The Rise of Polyploids During Environmental Upheaval

Hengchi Chen[1,2,*], Fabricio Almeida-Silva[1,2], Garben Logghe[3,4], Steven Maere[1,2,*], Dries Bonte[3,*], and Yves Van de Peer[1,2,5,6,*]

[1] Department of Plant Biotechnology and Bioinformatics, Ghent University, 9052 Ghent, Belgium
[2] VIB Center for Plant Systems Biology, VIB, 9052 Ghent, Belgium
[3] Department of Biology, Terrestrial Ecology Unit, Ghent University, 9000 Ghent, Belgium
[4] Research Institute for Nature and Forest (INBO), Herman Teirlinckgebouw, Havenlaan 88 Bus 73, 1000 Brussels, Belgium
[5] Centre for Microbial Ecology and Genomics, Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Pretoria 0028, South Africa
[6] College of Horticulture, Academy for Advanced Interdisciplinary Studies, Nanjing Agricultural University, Nanjing 210095, China

* Correspondence: hengchi.chen@uni-goettingen.de (H.C.); steven.maere@psb.ugent.be (S.M.); dries.bonte@ugent.be (D.B.); yves.vandepeer@psb.ugent.be (Y.V.d.P.)

SUMMARY

Polyploidy, or whole-genome duplication (WGD), serves as both a significant evolutionary force and a potential evolutionary dead end, occurring extensively across the tree of life, particularly among angiosperms. Despite the prevalence of polyploid organisms, instances of ancient polyploidy are surprisingly rare, presenting a paradox that remains poorly understood. In this study, we constructed a comprehensive genomic dataset of 470 angiosperm species to address this issue. We developed a highly congruent evolutionary timescale and dated 132 ancient WGD events that are non-randomly distributed, revealing a clustering around pivotal periods of environmental upheaval and extinction. Notably, our findings highlight a strong correlation between waves of paleopolyploidization and significant events such as the Middle Miocene Disruption, the Eocene–Oligocene transition (EOT), the Paleocene–Eocene Thermal Maximum (PETM), the Cretaceous–Paleogene (K–Pg) extinction, and different oceanic anoxic (OAE) events, several of which can be linked to extinction events impacting flowering plant genera. By integrating multiple lines of evidence, we propose that polyploid organisms have an increased chance of survival during times of great environmental turmoil, a conclusion with important implications in the context of contemporary climate change and rapid global warming.

KEYWORDS

WGD, Polyploidy, Paleopolyploidization, Angiosperms, Environmental turmoil, Extinction

INTRODUCTION

The realization that the evolutionary trajectory of most, if not all, flowering plants has been profoundly influenced by multiple rounds of polyploidization - the result of whole genome duplication (WGD) - has been a cornerstone of early plant genomics research[1]. These "remarkable cycles of polyploidy" had been predicted by Haldane (1933)[2] in the early days of polyploidy research, but were later often considered as "evolutionary noise"[3] or even as evolutionary "dead ends"[4,5]. Such skepticism seemed partially justified: although extant polyploid organisms significantly enrich botanical diversity[6], and polyploidization is acknowledged as a major driver of plant speciation[7], the long-term persistence of polyploids appears to be a rare and infrequent phenomenon[8]. This infrequency suggests that substantial numbers of polyploid lineages may falter, failing to endure through evolutionary timescales. The presence of numerous instances of recurring polyploidy and recently originated polyploids, against a backdrop of scarce evidence for ancient polyploidy, presents an interesting evolutionary paradox[8,9].

Whole genome duplication carries transformative implications as a significant mutational occurrence. WGD can lead to immediate detrimental fitness effects, including complications in cell cycling, stunted growth, and reproductive incompatibility with ancestral (diploid) forms. Simultaneously, WGDs elicit an instantaneous increase in cell dimensions, various genomic and epigenomic alterations, along with an array of downstream effects that directly influence fitness[10]. Both disadvantages and potential benefits of WGDs have been extensively documented[11-13]. Although the long-term consequences of WGD are relatively easy to understand (i.e., extra genetic material created by WGD can lead to novel gene functions and more robust regulatory networks[14]), the immediate or short-term advantages following WGD are less apparent. Several recent studies have meticulously described the immediate effects of polyploidy[12,15,16]. Among the most consistent outcomes of WGD is amplified cell size, coupled with physiological modifications that can lead to increased resilience in stress conditions. For instance, neo-autotetraploid cultivars of *Arabidopsis thaliana* display increased tolerance to salt and drought stress compared to their diploid progenitors[17,18]. Similarly, tetraploid cultivars of rice (*Oryza sativa*) and citrange (*Citrus sinensis* × *Poncirus trifoliata*) exhibit increased tolerance to salt and drought stress due to WGD's modulation of genes involved in stress response and phytohormone signaling pathways[19,20]. As these examples of physiological and cellular responses to stress are frequently observed in cultivated polyploids, they have led to the hypothesis that polyploidization generally aligns with an increased resilience to stress[11,13].

An enhanced stress tolerance or adaptive flexibility is hypothesized to transiently elevate polyploid establishment rates beyond expected norms during periods of environmental change. While classical birth–death models have successfully linked speciation rates to environmental variables like temperature[21,22], validating similar hypotheses for polyploid establishment rates remains challenging because the environmental drivers and responses are highly diverse, likely often interacting. In 2009, Fawcett and co-authors[23] estimated the absolute ages of nine ancient WGD events in angiosperms and found that these events appeared to roughly coincide with the K–Pg boundary, based on which they hypothesized that

90 polyploidization may have helped plant lineages survive the K-Pg mass extinction event. Some of the same
91 authors[24] expanded the list of nine WGDs with another 11 and confirmed a clustering around the K–Pg
92 boundary. These two studies, although invigorating a range of hypotheses about the evolutionary impact
93 of polyploidy[13,25-27], were not only constrained by limited phylogenetic sampling, but also remained
94 contentious in light of recent large-scale WGD studies, where a clustering of WGDs in time, or a link with
95 environmental disturbance, was not observed or reported[26-28]. Moreover, the inference and dating of ancient
96 WGDs is notoriously difficult[29] and numerous previously identified ancient WGD events have been shown
97 to be either false positives[30,31], or wrongly dated, or both[32,33].

98 Here, we conducted an extensive analysis of a dataset encompassing 470 angiosperm genomes to identify
99 and date 132 independent ancient polyploidization events using the Bayesian relaxed molecular clock
100 model and 44 fossil calibrations for angiosperms that were highly curated following best practices[34]. To
101 investigate patterns of paleopolyploidization across lineages in conjunction with paleoclimate dynamics and
102 periods of mass extinction, we first meticulously assembled a robust phylogenetic framework and
103 evolutionary timeline for the angiosperms. Our findings show that paleopolyploidization occurs across each
104 major clade of angiosperms, including the ANA grade, Ceratophyllales, Chloranthales, eudicots, monocots,
105 and magnoliids, and spans an evolutionary timeline from the Lower Cretaceous to the Neogene. Although
106 the frequency of paleopolyploidization across different lineages and time varies dramatically, we find that
107 ancient WGD establishment is correlated with periods of environmental turmoil or mass extinction, and
108 more generally with periods of reduced lineage diversity and associated reduced competition. To facilitate
109 data reuse, we built a web application (available at https://bioinformatics.psb.ugent.be/AngioWGD/,
110 Supplementary Note 1) where users can explore WGDs across the angiosperm phylogeny, download
111 inferred dates for each WGD, and access original genomic data included in this work.

112

113 ## RESULTS

114 ### *Genome assessment based on BUSCO and OMArk*

115 Because the quality of genome assemblies is essential for reliable downstream comparative analyses, we
116 assessed all genome assemblies, i.e., 470 angiosperm genomes and the genome of the outgroup species
117 *Cycas panzhihuaensis*. Complete BUSCO[35] scores are generally high (median, Q1, and Q3 = 92%,
118 85.15%, and 96.45%, respectively), and so are OMArk[36] scores (median, Q1, and Q3 = 95.45%, 90.41%,
119 and 96.52%, respectively) (Figures S1 and S2 in Supplementary Note 2). The percentages of missing
120 BUSCOs are low (mean and median = 7.93% and 4.10%, respectively), and so are percentages of OMArk
121 missing families (mean and median = 7.85% and 5.55%, respectively). Taken together, these statistics
122 indicate the high quality of our datasets.

123 ### *Reconstruction of an evolutionary timescale of angiosperms*

124 Using 32 gene tree datasets, each with 1,614 reference single-copy gene families obtained from the
125 *embryophyta_odb10* dataset[37], we inferred species trees and assessed their overall accuracy and

consistency, and ultimately reconstructed a 'consensus tree' (see Method details). The resulting consensus tree (hereafter referred to as 'main tree'), covering 45 orders, 131 families, and all major clades of Mesangiospermae and ANA grade, displays high dataset- and branch-wise concordance (Figure S2 and S3-7 in Supplementary Note 3). High congruence was achieved across alternative Multiple Sequence Alignment (MSA) datasets, while two discordances pertaining to Chloranthales and Crossosomatales were further tested, concluding that the phylogenetic relationships that Chloranthales is sister to magnoliids, and Crossosomatales is sister to Fabales and Myrtales, were most supported (Figure S3 and S5 in Supplementary Note 3). We also show that our phylogenetic inference is robust to random sampling bias of gene trees (Figure S7 in Supplementary Note 3).

In comparison to a recent large-scale angiosperm phylogeny based on 353 nuclear orthologues[38], our main tree shows concordance in most of the ordinal-level relationships (Figure S8 in Supplementary Note 3), and in particular the results of our statistical topology tests on Chloranthales and Crossosomatales are supported. However, the phylogenetic positions of magnoliids, Ceratophyllales, and Caryophyllales remain uncertain. Although the position of magnoliids has long been disputed, several recent large-scale phylogenetic studies[26,39,40] agree with our main tree, supporting monocots being sister to magnoliids and eudicots. Phylogenetic inference from 30 out of 32 datasets supports our main tree, while the only two datasets that did not (the codon-level and TAPER-masked[41] datasets at the branch filtering cut-off of 70%) also fail to recover monophyly of several orders and denoted numerous branches as highly uncertain (Supplementary Note 3), probably due to an insufficient number of remaining branches after stringent filtering. The phylogenetic location of Ceratophyllales in Zuntini et al. (2024)[38] is highly uncertain (local posterior probability (LPP) below 0.5), and it was claimed by the authors themselves as inconclusive, while our main tree fully supports Ceratophyllales as sister to the eudicots (LPP = 1). Our tree also supports Caryophyllales as sister to the Lamiids and Asterids (LPP = 0.83) in 31 out of 32 datasets (except for the codon-level dataset at the branch filtering cut-off of 70%), in line with previous studies[26,39,40,42], but disagreeing with Zuntini and co-authors[38], who showed high gene tree conflict at the node pertaining to Caryophyllales.

Using the topology of our main tree, we reconstructed a time-calibrated phylogenetic tree based on our angiosperm fossil calibration dataset (Supplementary Note 4). Given that only 1 Single-Copy Orthogroup (SOG, see Method details) exist including all 470 angiosperm species, we slightly relaxed the species coverage to 99%, resulting in 12 SOGs encompassing 466 species. The uncorrelated lognormal (UCLN) relaxed molecular clock model was applied with soft fossil calibration bounds to consider both rate uncertainty and time uncertainty (Supplementary Note 4), resulting in an evolutionary time tree of 466 angiosperm species (Figure 1). The stem order and family ages in our time tree (Table S3.2 and 3.3) are highly correlated with those obtained by Ramírez-Barahona and co-authors[43] (Pearson correlation coefficients = 0.8157 and 0.7542; $P < 0.0001$; Figure S9 in Supplementary Note 5), suggesting comparable divergence time estimates.

Dating WGDs reveals temporal and phylogenetic dynamics of paleopolyploidization

We identified 132 independent ancient WGD events including both previously reported and novel ones, and conducted standardized WGD dating analysis on each one of them (Figure 1; Supplementary Note 6). Eudicots display the largest number of ancient WGD events (*N* = 95), followed by monocots (*N* = *25*), magnoliids (*N* = 5), ANA grade (*N* = 4), Ceratophyllales (*N* = 2), and Chloranthales (*N* = 1). WGD frequency among different orders and families varies greatly. For instance, Lamiales harbors 10 independent ancient WGD events, which is greater than the 95th percentile of all orders with ancient WGDs (Figure S1b in Document S1; Figure 2). Likewise, the plant families Poaceae, Fabaceae, and Brassicaceae contain 8, 5, and 5 independent ancient WGD events, respectively, which is greater than the 95th percentile of all families with ancient WGDs (Figure S1c in Document S1; Figure 2).

The most ancient WGD event dated in monocots is the well-known τ event[44], shared by all non-Alismatales monocots (excluding Acorales) and dated at 121.31 Ma (90% HCR = 112.87-129.35 Ma) (Figure S1d in Document S1; Supplementary Note 6). In eudicots, the oldest dated WGD events are the 'RANU' event (WGD nomenclature in Supplementary Note 6), shared by Ranunculales (Figure S1e in Document S1) and dated at 113.92 Ma (90% HCR = 108.24-122.32 Ma), and the γ event[45], shared by core eudicots (Figure S1d in Document S1) and dated at 107.58 Ma (90% HCR = 94.96-119.90 Ma). In magnoliids, the oldest WGD event is the 'LAUR β' event, shared by Laurales and Magnoliales (Figure S1d in Document S1) and dated at 109.98 Ma (90% HCR = 98.56-120.51 Ma). There are 10 independent ancient WGD events shared by entire plant orders, and 10 ancient WGD events shared by 'partial' orders (i.e., shared by multiple plant families, but not the entire order; Figure S1e and S1f in Document S1). We also identified 31 independent ancient WGD events shared by entire plant families (Figure S1g in Document S1), and 20 ancient WGD events shared by 'partial' families (i.e., shared by multiple plant genera, but not the entire family; Figure S1h in Document S1). The remaining 58 independent paleopolyploidy events are shared by entire genera or unique to specific species (Figure S1i in Document S1).

