals conflict, concordance, and gene duplications with examples from animals and
965 plants. *BMC Evol. Biol*. **15**, 150 (2015).
966 81. Smith, S. A. & Walker, J. F. Pyphlawd: A python tool for phylogenetic dataset
967 construction. *Methods Ecol. Evol*. **10**, 104–108 (2019).
968 82. Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T. & Calcott, B. PartitionFinder 2: new
969 methods for selecting partitioned models of evolution for molecular and morphological
970 phylogenetic analyses. *Mol. Biol. Evol*. **34**, 772–773 (2017).

971 83. Kozlov, A.M., Darriba, D., Flouri, T., Morel, B. & Stamatakis, A. RAxML-NG: a fast,
972 scalable and user-friendly tool for maximum likelihood phylogenetic
973 inference. *Bioinformatics* **35**, 4453–4455 (2019).
974 84. Smith, S. A. & O'Meara, B. C. treePL: divergence time estimation using penalized likelihood
975 for large phylogenies. *Bioinformatics* **28**, 2689–2690 (2012).
976 85. Glick, L. & Mayrose, I. ChromEvol: Assessing the pattern of chromosome number Evolution
977 and the inference of polyploidy along a phylogeny. *Mol. Biol. Evol.* **31**, 1914–1922 (2014).
978 86. R Core Team. R: A language and environment for statistical computing. R Foundation for
979 Statistical Computing, Vienna, Austria. URL https://www.R-project.org (2020).
980 87. Revell, L. J. Phytools: An R package for phylogenetic comparative biology (and other
981 things). *Methods Ecol. Evol.* **3**, 217–223 (2012).
982 88. Rabosky, D. L. et al. Rates of speciation and morphological evolution are correlated across
983 the largest vertebrate radiation. *Nat. Commun.* **4**, 1–8 (2013).
984 89. Rabosky, D. L. Automatic detection of key innovations, rate shifts, and diversity-dependence
985 on phylogenetic trees. *PLoS One* **9**, e89543 (2014).
986 90. Rabosky, D. L. et al. BAMM tools: an R package for the analysis of evolutionary dynamics
987 on phylogenetic trees. *Methods Ecol. Evol.* **5**, 701–707 (2014).
988 91. Zizka, A. et al. CoordinateCleaner: Standardized cleaning of occurrence records from
989 biological collection databases. *Methods Ecol. Evol.* **10**, 744–751 (2019).
990 92. Hart, J. A. A cladistics analysis of conifers: preliminary results. *J. Arnold Arb.* **68**, 269–307
991 (1987).
992 93. Hilton, J. & Bateman, R. M. Pteridosperms are the backbone of seed-plant phylogeny. *J.*
993 *Torrey Bot. Soc.* **133**, 119–168.
994 94. Mao, K. et al. Distribution of living Cupressaceae reflects the breakup of Pangea. *Proc. Natl*
995 *Acad. Sci. USA* **109**, 7793–7798 (2012).
996 95. Escapa, I. H. & Catalano, S. A. Phylogenetic analysis of Araucariaceae: integrating
997 molecules, morphology, and fossils. *Int. J. Plant Sci.* **174**, 1153–1170 (2013).
998 96. Coiro, M. & Pott, C. *Eobowenia* gen. nov. from the Early Cretaceous of Patagonia: indication
999 for an early divergence of Bowenia? *BMC Evol. Biol.* **17**, 97 (2017).
1000 97. Herrera, F. et al. 2017. Cupressaceae conifers from the Early Cretaceous of Mongolia. *Int. J.*
1001 *Plant Sci.* **178**, 19–41 (2017).

98. Herrera, F. et al. 2020. Reconstructing *Krassilovia mongolica* supports recognition of a new and unusual group of Mesozoic conifers. *PLoS One* **15**, e0226779.

99. Gernandt, D. S. et al. Incorporating fossils into the Pinaceae tree of life. *Am. J. Bot.* **105**, 1329–1344 (2018).

100. Andruchow-Colombo, A., Wilf, P. & Escapa, I. H. A South American fossil relative of *Phyllocladus*: *Huncocladus laubenfelsii* gen. et sp. nov. (Podocarpaceae), from the early Eocene of Laguna del Hunco, Patagonia, Argentina. *Aust. Syst. Bot.* **32**, 290–309 (2019).

101. Nixon, K. C., Crepet, W. L., Stevenson, D. & Friis, E. M. A reevaluation of seed plant phylogeny. *Ann. Missouri Bot. Gard.* **81**: 484–533 (1994).