The WGD establishment rate is negatively correlated with species richness across time

Linear regression analysis revealed a strong linearity between the ancient WGD establishment rate (i.e., the number of ancient WGDs observed per lineage and per time unit) and the number of lineages across time (R^2 varying from 0.2734 to 0.8217 when using different time window sizes, P<0.0001; Figure S3 in Document S1), and a linear model fit better than a power-law model (Figure S4 in Document S1). As the number of lineages increases exponentially in time (R^2=0.9915, P<0.0001; Figure S5 in Document S1), the negative linear dependence of the WGD establishment rate on species richness translates to an exponential relationship between the WGD establishment rate and time.

Significant deviations from the exponential relationship between WGD establishment rate and time unveil WGD clustering in time

When modeling the WGD establishment rate as a linear function of species richness and time and fitting a smoothing spline to the resulting relative residuals, 10 WGD peaks are recovered in the past 115 Ma (Figure

198 3A). An alternative approach using exponential-Gaussian mixture modeling of the WGD establishment rate
199 profile with a varying number of Gaussians recovers all these peaks except the most recent one (around 8
200 Ma) in the optimal mixture model (adjusted R^2=0.9987, Figure 3D, see Table S4). In addition, comparison
201 of the observed peaks with a null distribution of the WGD establishment rate across time, as obtained from
202 a stochastic WGD model assuming constant establishment rates in major clades, shows that 6 of the peaks
203 rise above the 95% confidence interval (CI) of the null distribution (Figure 3G), namely those located at 35,
204 54, 64, 73, 85 and 99 Ma. These results suggest that at least 6 of the peaks recovered in the relative
205 residuals profile may be evolutionarily significant. A null model including negative lineage diversity
206 dependence of the WGD establishment rate showed two peaks at 73 Ma and 54 Ma beyond the 95% CI
207 (Figure S6 in Document S1). The better model fit confirms again that a lineage diversity-dependent WGD
208 establishment rate model is a better fit to the data than a constant rate model, supporting the hypothesis
209 that the background WGD establishment rate is lineage diversity-dependent. For the two peaks that reach
210 the upper 95% CI threshold, the lineage diversity-dependent null model provides solid evidence that other
211 factors than random lineage diversity-dependent WGD establishment contributed to their formation. For the
212 four peaks at 35, 64, 85, and 99 Ma that exceed the upper 50% CI threshold, this cannot be assessed with
213 much confidence from this model, but one must consider that the null model WGD establishment rate
214 parameters were fitted to all the data, including the peaks. If the observed peaks are non-random and
215 superimposed on a random lineage diversity-dependent background, the background WGD establishment
216 rates in the null model and the associated upper CI thresholds would have been overestimated.

217 ***Significant coupling between ancient WGD establishment rate peaks and environmental upheaval***

218 To assess the evolutionary significance of the nine peaks in WGD establishment rate (further referred to as
219 WGD peaks) recovered in the linear and exponential-Gaussian mixture models, we compared the timing of
220 these peaks to the timing of geological events, mass extinction events, and significant maxima or minima
221 in time profiles of paleoclimatic variables (Figure 3). Monte Carlo randomization analyses of WGD peak
222 timing uncovered that the observed WGD peaks in the past 115 Ma are located significantly closer to known
223 extinction events and oceanic anoxic events (OAEs) than expected by chance (P=1.11e-03, Figure 3A).
224 The observed peaks are also significantly linked to genus-level angiosperm extinction rate peaks inferred
225 from the angiosperm fossil record, as documented in the Paleobiology database (P=9.20e-03, Figure 3D).
226 We also assessed links between the observed WGD peaks and minima/maxima in smoothed time profiles
227 of benthic foraminiferal $\delta^{13}C$ and $\delta^{18}O$ isotopic signatures over the past 115 Ma (Figure 3E, F), and time
228 profiles of inferred sea surface temperature (Figure 3G), global mean surface temperature (Figure 3H) and
229 atmospheric CO_2 concentration (Figure 3I) over the past 115 Ma. Given that the paleoclimatic profiles
230 contain large numbers of minima and maxima, we focused on the top-18 minima/maxima for each profile
231 in terms of peak importance value (i.e., twice the number of WGD peaks) and compared randomized WGD
232 peaks to the top-1, top-2 … top-18 minima/maxima for each (see Table S5.1). The minimum P-values
233 obtained for each of the data layers and the corresponding top minima/maxima are displayed in Figure 3
234 and indicate significant coupling of the observed WGD peaks to important minima/maxima for all

235 paleoclimatic variables at P<0.01, except global mean surface temperature (P=0.116) and CO_2
236 concentration (P=0.0554). Because uniform randomization of peaks across the last 115 Ma may not
237 produce sufficiently realistic inter-peak distance profiles (e.g., 9 random peaks may be located at 1,2,...,9
238 Ma whereas this is highly unlikely in smoothed WGD establishment rate profiles), we performed alternative
239 randomization analyses where inter-peak distances were sampled from a Gaussian distribution (μ=11.57,
240 σ=3.37 Ma) fit to the observed distances between the consecutive WGD peaks in Figure 3A. These
241 randomization analyses are highly restrictive, as they essentially assume that peaks in establishment rate
242 profiles are always located ~10 Ma apart on average, whereas the observed regular WGD peak spacing
243 they mimic may be caused by geological or biological rather than technical effects. It has been conjectured
244 for instance that mass extinctions, associated geological events such as continental flood basalt eruptions
245 and OAEs, and paleoclimatic events such as glaciations may happen at regular intervals, linked to
246 astronomical cycles such as the Milankovitch Earth orbital eccentricity cycle and Galactic cycles[46].
247 Nevertheless, the restrictive randomization analyses still result in significant links (P<0.05) between the
248 observed WGD peaks and geological events (P=6.62e-03), and minima/maxima in sea surface temperature
249 (P=1.20e-02) and $\delta^{13}C$ isotopic signatures (P=3.49e-02), whereas links between observed WGD peaks and
250 angiosperm extinction rate peaks or minima/maxima in $\delta^{18}O$ and CO_2 are still significant at P<0.1 (Figure
251 S7 in Document S1, full results in Table S5.2).

252

253 ## DISCUSSION

254 ### *Waves of paleopolyploidizations during angiosperm evolution*

255 Analyzing an extensive dataset containing 470 complete angiosperm genomes, we found evidence for 132
256 ancient polyploidization events. This seems a lot but if we consider individual evolutionary lineages, most
257 plant lineages seem to contain remnants of only one, sometimes two, and in rare cases, three, four, or five
258 WGD events (Figure 2). The contrast between numerous instances of recurring polyploidy and recently
259 originated polyploids, and the few cases - within the same evolutionary lineage - of established ancient
260 polyploidies over the past 150 million years (Figures 1 and 2), is striking[8,9]. This observation can only be
261 understood when the high rate of polyploid emergence is countered by a reduced establishment rate at
262 longer evolutionary time scales. Furthermore, one can wonder whether the sparsely distributed WGDs that
263 survived in the long run got established at random times, or whether these WGDs survived because they
264 occurred at decisive moments in evolution. As we show here, different model-based analyses support
265 'waves' of WGD establishment around 14, 24, 35, 54, 64, 73, 85, 99, and 108 Ma (Figure 3), many of which
266 are associated with periods of mass extinction and major climate change.

267 ### *Ancient polyploidy correlates with periods of environmental upheaval*

268 The first three WGD establishment waves (from more ancient to more recent) at ~108, 99, and 85 Ma
269 approximately coincide with the oceanic anoxic events OAE1c (~106 Ma), OAE1d (~101 Ma) and OAE3
270 (~87 Ma), respectively[47]. Oceanic anoxic events are significant periods in Earth's history when large

portions of the ocean became depleted of oxygen, leading to widespread marine dysoxia or anoxia[48]. These events are often associated with climate shifts (thermal maxima), geological events such as increased large igneous province (LIP) volcanism and extinction events. OAE3 was the last and one of the longest of such events in the Cretaceous, associated with high temperatures and atmospheric CO_2 levels (Figure 3G-I), but not with an extinction peak for angiosperm genera (Figure 3D). The comparatively small WGD wave at ~108 Ma and the linked OAE1c event are associated with lower temperatures and high atmospheric CO_2 levels (Figure 3G-I), but again not with a peak in angiosperm extinctions (Figure 3D). The WGD wave at ~99 Ma and the close by OAE1d event on the other hand are associated with a major angiosperm extinction rate peak and substantial excursions in $\delta^{13}C$ and $\delta^{18}O$ isotopic signatures (Figure 3D-F). Interestingly, no WGD wave seems to be linked to OAE2, the most significant oceanic anoxic event in the Cretaceous associated with a mass extinction event at the Cenomanian–Turonian boundary (~93.9 Ma)[49]. It is tempting to speculate that the observed WGD peak at ~99 Ma and the angiosperm extinction rate peak at ~97.2 Ma (Figure 3D) may be linked to OAE2 rather than OAE1d. Alternatively, the WGD peak around OAE1d may have reduced the need for additional angiosperm polyploidizations at the Cenomanian-Turonian boundary.

The fourth, and most prominent, wave of ancient WGDs occurred at ~73 Ma. Interestingly, this wave is not located close to a major established extinction event or an OAE (Figure 3A), but it is located close to a peak in the genus-level angiosperm extinction rate at ~78 Ma (Figure 3D, note that the temporal resolution of the angiosperm extinction rate profile and the associated extinction peaks is low). At the end of the Cretaceous (~80–66 Ma), a period with cooling temperatures[50] (Figure 3G,H) and decreasing CO_2 levels (Figure 3I) after the preceding Cretaceous 'hothouse' period, overall plant diversity decreased substantially and the rise of the angiosperms temporarily stalled[51]. In Antarctica, angiosperm species diversity exhibited a pronounced dip around the Campanian-Maastrichtian boundary at ~72.1 Ma[51], very close to the inferred WGD wave's location at ~73 Ma. The Campanian–Maastrichtian Boundary Event (CMBE) is characterized by substantial negative and positive excursions in the $\delta^{13}C$ and $\delta^{18}O$ isotopic signatures, respectively (Figure 3E,F). What caused these excursions is still debated, with two of the competing hypotheses being the buildup of temporary ice sheets in Antarctica associated with a drop in sea levels[52], or a major change in intermediate to deep water circulation caused by the global cooling trend[53]. Despite not being associated with a major extinction event, the presence of a pronounced WGD peak around 73 Ma suggests that the period around the Campanian–Maastrichtian boundary may have been more important for angiosperm evolution than previously realized. Alternatively, the long-term evolutionary success of some of the polyploidizations that happened ~73 Ma may have been catalyzed by the increased potential of the affected lineages to weather the subsequent climatological events in this turbulent geological time period, such as the Middle Maastrichtian Event (~69.5 Ma), the Late Maastrichtian Warming Event (~66.3 Ma) associated with increased volcanic activity of the Deccan Traps[54], and ultimately the K–Pg extinction event (~66 Ma). This hypothesis of increased adaptive potential of species due to earlier polyploidizations may also help explain why the ancient WGD peak and angiosperm genus extinction rate peak at the K–Pg boundary itself (see below) is smaller in size than might be expected for a major mass extinction event.

308 The fifth WGD establishment wave, at ~64 Ma, occurred close to the K–Pg boundary. The K–Pg mass
309 extinction, ~66 Ma, is the most recent of the 'Big Five' mass extinctions, best known for the extinction of
310 the non-avian dinosaurs[55,56]. Previously, Fawcett and co-authors[23] and Vanneste and coauthors[24] already
311 suggested a potential clustering of WGDs around the K–Pg boundary. Here, using a genomic data set
312 twelve times larger (N = 470, with 132 ancient WGDs), we still find support for this observation, although
313 the 64 Ma WGD peak recovered in Figure 3 A,B is substantially smaller than the peaks before and after it
314 (73 and 55 Ma). The angiosperm extinction rate exhibits a peak at 66 Ma (Figure 3D), but it is the smallest
315 of the five peaks recovered from the divDyn analyses (see Methods), suggesting that the K–Pg extinction
316 event had a comparatively minor effect on the survival of angiosperm genera. At least one phylogenetic
317 study points in the same direction[57]. Given the large uncertainties in ancient WGD dating, Fawcett and co-
318 authors[23] and Vanneste and co-authors[24] assumed that all WGDs that were dated between 55 and 75 Ma
319 were likely associated with the K–Pg boundary. Now, using a dataset with many more species and ancient
320 WGDs, and hence enhanced resolution, the evidence suggests that the previously observed clustering of
321 ancient WGDs at 55–75 Ma is in fact associated with three separate geological events: the Paleocene–
322 Eocene Thermal Maximum (PETM, see below), the K–Pg boundary and the CMBE) of which K–Pg may
323 have been the least consequential in terms of angiosperm genus extinction and ancient WGD
324 establishment.

325 The sixth wave of WGDs at ~54 Ma is located near the Paleocene–Eocene Thermal Maximum (PETM
326 event), ~55.8 Ma, a period characterized by a 5 to 8°C increase in global mean surface and sea surface
327 temperatures (Figure 3G-H)[58-60], and a concomitant significant negative excursion in $\delta^{18}O$ in lake bed and
328 ocean sediments (Figure 3F). The PETM event is considered one of the most extreme climate changes in
329 Earth's history, and was characterized by a massive release of $\delta^{13}C$-depleted carbon to the ocean–
330 atmosphere system[61-63], causing a prominent negative excursion in $\delta^{13}C$ in e.g. benthic foraminifera from
331 ocean sediment cores (Figure 3E). The causes of this massive carbon release are still heavily debated, but
332 may include volcanic intrusions in carbon-rich sediments followed by massive thermogenic $\delta^{13}C$-depleted
333 methane release[60,64] or increased volcanism causing an initial phase of global warming, followed by
334 subsequent warming-induced release of $\delta^{13}C$-depleted methane from submarine clathrates on continental
335 margins[65]. The PETM had significant impacts on biodiversity and many species of both terrestrial and
336 marine organisms became extinct, including angiosperms (Figure 3D), or experienced substantial shifts in
337 their distribution. The PETM saw the flourishing of subtropical forests, with many lineages adapting to the
338 warmer conditions[65,66]. Previously, Cai and coauthors[67] also reported a clustering of ancient WGD events
339 around the Paleocene–Eocene transition, based on the dating of 22 independent ancient WGD events from
340 genomic and transcriptomic assemblies of 36 Malpighiales species.

341 The seventh WGD establishment wave at ~35 Ma is located very close to the Eocene–Oligocene transition
342 (EOT), associated with the Eocene–Oligocene extinction event or 'Grande Coupure', about 33.5 Ma, which
343 was characterized by a rapid shift to a colder climate that led to the development of a permanent Antarctic
344 ice sheet and significant environmental changes. The transition from the warm conditions of the Eocene to

the cooler, more variable climates of the Oligocene is one of the most important climate shifts in Earth's history. This shift is associated with a significant drop in global temperatures, because of both tectonic activities and changes in oceanic circulation patterns[68,69]. Indeed, the EOT and associated WGD wave correspond to a peak in $\delta^{18}O$ indicative of a cooling event and increasing ice volume[70] (Figure 3F). That the EOT and the associated WGD wave seem to correspond to a sea surface temperature peak (Figure 3G) rather than the transition to colder temperatures (between 32 and 30 Ma in Figure 3G) may be due to dating uncertainties. The EOT is also associated with a mass extinction event, mainly among marine species[71], but also greatly affecting flowering plants (see Figure 3D). Indeed, on land, the cooler temperatures and changing habitats led to shifts in vegetation and the extinction of plants and animals adapted to the warmer, more humid conditions of the Eocene[72]. Overall, the EOT is considered a critical period in Earth's history, marking significant changes in climate and ecology.

The eighth WGD establishment wave around 24 Ma does not coincide with a documented major extinction event or OAE (Figure 3A) or a peak in the angiosperm extinction rate (Figure 3B), but approximately coincides with the Oligocene-Miocene Transition (OMT), a transient global cooling event ~ 23 Ma that is associated with significant Antarctic ice sheet expansion. However, the $\delta^{13}C$ (Figure 3E) and sea surface temperature (Figure 3C) profiles exhibit a maximum around this time, and the $\delta^{18}O$ a minimum (Figure 3H), which may reflect a warming event in the late Oligocene rather than the OMT[73]. The discrepancies between the estimated OMT age and the observed excursions in $\delta^{13}C$, $\delta^{18}O$ and sea surface temperature profiles may be due to dating uncertainties. As the 24 Ma WGD peak is the least prominent of the WGD establishment rate peaks supported by mixture modeling (Figure 3B) and is not supported by our null model simulations (Figure 3C, Figure S6) and not strongly connected to a known mass extinction event or major climate transition, this is the WGD peak we are least confident about. Nevertheless, its timing fits nicely in a continuous series of 6 WGD peaks associated with successive geological boundaries, including, in succession, the WGD peak at ~85 Ma close to the Coniacian-Santonian and Santonian-Campanian age boundaries in the Upper Cretaceous (~86.3 and ~83.6 Ma, respectively), the WGD peak at ~73 Ma close to the Campanian-Maastrichtian age boundary at ~72.1 Ma, the ~64 Ma peak close to the Cretaceous-Paleogene (K-Pg) period boundary (~66 Ma), the ~54 Ma and ~35 Ma WGD peaks close to the Paleocene-Eocene and Eocene-Oligocene epoch boundaries in the Paleogene (~56 Ma and ~33.9 Ma, respectively), and the ~24 Ma peak close to the Paleogene-Neogene period boundary (which coincides with the OMT, ~23 Ma). Of the three remaining WGD peaks, only one is close to a geological boundary, namely the WGD peak at ~99 Ma close to the period boundary between the Lower and Upper Cretaceous (~100.5 Ma).

The ninth and most recent wave of WGD establishment at ~14 Ma coincides with the Middle Miocene Climatic Transition (MMCT), also known as the Middle Miocene disruption. The MMCT occurred approximately 13 to 15 Ma and marks a significant shift from a warmer climate to a cooler one, leading to substantial environmental changes greatly affecting ecosystems and biodiversity[74,75]. As the climate cooled, forests began to decline in many regions, giving way to grasslands and open savannas. This shift led to adaptations in flora and the evolution of new faunal communities, including grazers that thrived in the new

grassland habitats[76]. The MMCT is sometimes considered a significant extinction event[77,78], while others see it rather as a transformative period that significantly affected species distributions and set the groundwork for the ecosystems known in the later stages of the Miocene and into the Pleistocene[79,80]. The MMCT is not associated with a peak in the genus-level angiosperm extinction rate (Figure 3B), but does correspond with a drop in sea surface temperature (Figure 3C), a sharp drop in $\delta^{13}C$ (Figure 3E), indicative of a decrease in vegetation abundance, and a sharp rise in $\delta^{18}O$ (Figure 3H), indicative of a decrease in seawater temperature and/or growing ice sheets[81].

Revisiting the polyploid paradox

Nature is teeming with polyploid species that have extra sets of chromosomes. Yet, ancient WGDs that survive over the long term are surprisingly scarce (Figure 2). The apparent paucity of (established) ancient genome duplications and the existence of so many species that are currently polyploid provides a fascinating but puzzling paradox[8,82]. The rarity of polyploidy events that get 'fixed' on the long run seems to suggest that polyploidy is usually disadvantageous, and an evolutionary dead end. However, there is growing evidence that polyploidization can sometimes help overcome environmental change (stress)[13,83,84]. At decisive moments in evolution, as again shown here, polyploids seem to have had an adaptive advantage and likely outcompeted many of their diploid progenitors. Despite much research, the mechanistic underpinnings of why and how polyploids might be able to outcompete or outlive non-polyploids remain elusive.

Here, we demonstrate for the first time that WGD establishment is enhanced under conditions of low lineage diversity, both early on in angiosperm evolution and during periods of major environmental upheaval and mass extinction. The negative lineage diversity dependence we observe for the background WGD establishment rate is very similar to the well-known negative lineage diversity dependence of speciation rates. Indeed, empirical evidence suggests that diversification tends to decelerate as evolution progresses[21,85-87]. This slowdown is traditionally attributed to diversity-dependent mechanisms, where increasing species richness intensifies competition for resources and space, leading to niche saturation[88,89]. Recent studies highlight the role of environmental perturbations such as fluctuations in sea level, $\delta^{13}C$, CO_2 concentrations, and temperature, as additional drivers of diversification dynamics[22,90,91]. Very similar dynamics are observed here for WGD establishment rates, suggesting that, as for speciation, increased diversity reduces the likelihood of successful WGD establishment due to competitive exclusion. Conversely, in periods of reduced diversity and competition, e.g., when major environmental changes cause previously well-adapted species to go extinct and new niches to open up, WGD establishment rates may temporarily increase. In first instance, environmental stress may lead to an increase in polyploid formation due to increased unreduced gamete formation[92]. These polyploids may then be better able to adapt to the changed environment than their diploid progenitors, either due to immediate adaptive advantages obtained through polyploidization[93] or heterosis, or through higher adaptability on the longer term[6,94]. Examples of the first are duplicates of RNA-binding proteins involved in cold responses that exhibit convergent retention across diverse angiosperm lineages following independent WGD events occurring around the K–Pg boundary[95],

419 or gene duplicate advantages in low temperature and darkness around the K–Pg boundary, when global
420 cooling and darkness were the two main stresses[96]. To take advantage of potential higher long-term
421 adaptability however, the polyploid lineage must be able to survive long enough to allow for any initial fitness
422 disadvantages to be overcome[11,12] and for adaptive advantages to arise, e.g. through genome fractionation
423 and diploidization[97] and duplicate gene divergence. Here, reduced diversity and reduced competition may
424 be crucial in facilitating newly emerged polyploids to persist long enough to gain the upper hand[10,94,98]. In
425 some cases, lag times of several million years have been identified between ancient WGD events and the
426 subsequent diversification of lineages thought to have been enabled by beneficial traits obtained from those
427 WGDs[99-101]. Such long lag times suggest that it may often take a lot of time for polyploids to evolve adaptive
428 traits, which may be challenging in stable environments with high competition. Alternatively, an established
429 (paleo)polyploid lineage may have had to wait for millions of years to diversify until the ecological
430 opportunity presented itself in the form of severe environmental change. Both scenarios are not mutually
431 exclusive. In the aftermath of environmental turmoil, as niches become filled due to radiation and
432 competition increases again, the window of opportunity for polyploids to get established on the long term is
433 expected to close again. In summary, our data indicates that WGDs are a dead-end under relatively stable
434 conditions (hence their scarcity on the evolutionary tree) but could confer a selective advantage in times of
435 severe environmental change, providing a potential explanation for the polyploid paradox.

436

437 ## CONCLUSIONS

438 The advent of genomics made it clear that the significance of polyploidy extends across all eukaryotes, and
439 that most, if not all, extant species (including our own) carry the signature of at least one ancient whole
440 genome duplication (WGD) in their history[1,8,102-104]. Nevertheless, finding evidence for ancient WGD can be
441 complicated[105] and dating these events, going back tens to hundreds of millions of years ago, is notoriously
442 difficult. Here, we assembled and assessed a large angiosperm genome dataset comprising 470 different
443 angiosperm species and used state-of-the-art methods to identify and date 132 independent, ancient
444 WGDs. Analyzing such a large number of species and WGDs is necessary to be able to draw reliable
445 conclusions about potential correlations between WGDs and 'decisive' moments in evolution, which was a
446 limitation of some previous studies based on much smaller numbers of species and WGDs[23,24].
447 Furthermore, our large dataset allows for a resolution that was impossible to achieve previously. Model-
448 based analyses and simulations clearly reveal that the inferred WGD events are not randomly distributed
449 but rather cluster around periods of environmental upheaval. We find significant clustering of WGDs around
450 the Middle Miocene Disruption, the Eocene–Oligocene Extinction Event, the Paleocene–Eocene Thermal
451 Maximum, the K–Pg boundary, and several oceanic anoxic events. Our study again provides support for,
452 what we call "the polyploid paradox": neopolyploids abound, while the long-term survival of polyploidy
453 seems to be rare (Figure 2). We hypothesize that, in general, and under relatively stable environments, the
454 disadvantages and the costs of being polyploid do not outweigh the sometimes-observed advantages such

as increased stress resilience[12,13]. However, when the circumstances are such that the conditions to survive are extremely challenging, polyploids could be the 'hopeful monsters'[93,106-109] that can outcompete their diploid progenitors. In this respect, we feel that polyploidy can indeed often be considered an evolutionary dead end, at least when interpreted as 'rarely successful'[110-112], but that it can be an evolutionary and ecological force in (very) unstable times. In this respect it is interesting to speculate that, in the current Anthropocene, a time during which humans are having a substantial impact on our planet, several new 'ancient' polyploidy events might be in the making.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Yves Van de Peer (yves.vandepeer@psb.ugent.be).

Materials availability

All results generated in this study are available from the lead contact upon request.

Data and code availability

- The genomic data involved in this study have been deposited at figshare at https://doi.org/10.6084/m9.figshare.27011128.v1 and https://doi.org/10.6084/m9.figshare.27011134.v1.
- The posterior samples of WGD dates for each independent WGD event have been deposited at figshare at https://doi.org/10.6084/m9.figshare.27596856.v1 and the associated mixed K_S distribution figures have been deposited at figshare at https://doi.org/10.6084/m9.figshare.29553929.v1.
- The code for the genome assessment, phylogenetic inference, molecular dating, WGD identification and dating, Poisson process models of WGD simulation, leave-one-out cross-validation (loo-cv) of KDE bandwidth, residual analysis, mixture modeling analysis, coupling analysis, and the extinction rate estimation is available at https://github.com/heche-psb/Angiosperm_WGDs.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGEMENTS

H.C. acknowledges funding from the Research Foundation—Flanders (FWO) (No. 3G032219). Y.V.d.P. acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (No. 833522). Y.V.d.P. and F.A-S. acknowledge funding from Ghent University (Methusalem funding, BOF.MET.2021.0005.01).

AUTHOR CONTRIBUTIONS

Y.V.d.P. and H.C. conceived and managed the project. H.C. collected the genomic data and conducted the assessment analysis. H.C. conducted the phylogenetic and molecular dating analysis. H.C. conducted the WGD inference and dating analysis. H.C. conducted the WGD simulation analyses. G.L., D.B. and F.A.S. conducted the Bayesian phylogenetic clustering analysis in the first version of the manuscript. F.A.S. constructed the website. S.M. conducted the mixture modeling analysis of the WGD establishment rate profile and the coupling analyses between WGD establishment rate peaks and other data layers. D.B. conducted the diversity dependence analysis of the ancient WGD establishment rate. H.C. conducted the loo-cv analysis and the extinction rate estimation analysis. H.C., F.A.S., D.B., S.M., and Y.V.d.P. wrote the manuscript. All authors read and approved the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

SUPPLEMENTAL INFORMATION

Document S1. Figures S1–S8

Table S1. Tukey's HSD test of different datasets at different branch support filtering cut-offs, related to Figure 1

Table S2. Age constraints for the stem group of different angiosperm orders, related to Figure 1

Table S3.1. Functional descriptions of the 12 SOGs adopted in molecular dating, related to Figure 1

Table S3.2. Posterior mean, 95% HPD of stem age of angiosperm orders, related to Figure 1

Table S3.3. Posterior mean, 95% HPD of stem age of angiosperm families, related to Figure 1

Table S3.4–3.8. Tukey's HSD test of the relative nucleotide and amino acid substitution rate, related to Figure 1

Table S4. Adjusted R^2 for the Exponential-Gaussian mixture modeling, related to Figure 3

Table S5.1. P-values for the association test under the null hypothesis of random uniform peak distribution, related to Figure 3

Table S5.2. P-values for the association test under the null hypothesis of restricted distances between randomized WGD peaks, related to Figure 3

Table S6. K_S peak values, credible ranges and WGD dates for the absolute dating of each independent WGD, related to Figure 1

Table S7. WGD establishment rate parameter in the homogeneous Poisson process model, related to Figure 3

Supplementary Note 1. The *AngioWGD* web application, related to Figure 1

525 **Supplementary Note 2. Assessment of genome assembly, related to Figure 1**
526 **Supplementary Note 3. Phylogeny reconstruction, related to Figure 1**
527 **Supplementary Note 4. Fossil calibrations, related to Figure 1**
528 **Supplementary Note 5. Evolutionary timeline and rate variation, related to Figure 1**
529 **Supplementary Note 6. WGD identification and dating, related to Figure 1**
530 **Supplementary Information 1. Results of species tree inference, related to Figure 1**
531 **Supplementary Information 2. Fossil calibration map, related to Figure 1**
532

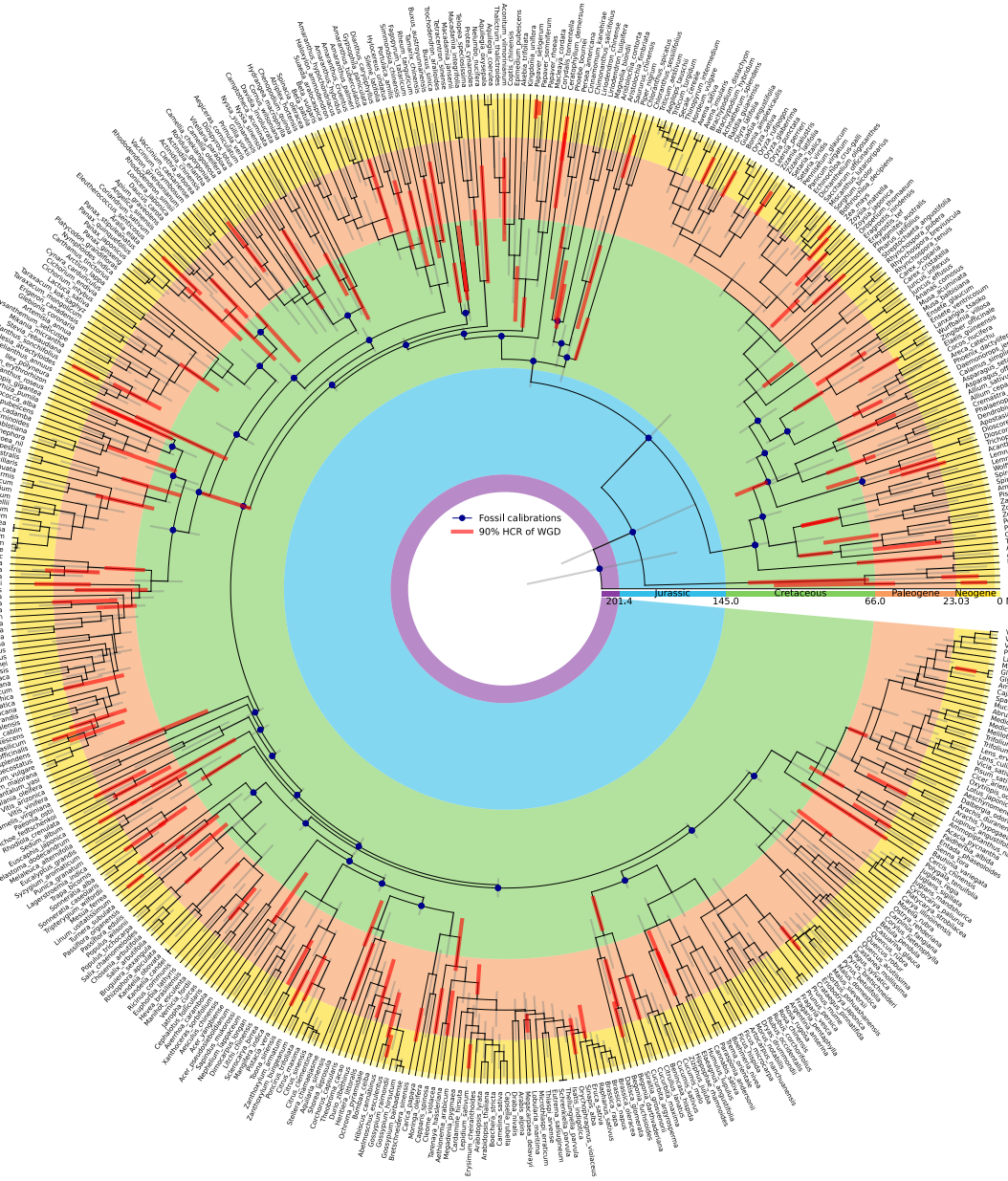

Figure 1. Reconstructed evolutionary time tree of 466 angiosperms. The 95% Highest Posterior Density (HPD) of divergence time estimates are represented as light gray lines on the internodes, while the 90% Highest Convergence Region (HCR) (see Methods) of WGD date estimates are represented as red rectangles. Applied fossil calibrations are represented as dark blue dots. Geological epochs are marked in distinct colors.

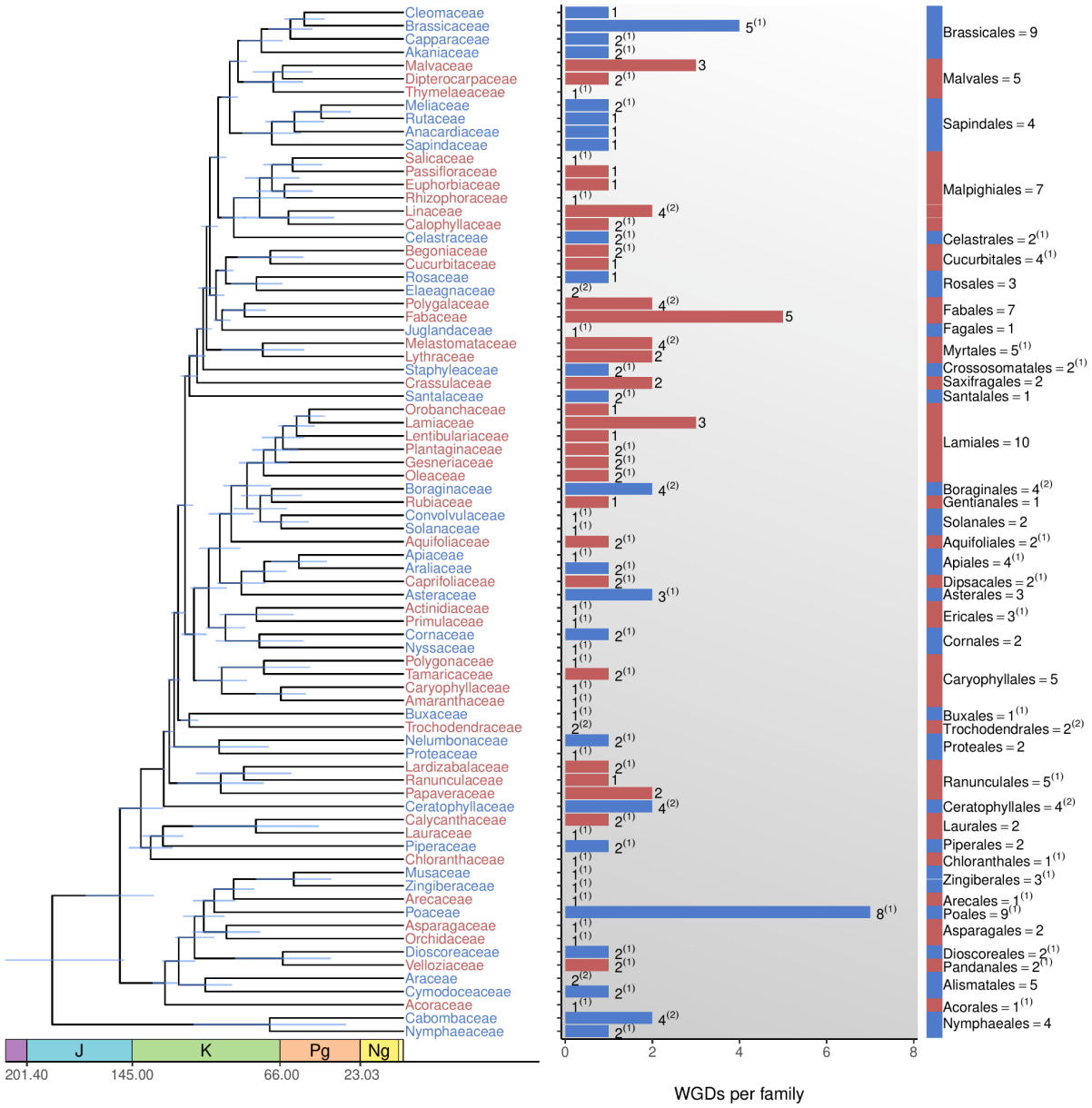

Figure 2. Distribution and number of WGDs in the plant families and orders considered in the current study. Numbers indicate WGDs observed in the clade. Numbers in superscript in parentheses indicate the number of WGDs at the base of each clade (MRCA), i.e., shared by all descendants. For instance, 'Myrtales = $5^{(1)}$ indicates five WGDs in Myrtales, one of which occurred in the most recent common ancestor of this order.

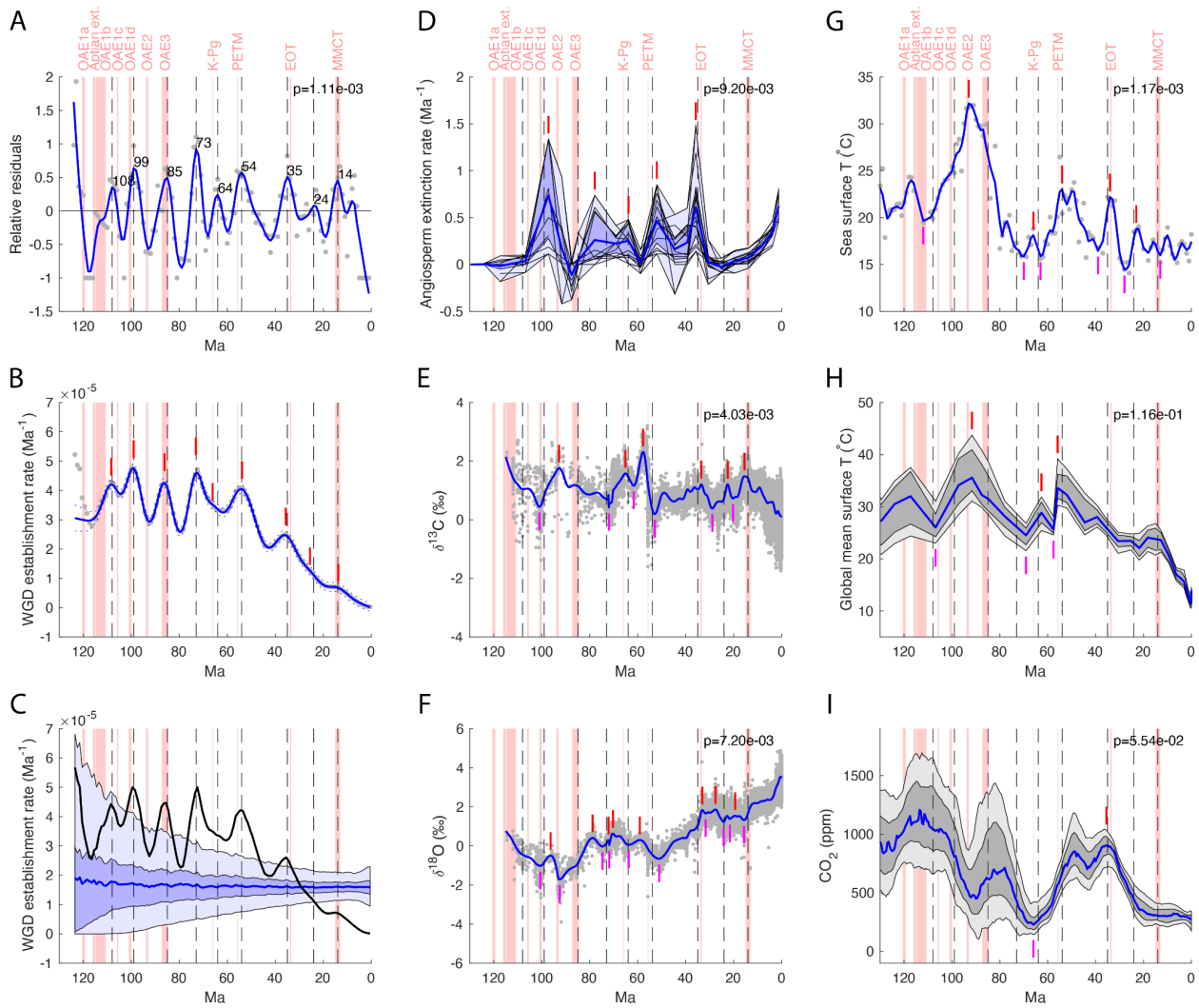

Figure 3. Association between WGD establishment rate peaks, geological events, extinction events and paleoclimatic extrema. Pink bars in all panels indicate significant extinction events or oceanic anoxic events (OAEs) in the past 130 Ma, as labeled in the top panels. Aptian ext.: Aptian extinction; K–Pg: Cretaceous–Paleogene extinction; PETM: Paleocene–Eocene Thermal Maximum; EOT: Eocene–Oligocene Transition; MMCT: Middle Miocene Climatic Transition. Vertical black dashed lines in all panels indicate WGD establishment rate peak locations in panel A. **A**. Gray dots represent the relative residuals of a linear model in which the moving average of the WGD establishment rate (window size 4 Ma) is modeled as a function of time and the number of angiosperm lineages in the 466-species tree through time (Figure 1). The blue curve is a LOESS fit with span set to 10% of the total number of data points. Peaks in the blue curve that are recovered in the optimal exponential-Gaussian mixture model (panel B) are indicated with vertical black dashed lines and the associated numbers indicate their ages in Ma. The extinction events and OAEs displayed in pink are located significantly closer to the WGD peaks than expected under a null model of uniform random WGD peak placement across the past 115 Ma, as indicated by the P-value in the upper right corner of the panel. **B**. Optimal exponential-Gaussian mixture model for WGD establishment

rate data. Gray dots are kernel density estimates (KDE) of the WGD establishment rate (KDE bandwidth = 3.1). The blue line is the optimal exponential-Gaussian mixture fit (adjusted R^2 = 0.9986) to the KDE data for the past 115 Ma, containing an exponential background and 9 Gaussian components of which the means are indicated by vertical red bars. **C**. Comparison of the observed KDE profile for WGD establishment rate (black line) with a null model assuming constant clade-specific WGD establishment rates (see Methods). The blue line represents the average WGD establishment rate profile obtained across 10,000 null model simulations, while the dark and light blue bands depict the 50% and 95% confidence intervals, respectively. Six of the nine peaks identified in panels A and B surpass the 95% confidence threshold. **D**. Genus-level angiosperm extinction rate profiles for the past 130 Ma. Black lines are extinction rate profiles obtained from the angiosperm fossil record in the Paleobiology database[113] using three methods of sampling standardization and four macroevolutionary rate models in divDyn (v0.8.3)[114]. The blue line is the average estimated angiosperm extinction rate across the 12 models used, with dark blue and light blue bands depicting the interquartile range and total range of the values. Five extinction rate peaks are indicated with vertical red bars. These angiosperm extinction rate peaks are located significantly closer to a WGD peak in panel A (dashed vertical black lines) than expected under a null model of uniform random WGD peak placement across the past 115 Ma, as indicated by the P-value in the upper right corner of the panel. **E**. Benthic foraminiferal $\delta^{13}C$ isotopic signature over the past 115 Ma, data from Friedrich et al. (2012)[115]. Gray dots are $\delta^{13}C$ measurements, the blue line is a smoothing spline with smoothing parameter 0.05. **F**. Benthic foraminiferal $\delta^{18}O$ isotopic signature over the past 115 Ma, data from Friedrich et al. (2012)[115]. Gray dots are $\delta^{18}O$ measurements, the blue line is a smoothing spline with smoothing parameter 0.05. **G**. Low-latitude sea surface temperature in the past 130 Ma, data from Song et al. (2019)[116]. Gray dots represent mean sea surface temperature estimates, and the blue curve is a LOESS fit with the span set to 2% of the total number of data points. **H**. Global mean surface temperature in the past 130 Ma, data from Judd et al. (2024)[117]. The blue line represents the median estimate while the dark and light gray bands represent the 68% and 90% confidence intervals, respectively. **I**. Atmospheric CO_2 concentration in the past 130 Ma, data from Foster et al. (2017)[118]. The dark blue solid curve represents the probability maximum while the dark and light gray bands represent the 68% and 95% confidence intervals, respectively. Vertical red and purple bars in panels E-I represent the maxima and minima of the blue lines, respectively, for which the linkage to the observed WGD peaks in panel A was the strongest (minimum P-value in top-right corner of panels). For each of the panels E-I, sets consisting of the top-1 up to the top-18 extrema were tested by comparing the observed average nearest distance to a WGD peak for any given set of extrema to a null distribution of average nearest distances obtained by uniformly sampling 9 WGD peaks across the past 115 Ma a million times. The same randomization procedure was used for panels A and D but there only the full set of geological events (panel A) or angiosperm extinction rate peaks (panel D) was tested.

602 **STAR★METHODS**

603 **KEY RESOURCES TABLE**

REAGENT OR RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Coding sequence of 470 angiosperm species	Public data, reassembled in this paper	https://doi.org/10.6084/m9.figshare.27011128.v1
General Feature Format files of 470 angiosperm species	Public data, reassembled in this paper	https://doi.org/10.6084/m9.figshare.27011134.v1
Posterior samples of WGD dates for 132 independent WGD events in angiosperms	This paper	https://doi.org/10.6084/m9.figshare.27596856.v1
Software and algorithms		
BUSCO v5.4.3	Simão et al.[35]	https://busco.ezlab.org/
OMArk	Nevers et al.[36]	https://omark.omabrowser.org/
HMMER v3.1b2	Potter et al.[119]	https://www.ebi.ac.uk/Tools/hmmer/
MAFFT v7.310	Katoh et al.[120]	https://mafft.cbrc.jp/
TAPER	Zhang et al.[41]	https://github.com/chaoszhang/TAPER
IQTREE v1.6.5	Nguyen et al.[121]	http://www.iqtree.org/
ModelFinder	Kalyaanamoorthy et al.[122]	http://www.iqtree.org/ModelFinder/
ASTRAL-Pro	Zhang et al.[123]	https://github.com/chaoszhang/ASTER
PAML v4.9j	Ziheng Yang[124]	http://abacus.gene.ucl.ac.uk/software/paml.html
i-ADHoRe v3.0.01	Proost et al.[125]	https://github.com/VIB-PSB/i-ADHoRe
wgd v2	Chen et al.[126]	https://github.com/heche-psb/wgd
divDyn v0.8.3	Kocsis et al.[114]	https://cran.r-project.org/web/packages/divDyn/
SciPy v1.10.1	Virtanen et al.[127]	https://scipy.org/
Python v3.8.3	Python Software Foundation	https://www.python.org/
MATLAB R2024a	MathWorks	https://www.mathworks.com/

604

605

606 ## METHOD DETAILS

607 ### *Collection and quality assessment of genome assemblies*

608 We searched for all published angiosperm genome assemblies on the platform PlabiPD (January 5th, 2024).
609 We downloaded at least one representative genome for most genera via the link provided in the data
610 availability section of the corresponding article, leading to 470 genomes in total. To assess the quality of
611 the genome assemblies and annotations, we used BUSCO (v5.4.3)[35] ('protein' mode, embryophyta_odb10
612 database) and OMArk[36] (Viridiplantae.h5 database).

613 ### *Phylogenetic inference of angiosperms*

614 Reconstructing species trees from gene trees involves uncertainties and biases such as incomplete lineage
615 sorting (ILS) and sampling, horizontal gene transfer (HGT), gene flow, etc. Inferring orthologues de novo
616 from a sequence similarity graph, as implemented in some ortholog inference algorithms[128,129], is sensitive
617 to parameters such as the inflation factor[130], especially for single-copy orthogroups (SOGs). To avoid such
618 biases, we used the "reference orthologues" from the embryophyta_odb10 BUSCO dataset[37] to retrieve the
619 orthologues in each one of the 470 angiosperm genomes, which were used to build individual gene trees,
620 and summarize the species tree under the multispecies coalescence (MSC) model. Briefly, for each SOG
621 in the *embryophyta_odb10* database, we first used HMMER (v3.1b2)[119] to find corresponding orthologues
622 using the pre-built profile hidden Markov model and the cut-off table from BUSCO. When multiple hits
623 occurred, only the best hit was used for subsequent analysis. Secondly, a multiple sequence alignment
624 (MSA) of nucleotides was inferred for each SOG using MAFFT (v7.310)[120]. IQTREE (v1.6.5)[121] was then
625 applied to infer the maximum likelihood (ML) gene tree for each MSA with the automatic detection of the
626 best-fit substitution model via ModelFinder[122]. The uncertainties of branches were measured by the SH-
627 aLRT[131] and aBayes test with 1000 ultrafast bootstrap replicates, which were further optimized by a hill-
628 climbing nearest neighbor interchange (NNI) search.

629 To find the optimal branch support cut-off to use for filtering branches, we tested cut-offs of 0%, 10%, 20%,
630 30%, 40%, 50%, 60% and 70%. To account for the effect of different MSA data types on the branch length
631 and topology of inferred gene tree[132], we conducted the gene tree inference using i) codon-level nucleotide;
632 ii) 1st- and 2nd-position codon nucleotide; iii) TAPER-masked[53] nucleotide; and iv) raw nucleotide datasets.
633 Thus, we generated 32 gene tree datasets (eight branch support cut-offs x four MSA data types), each
634 containing gene trees inferred from 1614 SOGs. The 1614 individual gene trees of each dataset were then
635 supplied to ASTRAL-Pro[123] to summarize the species tree under default parameters (Supplementary
636 Information 1). We adopted the phylogenetic inference from the raw nucleotide dataset with the branch
637 support cut-off of 50% based on its concordance with the statistical topological test and the optimized trade-
638 off between high branch support of gene trees and high statistical support of the inferred phylogeny in terms
639 of local posterior probability (LPP) (Supplementary Note 3).

640 ### *Divergence time tree inference of angiosperms*

641 Using the species tree obtained with ASTRAL-Pro, we conducted molecular dating analyses based on the
642 concatenated amino acid MSA of 12 SOGs that covered at least 466 species (species coverage 99%) to
643 obtain a time-calibrated tree. The time tree was inferred using the uncorrelated lognormal (UCLN) relaxed
644 molecular clock model implemented in mcmctree (v4.9j)[124], with soft fossil calibration bounds. The model
645 parameters in mcmctree were set as follows: the clock model was set to 'independent rates model'; a log-
646 normal distribution of evolutionary rates across branches was set with a LG amino acid substitution matrix
647 and a gamma model with five rate categories and $\alpha = 0.5$; parameters for the birth-death process were set
648 as 1 1 0.1 to generate uniform age priors on nodes that did not have a fossil calibration; gamma priors for
649 the transition/transversion rate ratio and the shape parameter for variable rates among sites were set as 6
650 2 and 1 1; a Dirichlet-gamma prior was set with the parameters controlling the mean rate across loci and
651 the logarithmic variance as 2 20 1 and 1 10 1. The first 2000 generations were discarded as burn-in after
652 which 2,000,000 generations were pulled with sampling per 100 generations. To ensure accuracy and
653 consistency, we assembled an angiosperm fossil calibration dataset (Supplementary Note 4) and the full
654 age constraints implemented in our molecular clock analysis are shown in the calibration map
655 (Supplementary Information 2). The effective sample size (ESS) was higher than 200 for all parameters,
656 suggesting good convergence. Given the inferred time tree, the nucleotide substitution rate (Figure S2 in
657 Document S1) was further estimated based on the concatenated nucleotide MSA of the same 12 SOGs
658 using the best-fit substitution model predicted by ModelFinder[122].

659 ***Identification and dating of putative WGD events***

660 Intragenomic collinear segments and a burst of genes that seemed duplicated at the same time can be
661 indicative of a putative WGD event. Here, we used i-ADHoRe (v3.0.01)[125] and wgd v2[126] to search for
662 genomic collinearity and for bursts of gene duplication events in K_S-based age distributions, where the
663 number of duplicated genes is plotted against their age[105]. In short, the program "wgd dmd" was applied
664 with default parameters to construct the whole paranome (collection of all duplicated genes in a genome)
665 and the program "wgd syn" was applied with default parameters to infer genomic collinearity using i-
666 ADHoRe. The whole paranome and 'anchor pair' (pairs of duplicated genes derived from large-scale gene
667 duplication events and residing in duplicated segments) K_S distributions were then constructed using the
668 program "wgd ksd" with default parameters[126]. We applied exponential-lognormal mixture models (ELMM)
669 and log-scale Gaussian mixture models (GMM), as implemented in the program "wgd viz", to delineate
670 potential WGD components from the whole paranome and anchor pair K_S distributions, respectively. The
671 phylogenetic location of putative WGD events can be inferred by ranking the relative timing of WGD events
672 and speciation events within a K_S timeframe after proper correction of substitution rate variation[133]. The
673 former are represented as peaks of the delineated log-normal components in the whole paranome and
674 anchor pair K_S distributions while the latter are represented as peaks of the kernel density estimate (KDE)
675 of the K_S distributions for orthologous gene pairs, identified as the reciprocal best hits (RBH) across all
676 species.

677 For each putative WGD event, we compared its relative timing to a sequence of speciation events based
678 on prior investigations of earlier studies to delineate its phylogenetic location (detailed species composition
679 for each putative WGD event in Supplementary Note 6). The absolute dating of the identified WGD event
680 was conducted in a standardized way. In short, given an identified WGD event for a group of species
681 belonging to a certain order, we assembled an orthologous gene family dataset consisting of the anchor
682 pairs and their orthologues from species of other orders (detailed order composition and species
683 composition for each WGD event in Supplementary Note 6), in total 16-18 species per orthologous gene
684 family following Chen et al. (2024)[126]. Based on the constructed orthologous gene family and our main tree
685 topology, we conducted phylogenetic WGD dating on the concatenated amino acid MSA using mcmctree
686 (v4.9j) with the same composition of ordinal age constraints and the same parameters as implemented in
687 the divergence time tree inference section, except that the number of generations was increased to
688 20,000,000 and the sampling frequency to 1,000. For each WGD event, we dated at least one species, and
689 we adopted the consensus mean, calculated as the overall mean of all WGD age estimates for a WGD
690 event, and the 90% Highest Convergence Region (HCR), calculated as the intersection of the 90% Highest
691 Posterior Density (HPD) of the WGD date estimates across species for a WGD event, to represent the age
692 and the associated uncertainty of a WGD event. The estimated K_S peak and WGD date for each
693 independent WGD event are summarized in Table S6.

694 ***Diversity dependence of WGD establishment rate and the deviation***

695 The WGD establishment rate, measured by the number of established ancient WGDs per lineage per million
696 years (My), likely covaries with the number of lineages, as ecological factors such as competition and niche
697 overlapping are known to impact the establishment of polyploidy[10]. We applied a moving-window method
698 to estimate WGD establishment rates over time. We found a significant linearity between the estimated
699 WGD establishment rate profile and the number of lineages in the 466-species tree over time (R^2 varying
700 from 0.2734 to 0.8217 with varying window size, P<0.0001; Figure S3 in Document S1) and we showed
701 that a linear model fit better than a power-law model in terms of AIC and BIC (window sizes starting from 8
702 My to prevent singularities derived from log-transformation; Figure S4 in Document S1), supporting a linear
703 dependence of the WGD establishment rate on species richness. As the number of lineages grows
704 exponentially in time (R^2=0.9915, P<0.0001; Figure S5 in Document S1), the relationship between the WGD
705 establishment rate and time is approximately an exponential of the form a.exp(bt)+c with a<0 and b,c>0.
706 Substantial positive deviations of the WGD establishment rate profile from this relationship may indicate
707 peaks in WGD establishment activity related to other (geological, environmental, biological) phenomena.
708 To identify such peaks, a linear model was made of the WGD establishment rate moving average (window
709 size 4 My) as a function of time and the number of lineages through time. The resulting relative residuals
710 (observed data minus linear model estimates, normalized on the linear model estimates) are depicted in
711 Figure 3A. Peaks in the relative residuals profile were identified using the 'findpeaks' function in MATLAB
712 R2024a on a smoothened relative residuals profile (LOESS fit using the 'smooth' function in MATLAB
713 R2024a with span set to 10% of the total number of data points).

714 ***Leave-one-out cross-validation for optimal KDE bandwidth***

715 As an alternative to the moving-average analysis, we used kernel density estimation (KDE) to estimate the
716 WGD establishment rate profile. To ensure proper smoothness while avoiding overfitting, we used leave-
717 one-out cross-validation to find the optimal KDE bandwidth to model the WGD establishment rate over time.
718 We compared 200 bandwidths evenly distributed from 1 to 20 My based on the average log-likelihood per
719 test point. WGD age data was reflected around 0 Ma and the root age of the phylogenetic tree before KDE
720 to avoid boundary effects. We found that 1) starting from bandwidth 3.1, the increase in average log-
721 likelihood between adjacent bandwidths significantly slowed, i.e., became smaller than 5% of the total range
722 of average log-likelihoods for 10 consecutive steps; 2) after bandwidth 10.1, the average log-likelihood
723 declined monotonically, as shown in Figure S8 in Document S1. Given the marginal improvement of log-
724 likelihood in between, we used a lower bandwidth (3.1) that gave higher resolution to detect WGD peaks.
725 The resulting KDE curve was converted to a curve of WGD establishment rate per lineage by dividing the
726 kernel density estimates by the number of lineages in the species tree at each time point.

727 ***Exponential-Gaussian mixture model of WGD establishment rate***

728 To assess how many peaks occur in the WGD establishment rate profile obtained through KDE, we fit
729 exponential-Gaussian mixture models with an exponential background and 6 to 10 Gaussian components
730 to the profile using the CurveFitter toolbox in MATLAB R2024a. An exponential background of the form
731 a.exp(bt)-a with a<0 and b>0 was used to model the observed lineage diversity dependence of the WGD
732 establishment rate (with the term -a used to force the background to go through the origin at 0 Ma). The
733 highest adjusted R^2 value was achieved for the model with 9 Gaussian components, including one
734 underlying a non-visible peak at ~66 Ma that is masked by the peaks before and after it.

735 ***Homogeneous Poisson process model of WGD simulation***

736 Inspired by Vanneste and co-authors[24], we devised a basic null model of WGD simulation in Python 3.8.3
737 whereby WGD events along each branch of the time-calibrated phylogenetic tree were modeled as an
738 independent homogeneous Poisson process with the time interval between events exponentially distributed
739 and controlled by a WGD establishment rate parameter. To accommodate potential variation of WGD
740 establishment rates across different clades, we allowed clade-specific rate parameters for the ANA grade,
741 Ceratophyllales, Chloranthales, eudicots, monocots and magnoliids, respectively. We calculated the clade-
742 specific rate parameters that maximized the likelihood given the independent homogeneous Poisson
743 process model and the observed WGD events as the phylogenetic background rate parameters for each
744 clade (Table S7). We also validated that this variable-rate model fit the data significantly better than the
745 alternative simpler model with a constant rate across all clades (likelihood ratio test, P <0.0001). Under this
746 independent homogeneous Poisson process model and the clade-specific background rate parameters,
747 we simulated WGDs throughout the main tree and documented the randomly generated WGD ages for
748 10,000 replicates to construct the null distribution. KDE profiles of the WGD establishment rate were
749 constructed for each set of simulated WGD ages as described above, using the same bandwidth (3.1). The

750 mean, median, 50%CI, and 95%CI of the simulated WGD establishment rate profiles were calculated over
751 time and compared to the observed WGD establishment rate profile.

752 ***Non-homogeneous Poisson process model of WGD simulation***

753 As the WGD establishment rate is related exponentially to time because of the observed negative linear
754 lineage diversity-dependence of WGD establishment and the exponential increase of lineages over time,
755 we devised a non-homogeneous Poisson process (NHPP) model to build an alternative null distribution of
756 WGDs accounting for the WGD establishment rate variation across time using a standard thinning method.
757 In this model, WGDs are generated following a Poisson process for each phylogenetic branch with
758 exponentially decaying rate from the root to the tips of the tree. The WGD establishment rate is assumed
759 to follow the functional form a.exp(bt) + c over time t, with parameters a < 0, b, c > 0. We first applied a
760 global optimization approach using the differential evolution algorithm (maximum 10,000 iterations) from
761 the SciPy library in Python 3.8.3 to estimate the parameters a,b and c of the NHPP model based on the
762 observed WGD profile. We varied the random seed as an integer from 0 to 9,999 and repeated the
763 optimization process 10,000 times, the median estimate of which was then integrated into the WGD
764 establishment rate function for simulation, i.e., a = -0.01506, b = 0.01949, c = 0.01538. Subsequently,
765 10,000 NHPP model simulations were performed and KDE profiles of the WGD establishment rate were
766 constructed for each set of simulated WGD ages as described above, using the same bandwidth (3.1). The
767 mean, median, 50%CI, and 95%CI of the simulated WGD establishment rate profiles were calculated over
768 time and compared to the observed WGD establishment rate profile. We also increased the number of
769 simulations to one million and obtained nearly identical results, confirming the sufficiency of our simulation.

770 ***Fossil-based genus-level extinction rate estimation***

771 We downloaded the latest angiosperm fossil occurrence data from the Paleobiology database[113] on July
772 10th 2025, for the time frame spanning the Triassic until now. We circumvented potential homonymies by
773 combining the class names and genus names to create individual entries. Multiple species of the same
774 genus that were registered in the same collection are only counted as one occurrence. To take into account
775 variations derived from different methodologies, we implemented three methods of sampling
776 standardization (i.e., raw pattern, Classical Rarefaction[134] and Shareholder Quorum Subsampling[135]) and
777 four macroevolutionary rate models (i.e., per capita rates[136], corrected three-timer rates[137], gap-filler
778 equations[138] and second-for-third substitution rates[139]) using divDyn (v0.8.3)[114] to estimate angiosperm
779 genus−level extinction rates based on the time-scale of geological stages[140].

780 ***Statistical tests for association between WGD establishment rate peaks, geological events,***
781 ***extinction events and paleoclimatic extrema***

782 Nine peaks in the relative residuals of the linear model for the WGD establishment rate as a function of time
783 and the number of lineages were supported by Gaussian components in the optimal exponential-Gaussian
784 mixture model. Associations between these peaks and geological events, genus-level angiosperm
785 extinction events and extrema in paleoclimatic variables, as depicted in Figure 3, were assessed through

786 two types of Monte Carlo randomization analysis. In the first set of analyses, for each data layer the WGD
787 peaks are compared to, 9 peak ages were randomly sampled a million times from a uniform distribution on
788 the interval [0 Ma, 115 Ma]. For each random sample, the average nearest distance of the sampled WGD
789 peak ages to the ages of 'events' in the other data layer were computed. Event ages are defined as the
790 mean ages for geological events, the ages of maxima in the genus-level angiosperm extinction rate curve,
791 and the ages of maxima and minima in the other data layers. Average nearest distances were computed
792 with respect to the peaks in the other data layer, not the WGD peaks. For the paleoclimatic profiles
793 exhibiting many peaks of varying peak importance ($\delta^{13}C$, $\delta^{18}O$, CO_2 sea surface temperature, global mean
794 surface temperature), only the top-18 minima and maxima in terms of peak importance were taken into
795 account, and association analyses were done for different peak sets containing the top-x peaks with
796 x=1...18. For each peak set tested, the average nearest distance of the observed WGD peak ages to the
797 peak set was compared to the sampled average nearest distances under the null hypothesis of uniformly
798 distributed peak locations, and a P-value for peak association was computed as P=(k+1)/(n+1) with k as
799 the number of times the sampled average nearest distance is lower than or equal to the observed average
800 nearest distance and n as the number of random samples (1 million)[141].

801 The second set of randomization analyses and P-value calculations was performed in the same way, except
802 that distances between WGD peaks were sampled instead of WGD peak ages. For each random sample,
803 an age in the interval [0 Ma, 30 Ma] was sampled to position the youngest random WGD peak, and
804 subsequent peaks were positioned at inter-peak distances sampled from a Gaussian distribution fit to the
805 inter-peak distances for the 9 observed WGD peaks (μ=11.57, σ=3.37 Ma).

806 All association analyses were done in MATLAB R2024a. Maxima were computed using the 'findpeaks'
807 function on the raw time profile (for the genus-level angiosperm extinction rate, global mean surface
808 temperature and CO_2 concentration), a smoothing spline (smoothing parameter 0.05 for $\delta^{13}C$ and $\delta^{18}O$
809 isotopic signature profiles) or a LOESS fit (with span = 2% for the sea surface temperature data) . Minima
810 were computed using the 'findpeaks' function on the opposite (additively inverted) profiles.

811

812 **QUANTIFICATION AND STATISTICAL ANALYSIS**

813 In the statistical test for the support of competing topologies pertaining to Chloranthales, we conducted one-
814 way ANOVA and Tukey's HSD test to determine which topology gets significantly higher support among
815 different datasets, details in Supplementary Note 3. In the statistical test for the support of competing
816 topologies pertaining to Crossosomatales, we conducted two-sided t-test and Mann-Whitney U test to
817 determine which topology gets significantly higher support among different datasets, details in
818 Supplementary Note 3. In the delineation of random sampling bias, we conducted ordinary linear regression
819 to demonstrate the significantly reduced random sampling bias of our main tree, details in Supplementary
820 Note 3. To compare our divergence time estimates with other studies, we calculated the Pearson correlation
821 coefficients of the estimated stem order and family ages between our main tree and that of Ramírez-

822 Barahona et al. (2020)[43], details in Supplementary Note 5. Relative nucleotide and amino acid substitution
823 rates across angiosperms were compared by one-way ANOVA and Tukey's HSD test, details in
824 Supplementary Note 5. Significance is defined by a P-value lower than 0.05, as denoted alongside the
825 associated statistical tests.

826

827 **ADDITIONAL RESOURCES**

828 Description: https://bioinformatics.psb.ugent.be/AngioWGD/

829 To adhere to the FAIR data principles (Findable, Accessible, Interoperable, and Reproducible), we
830 developed the *AngioWGD* website which can be used to visualize and download data on WGD events
831 across the plant phylogeny.

832

833

REFERENCES

1. Wendel, J.F. (2015). The wondrous cycles of polyploidy in plants. American Journal of Botany *102*, 1753-1756. https://doi.org/10.3732/ajb.1500320.
2. Haldane, J. (1933). The part played by recurrent mutation in evolution. Am. Nat. *67*, 5-19. https://doi.org/10.1086/280465.
3. Wagner Jr., W.H. (1970). Biosystematics and Evolutionary Noise. Taxon *19*, 146-151. https://doi.org/10.2307/1217945.
4. Stebbins, G.L. (1950). Chapter VI. Isolation and the Origin of Species. In Variation and Evolution in Plants, (Columbia University Press), pp. 189-250. https://doi.org/10.7312/steb94536-007.
5. Stebbins, G.L. (1971). Chromosomal evolution in higher plants.
6. Rice, A., Šmarda, P., Novosolov, M., Drori, M., Glick, L., Sabath, N., Meiri, S., Belmaker, J., and Mayrose, I. (2019). The global biogeography of polyploid plants. Nat. Ecol. Evol. *3*, 265-273. https://doi.org/10.1038/s41559-018-0787-9.
7. Landis, J.B., Soltis, D.E., Li, Z., Marx, H.E., Barker, M.S., Tank, D.C., and Soltis, P.S. (2018). Impact of whole-genome duplication events on diversification rates in angiosperms. American Journal of Botany *105*, 348-363. https://doi.org/10.1002/ajb2.1060.
8. Van de Peer, Y., Mizrachi, E., and Marchal, K. (2017). The evolutionary significance of polyploidy. Nature Reviews Genetics *18*, 411-424. 10.1038/nrg.2017.26.
9. Carretero-Paulet, L., and Van de Peer, Y. (2020). The evolutionary conundrum of whole-genome duplication. American Journal of Botany *107*, 1101-1105. https://doi.org/10.1002/ajb2.1520.
10. Mortier, F., Bafort, Q., Milosavljevic, S., Kauai, F., Prost Boxoen, L., Van de Peer, Y., and Bonte, D. (2024). Understanding polyploid establishment: temporary persistence or stable coexistence? Oikos *2024*, e09929. https://doi.org/10.1111/oik.09929.
11. Comai, L. (2005). The advantages and disadvantages of being polyploid. Nature Reviews Genetics *6*, 836-846. 10.1038/nrg1711.
12. Bomblies, K. (2020). When everything changes at once: finding a new normal after genome duplication. Proceedings of the Royal Society B: Biological Sciences *287*, 20202154. doi:10.1098/rspb.2020.2154.
13. Van de Peer, Y., Ashman, T.-L., Soltis, P.S., and Soltis, D.E. (2021). Polyploidy: an evolutionary and ecological force in stressful times. Plant Cell *33*, 11-26. https://doi.org/10.1093/plcell/koaa015.
14. Almeida-Silva, F., and Van de Peer, Y. (2023). Whole-genome Duplications and the Long-term Evolution of Gene Regulatory Networks in Angiosperms. Mol. Biol. Evol. *40*. https://doi.org/10.1093/molbev/msad141.
15. Baduel, P., Bray, S., Vallejo-Marin, M., Kolář, F., and Yant, L. (2018). The “Polyploid Hop”: Shifting Challenges and Opportunities Over the Evolutionary Lifespan of Genome Duplications. Front. Ecol. Evol. *6*. https://doi.org/10.3389/fevo.2018.00117.
16. Doyle, J.J., and Coate, J.E. (2019). Polyploidy, the Nucleotype, and Novelty: The Impact of Genome Doubling on the Biology of the Cell. Int J Plant Sci *180*, 1-52. 10.1086/700636.
17. Chao, D.-Y., Dilkes, B., Luo, H., Douglas, A., Yakubova, E., Lahner, B., and Salt, D.E. (2013). Polyploids Exhibit Higher Potassium Uptake and Salinity Tolerance in Arabidopsis. Science *341*, 658-659. 10.1126/science.1240561.
18. Del Pozo, J.C., and Ramirez-Parra, E. (2014). Deciphering the molecular bases for drought tolerance in Arabidopsis autotetraploids. Plant, Cell & Environment *37*, 2722-2737. https://doi.org/10.1111/pce.12344.
19. Ruiz, M., Quiñones, A., Martínez-Cuenca, M.R., Aleza, P., Morillon, R., Navarro, L., Primo-Millo, E., and Martínez-Alcántara, B. (2016). Tetraploidy enhances the ability to exclude chloride from

leaves in carrizo citrange seedlings. Journal of Plant Physiology *205*, 1-10. https://doi.org/10.1016/j.jplph.2016.08.002.

20. Yang, P.M., Huang, Q.C., Qin, G.Y., Zhao, S.P., and Zhou, J.G. (2014). Different drought-stress responses in photosynthesis and reactive oxygen metabolism between autotetraploid and diploid rice. Photosynthetica *52*, 193-202. https://doi.org/10.1007/s11099-014-0020-2.
21. Morlon, H., Potts, M.D., and Plotkin, J.B. (2010). Inferring the Dynamics of Diversification: A Coalescent Approach. PLoS Biol. *8*, e1000493. 10.1371/journal.pbio.1000493.
22. Condamine, F.L., Rolland, J., and Morlon, H. (2019). Assessing the causes of diversification slowdowns: temperature-dependent and diversity-dependent models receive equivalent support. Ecol. Lett. *22*, 1900-1912. https://doi.org/10.1111/ele.13382.
23. Fawcett, J.A., Maere, S., and Van de Peer, Y. (2009). Plants with double genomes might have had a better chance to survive the Cretaceous–Tertiary extinction event. Proc. Natl. Acad. Sci. U.S.A. *106*, 5737-5742. doi:10.1073/pnas.0900906106.
24. Vanneste, K., Baele, G., Maere, S., and Van de Peer, Y. (2014). Analysis of 41 plant genomes supports a wave of successful genome duplications in association with the Cretaceous–Paleogene boundary. Genome Res. *24*, 1334-1347. 10.1101/gr.168997.113.
25. Heslop-Harrison, J.S., Schwarzacher, T., and Liu, Q. (2022). Polyploidy: its consequences and enabling role in plant diversification and evolution. Ann. Bot. *131*, 1-10. 10.1093/aob/mcac132.
26. Leebens-Mack, J.H., Barker, M.S., Carpenter, E.J., Deyholos, M.K., Gitzendanner, M.A., Graham, S.W., Grosse, I., Li, Z., Melkonian, M., Mirarab, S., et al. (2019). One thousand plant transcriptomes and the phylogenomics of green plants. Nature *574*, 679-685. https://doi.org/10.1038/s41586-019-1693-2.
27. Ren, R., Wang, H., Guo, C., Zhang, N., Zeng, L., Chen, Y., Ma, H., and Qi, J. (2018). Widespread Whole Genome Duplications Contribute to Genome Complexity and Species Diversity in Angiosperms. Mol. Plant *11*, 414-428. https://doi.org/10.1016/j.molp.2018.01.002.
28. McKibben, M.T.W., Finch, G., and Barker, M.S. (2024). Species-tree topology impacts the inference of ancient whole-genome duplications across the angiosperm phylogeny. Am. J. Bot. *111*, e16378. https://doi.org/10.1002/ajb2.16378.
29. Clark, J.W., and Donoghue, P.C.J. (2018). Whole-Genome Duplication and Plant Macroevolution. Trends Plant Sci. *23*, 933-945. https://doi.org/10.1016/j.tplants.2018.07.006.
30. Roelofs, D., Zwaenepoel, A., Sistermans, T., Nap, J., Kampfraath, A.A., Van de Peer, Y., Ellers, J., and Kraaijeveld, K. (2020). Multi-faceted analysis provides little evidence for recurrent whole-genome duplications during hexapod evolution. BMC Biol. *18*, 57. 10.1186/s12915-020-00789-1.
31. Chen, H., Fang, Y., Zwaenepoel, A., Huang, S., Van de Peer, Y., and Li, Z. (2023). Revisiting ancient polyploidy in leptosporangiate ferns. New Phytol. *237*, 1405-1417. https://doi.org/10.1111/nph.18607.
32. Wang, X., Wang, J., Jin, D., Guo, H., Lee, T.-H., Liu, T., and Paterson, Andrew H. (2015). Genome Alignment Spanning Major Poaceae Lineages Reveals Heterogeneous Evolutionary Rates and Alters Inferred Dates for Key Evolutionary Events. Mol. Plant *8*, 885-898. https://doi.org/10.1016/j.molp.2015.04.004.
33. Ruprecht, C., Lohaus, R., Vanneste, K., Mutwil, M., Nikoloski, Z., Van de Peer, Y., and Persson, S. (2017). Revisiting ancestral polyploidy in plants. Science Advances *3*, e1603195. doi:10.1126/sciadv.1603195.
34. Parham, J.F., Donoghue, P.C.J., Bell, C.J., Calway, T.D., Head, J.J., Holroyd, P.A., Inoue, J.G., Irmis, R.B., Joyce, W.G., Ksepka, D.T., et al. (2011). Best Practices for Justifying Fossil Calibrations. Systematic Biology *61*, 346-359. 10.1093/sysbio/syr107.

35. Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics *31*, 3210-3212. https://doi.org/10.1093/bioinformatics/btv351.
36. Nevers, Y., Warwick Vesztrocy, A., Rossier, V., Train, C.-M., Altenhoff, A., Dessimoz, C., and Glover, N.M. (2024). Quality assessment of gene repertoire annotations with OMArk. Nat. Biotechnol., 1-10. https://doi.org/10.1038/s41587-024-02147-w.
37. Kriventseva, E.V., Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simão, F.A., and Zdobnov, E.M. (2018). OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. Nucleic Acids Res. *47*, D807-D811. https://doi.org/10.1093/nar/gky1053.
38. Zuntini, A.R., Carruthers, T., Maurin, O., Bailey, P.C., Leempoel, K., Brewer, G.E., Epitawalage, N., Françoso, E., Gallego-Paramo, B., McGinnie, C., et al. (2024). Phylogenomics and the rise of the angiosperms. Nature *629*, 843-850. https://doi.org/10.1038/s41586-024-07324-0.
39. Hu, H., Sun, P., Yang, Y., Ma, J., and Liu, J. (2023). Genome-scale angiosperm phylogenies based on nuclear, plastome, and mitochondrial datasets. Journal of Integrative Plant Biology *65*, 1479-1489. https://doi.org/10.1111/jipb.13455.
40. Yang, L., Su, D., Chang, X., Foster, C.S.P., Sun, L., Huang, C.-H., Zhou, X., Zeng, L., Ma, H., and Zhong, B. (2020). Phylogenomic Insights into Deep Phylogeny of Angiosperms Based on Broad Nuclear Gene Sampling. Plant Communications *1*, 100027. https://doi.org/10.1016/j.xplc.2020.100027.
41. Zhang, C., Zhao, Y., Braun, E.L., and Mirarab, S. (2021). TAPER: Pinpointing errors in multiple sequence alignments despite varying rates of evolution. Methods in Ecology and Evolution *12*, 2145-2158. https://doi.org/10.1111/2041-210X.13696.
42. APG, T. (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. Bot. J. Linn. Soc. *181*, 1-20. https://doi.org/10.1111/boj.12385.
43. Ramírez-Barahona, S., Sauquet, H., and Magallón, S. (2020). The delayed and geographically heterogeneous diversification of flowering plant families. Nature Ecology & Evolution *4*, 1232-1238. 10.1038/s41559-020-1241-3.
44. Jiao, Y., Li, J., Tang, H., and Paterson, A.H. (2014). Integrated Syntenic and Phylogenomic Analyses Reveal an Ancient Genome Duplication in Monocots. The Plant Cell *26*, 2792-2802. 10.1105/tpc.114.127597.
45. Jiao, Y., Leebens-Mack, J., Ayyampalayam, S., Bowers, J.E., McKain, M.R., McNeal, J., Rolf, M., Ruzicka, D.R., Wafula, E., Wickett, N.J., et al. (2012). A genome triplication associated with early diversification of the core eudicots. Genome Biology *13*, R3. 10.1186/gb-2012-13-1-r3.
46. Rampino, M.R., Caldeira, K., and Rodriguez, S. (2023). Cycles of ∼32.5 My and ∼26.2 My in correlated episodes of continental flood basalts (CFBs), hyper-thermal climate pulses, anoxic oceans, and mass extinctions over the last 260 My: Connections between geological and astronomical cycles. Earth-Sci. Rev. *246*, 104548. https://doi.org/10.1016/j.earscirev.2023.104548.
47. Scotese, C.R., Song, H., Mills, B.J.W., and van der Meer, D.G. (2021). Phanerozoic paleotemperatures: The earth's changing climate during the last 540 million years. Earth-Science Reviews *215*, 103503. https://doi.org/10.1016/j.earscirev.2021.103503.
48. Meyer, K.M., and Kump, L.R. (2008). Oceanic Euxinia in Earth History: Causes and Consequences. Annual Review of Earth and Planetary Sciences *36*, 251-288. https://doi.org/10.1146/annurev.earth.36.031207.124256.
49. Jenkyns, H.C. (2010). Geochemistry of oceanic anoxic events. Geochemistry, Geophysics, Geosystems *11*. https://doi.org/10.1029/2009GC002788.

50. Linnert, C., Robinson, S.A., Lees, J.A., Bown, P.R., Pérez-Rodríguez, I., Petrizzo, M.R., Falzoni, F., Littler, K., Arz, J.A., and Russell, E.E. (2014). Evidence for global cooling in the Late Cretaceous. Nat. Commun. *5*, 4194. 10.1038/ncomms5194.
51. Barreda, V.D., Palazzesi, L., and Olivero, E.B. (2019). When flowering plants ruled Antarctica: evidence from Cretaceous pollen grains. New Phytol. *223*, 1023-1030. https://doi.org/10.1111/nph.15823.
52. Miller, K.G., Wright, J.D., and Browning, J.V. (2005). Visions of ice sheets in a greenhouse world. Mar. Geol. *217*, 215-231. https://doi.org/10.1016/j.margeo.2005.02.007.
53. Koch, M.C., and Friedrich, O. (2012). Campanian-Maastrichtian intermediate- to deep-water changes in the high latitudes: Benthic foraminiferal evidence. Paleoceanography *27*. 10.1029/2011pa002259.
54. Barnet, J.S.K., Littler, K., Kroon, D., Leng, M.J., Westerhold, T., Röhl, U., and Zachos, J.C. (2017). A new high-resolution chronology for the late Maastrichtian warming event: Establishing robust temporal links with the onset of Deccan volcanism. Geology *46*, 147-150. 10.1130/g39771.1.
55. Renne, P.R., Deino, A.L., Hilgen, F.J., Kuiper, K.F., Mark, D.F., Mitchell, W.S., Morgan, L.E., Mundil, R., and Smit, J. (2013). Time Scales of Critical Events Around the Cretaceous-Paleogene Boundary. Science *339*, 684-687. doi:10.1126/science.1230492.
56. Schulte, P., Alegret, L., Arenillas, I., Arz, J.A., Barton, P.J., Bown, P.R., Bralower, T.J., Christeson, G.L., Claeys, P., Cockell, C.S., et al. (2010). The Chicxulub Asteroid Impact and Mass Extinction at the Cretaceous-Paleogene Boundary. Science *327*, 1214-1218. doi:10.1126/science.1177265.
57. Thompson, J.B., and Ramírez-Barahona, S. (2023). No phylogenetic evidence for angiosperm mass extinction at the Cretaceous–Palaeogene (K-Pg) boundary. Biol. Lett. *19*, 20230314. doi:10.1098/rsbl.2023.0314.
58. Zachos, J.C., Dickens, G.R., and Zeebe, R.E. (2008). An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. Nature *451*, 279-283. 10.1038/nature06588.
59. Dunkley Jones, T., Lunt, D.J., Schmidt, D.N., Ridgwell, A., Sluijs, A., Valdes, P.J., and Maslin, M. (2013). Climate model and proxy data constraints on ocean warming across the Paleocene–Eocene Thermal Maximum. Earth-Science Reviews *125*, 123-145. https://doi.org/10.1016/j.earscirev.2013.07.004.
60. Bowen, G.J., Maibauer, B.J., Kraus, M.J., Röhl, U., Westerhold, T., Steimke, A., Gingerich, P.D., Wing, S.L., and Clyde, W.C. (2015). Two massive, rapid releases of carbon during the onset of the Palaeocene–Eocene thermal maximum. Nat. Geosci. *8*, 44-47. 10.1038/ngeo2316.
61. Zeebe, R.E., Zachos, J.C., and Dickens, G.R. (2009). Carbon dioxide forcing alone insufficient to explain Palaeocene–Eocene Thermal Maximum warming. Nature Geoscience *2*, 576-580. 10.1038/ngeo578.
62. Zeebe, R.E., Ridgwell, A., and Zachos, J.C. (2016). Anthropogenic carbon release rate unprecedented during the past 66 million years. Nature Geoscience *9*, 325-329. 10.1038/ngeo2681.
63. Gutjahr, M., Ridgwell, A., Sexton, P.F., Anagnostou, E., Pearson, P.N., Pälike, H., Norris, R.D., Thomas, E., and Foster, G.L. (2017). Very large release of mostly volcanic carbon during the Palaeocene–Eocene Thermal Maximum. Nature *548*, 573-577. 10.1038/nature23646.
64. Svensen, H., Planke, S., Malthe-Sørenssen, A., Jamtveit, B., Myklebust, R., Rasmussen Eidem, T., and Rey, S.S. (2004). Release of methane from a volcanic basin as a mechanism for initial Eocene global warming. Nature *429*, 542-545. 10.1038/nature02566.
65. Korasidis, V.A., Wing, S.L., Shields, C.A., and Kiehl, J.T. (2022). Global Changes in Terrestrial Vegetation and Continental Climate During the Paleocene-Eocene Thermal Maximum. Paleoceanography and Paleoclimatology *37*, e2021PA004325. https://doi.org/10.1029/2021PA004325.

66. WAGNER, J.D., PEPPE, D.J., O'KEEFE, J.M.K., and DENISON, C.N. (2023). PLANT COMMUNITY CHANGE ACROSS THE PALEOCENE–EOCENE BOUNDARY IN THE GULF COASTAL PLAIN, CENTRAL TEXAS. Palaios *38*, 436-451. 10.2110/palo.2022.008.
67. Cai, L., Xi, Z., Amorim, A.M., Sugumaran, M., Rest, J.S., Liu, L., and Davis, C.C. (2019). Widespread ancient whole-genome duplications in Malpighiales coincide with Eocene global climatic upheaval. New Phytologist *221*, 565-576. https://doi.org/10.1111/nph.15357.
68. Costa, E., Garcés, M., Sáez, A., Cabrera, L., and López-Blanco, M. (2011). The age of the "Grande Coupure" mammal turnover: New constraints from the Eocene–Oligocene record of the Eastern Ebro Basin (NE Spain). Palaeogeogr., Palaeoclimatol., Palaeoecol. *301*, 97-107. https://doi.org/10.1016/j.palaeo.2011.01.005.
69. Ivany, L.C., Patterson, W.P., and Lohmann, K.C. (2000). Cooler winters as a possible cause of mass extinctions at the Eocene/Oligocene boundary. Nature *407*, 887-890. 10.1038/35038044.
70. Houben, A.J.P., van Mourik, C.A., Montanari, A., Coccioni, R., and Brinkhuis, H. (2012). The Eocene–Oligocene transition: Changes in sea level, temperature or both? Palaeogeogr., Palaeoclimatol., Palaeoecol. *335-336*, 75-83. https://doi.org/10.1016/j.palaeo.2011.04.008.
71. Hutchinson, D.K., Coxall, H.K., Lunt, D.J., Steinthorsdottir, M., de Boer, A.M., Baatsen, M., von der Heydt, A., Huber, M., Kennedy-Asser, A.T., Kunzmann, L., et al. (2021). The Eocene–Oligocene transition: a review of marine and terrestrial proxy data, models and model–data comparisons. Clim. Past *17*, 269-315. 10.5194/cp-17-269-2021.
72. Nge, F.J., Biffin, E., Thiele, K.R., and Waycott, M. (2020). Extinction pulse at Eocene–Oligocene boundary drives diversification dynamics of two Australian temperate floras. Proc. R. Soc. B Biol. Sci. *287*, 20192546. doi:10.1098/rspb.2019.2546.
73. O'Brien, C.L., Huber, M., Thomas, E., Pagani, M., Super, J.R., Elder, L.E., and Hull, P.M. (2020). The enigma of Oligocene climate and global surface temperature evolution. Proc. Natl. Acad. Sci. U.S.A. *117*, 25302-25309. doi:10.1073/pnas.2003914117.
74. Pearson, P.N., and Palmer, M.R. (2000). Atmospheric carbon dioxide concentrations over the past 60 million years. Nature *406*, 695-699. 10.1038/35021000.
75. Shevenell, A.E., Kennett, J.P., and Lea, D.W. (2004). Middle Miocene Southern Ocean cooling and Antarctic cryosphere expansion. Science *305*, 1766-1770. 10.1126/science.1100061.
76. Flower, B.P., and Kennett, J.P. (1994). The middle Miocene climatic transition: East Antarctic ice sheet development, deep ocean circulation and global carbon cycling. Palaeogeogr., Palaeoclimatol., Palaeoecol. *108*, 537-555. https://doi.org/10.1016/0031-0182(94)90251-8.
77. Melott, A.L., and Bambach, R.K. (2014). Analysis of periodicity of extinction using the 2012 geological timescale. Paleobiology *40*, 177-196. 10.1666/13047.
78. Raup, D.M., and Sepkoski, J.J. (1984). Periodicity of extinctions in the geologic past. Proc. Natl. Acad. Sci. U.S.A. *81*, 801-805. doi:10.1073/pnas.81.3.801.
79. Lewis, A.R., Marchant, D.R., Ashworth, A.C., Hedenäs, L., Hemming, S.R., Johnson, J.V., Leng, M.J., Machlus, M.L., Newton, A.E., Raine, J.I., et al. (2008). Mid-Miocene cooling and the extinction of tundra in continental Antarctica. Proc. Natl. Acad. Sci. U.S.A. *105*, 10676-10680. doi:10.1073/pnas.0802501105.
80. Zhang, R., Guo, J., Bradshaw, C.D., Xu, X., Shen, T., Li, S., Nie, J., Zhang, C., Li, X., Liu, Z., et al. (2025). Vegetation feedbacks accelerated the late Miocene climate transition. Science Advances *11*, eads4268. doi:10.1126/sciadv.ads4268.
81. Ravelo, A.C., and Hillaire-Marcel, C. (2007). Chapter Eighteen The Use of Oxygen and Carbon Isotopes of Foraminifera in Paleoceanography. In Developments in Marine Geology, C. Hillaire–Marcel, and A. De Vernal, eds. (Elsevier), pp. 735-764. https://doi.org/10.1016/S1572-5480(07)01023-8.

1068 82. Soltis, P.S., Marchant, D.B., Van de Peer, Y., and Soltis, D.E. (2015). Polyploidy and genome
1069 evolution in plants. Curr. Opin. Genet. Dev. *35*, 119-125.
1070 https://doi.org/10.1016/j.gde.2015.11.003.
1071 83. Carruthers, T., Gonçalves, D.J.P., Li, P., Chanderbali, A.S., Dick, C.W., Fritsch, P.W., Larson, D.A.,
1072 Soltis, D.E., Soltis, P.S., Weaver, W.N., and Smith, S.A. (2024). Repeated shifts out of tropical
1073 climates preceded by whole genome duplication. New Phytol. *244*, 2561-2575.
1074 https://doi.org/10.1111/nph.20200.
1075 84. Li, X., Wang, M., Zou, M., Guan, X., Xu, S., Chen, W., Wang, C., Chen, Y., He, S., and Guo, B. (2024).
1076 Recent and Recurrent Autopolyploidization Fueled Diversification of Snow Carp on the Tibetan
1077 Plateau. Mol. Biol. Evol. *41*, msae221. 10.1093/molbev/msae221.
1078 85. McPeek, Mark A. (2008). The Ecological Dynamics of Clade Diversification and Community
1079 Assembly. Am. Nat. *172*, E270-E284. 10.1086/593137.
1080 86. Phillimore, A.B., and Price, T.D. (2008). Density-Dependent Cladogenesis in Birds. PLoS Biol. *6*, e71.
1081 10.1371/journal.pbio.0060071.
1082 87. Moen, D., and Morlon, H. (2014). Why does diversification slow down? Trends Ecol. Evol. *29*, 190-
1083 197. 10.1016/j.tree.2014.01.010.
1084 88. Rabosky, D.L., and Lovette, I.J. (2008). EXPLOSIVE EVOLUTIONARY RADIATIONS: DECREASING
1085 SPECIATION OR INCREASING EXTINCTION THROUGH TIME? Evolution *62*, 1866-1875.
1086 10.1111/j.1558-5646.2008.00409.x.
1087 89. Aguilée, R., Gascuel, F., Lambert, A., and Ferriere, R. (2018). Clade diversification dynamics and
1088 the biotic and abiotic controls of speciation and extinction rates. Nat. Commun. *9*, 3013.
1089 10.1038/s41467-018-05419-7.
1090 90. Kergoat, G.J., Condamine, F.L., Toussaint, E.F.A., Capdevielle-Dulac, C., Clamens, A.-L., Barbut, J.,
1091 Goldstein, P.Z., and Le Ru, B. (2018). Opposite macroevolutionary responses to environmental
1092 changes in grasses and insects during the Neogene grassland expansion. Nat. Commun. *9*, 5089.
1093 10.1038/s41467-018-07537-8.
1094 91. Lewitus, E., Bittner, L., Malviya, S., Bowler, C., and Morlon, H. (2018). Clade-specific diversification
1095 dynamics of marine diatoms since the Jurassic. Nat. Ecol. Evol. *2*, 1715-1723. 10.1038/s41559-
1096 018-0691-3.
1097 92. De Storme, N., Copenhaver, G.P., and Geelen, D. (2012). Production of diploid male gametes in
1098 Arabidopsis by cold-induced destabilization of postmeiotic radial microtubule arrays. Plant Physiol
1099 *160*, 1808-1826. 10.1104/pp.112.208611.
1100 93. Ebadi, M., Bafort, Q., Mizrachi, E., Audenaert, P., Simoens, P., Van Montagu, M., Bonte, D., and
1101 Van de Peer, Y. (2023). The duplication of genomes and genetic networks and its potential for
1102 evolutionary adaptation and survival during environmental turmoil. Proc. Natl. Acad. Sci. U.S.A.
1103 *120*, e2307289120. doi:10.1073/pnas.2307289120.
1104 94. Kauai, F., Van de Peer, Y., and Bonte, D. (2025). Ecological opportunity and the onset of polyploid
1105 niche expansion waves. bioRxiv, 2025.2007.2011.664136. 10.1101/2025.07.11.664136.
1106 95. Guo, L., Wang, S., Jiao, X., Ye, X., Deng, D., Liu, H., Li, Y., Van de Peer, Y., and Wu, W. (2024).
1107 Convergent and/or parallel evolution of RNA-binding proteins in angiosperms after
1108 polyploidization. New Phytol. *242*, 1377-1393. https://doi.org/10.1111/nph.19656.
1109 96. Wu, S., Han, B., and Jiao, Y. (2020). Genetic Contribution of Paleopolyploidy to Adaptive Evolution
1110 in Angiosperms. Mol. Plant *13*, 59-71. 10.1016/j.molp.2019.10.012.
1111 97. Li, Z., McKibben, M.T.W., Finch, G.S., Blischak, P.D., Sutherland, B.L., and Barker, M.S. (2021).
1112 Patterns and Processes of Diploidization in Land Plants. Annu. Rev. Plant Biol. *72*, 387-410.
1113 https://doi.org/10.1146/annurev-arplant-050718-100344.

98. Kauai, F., Mortier, F., Milosavljevic, S., Van de Peer, Y., and Bonte, D. (2023). Neutral processes underlying the macro eco-evolutionary dynamics of mixed-ploidy systems. Proc. R. Soc. B Biol. Sci. *290*, 20222456. doi:10.1098/rspb.2022.2456.
99. Eric Schranz, M., Mohammadin, S., and Edger, P.P. (2012). Ancient whole genome duplications, novelty and diversification: the WGD Radiation Lag-Time Model. Curr. Opin. Plant Biol. *15*, 147-153. https://doi.org/10.1016/j.pbi.2012.03.011.
100. Robertson, F.M., Gundappa, M.K., Grammes, F., Hvidsten, T.R., Redmond, A.K., Lien, S., Martin, S.A.M., Holland, P.W.H., Sandve, S.R., and Macqueen, D.J. (2017). Lineage-specific rediploidization is a mechanism to explain time-lags between genome duplication and evolutionary diversification. Genome Biol. *18*, 111. 10.1186/s13059-017-1241-z.
101. Parey, E., Louis, A., Montfort, J., Guiguen, Y., Crollius, H.R., and Berthelot, C. (2022). An atlas of fish genome evolution reveals delayed rediploidization following the teleost whole-genome duplication. Genome Res. *32*, 1685-1697. 10.1101/gr.276953.122.
102. Soltis, P.S., and Soltis, D.E. (2016). Ancient WGD events as drivers of key innovations in angiosperms. Curr. Opin. Plant Biol. *30*, 159-165. https://doi.org/10.1016/j.pbi.2016.03.015.
103. Van de Peer, Y., Maere, S., and Meyer, A. (2009). The evolutionary significance of ancient genome duplications. Nature Reviews Genetics *10*, 725-732. 10.1038/nrg2600.
104. Fox, D.T., Soltis, D.E., Soltis, P.S., Ashman, T.-L., and Van de Peer, Y. (2020). Polyploidy: A Biological Force From Cells to Ecosystems. Trends Cell Biol. *30*, 688-694. https://doi.org/10.1016/j.tcb.2020.06.006.
105. Chen, H., and Zwaenepoel, A. (2023). Inference of Ancient Polyploidy from Genomic Data. In Polyploidy: Methods and Protocols, Y. Van de Peer, ed. (Springer US), pp. 3-18. 10.1007/978-1-0716-2561-3_1.
106. Goldschmidt, R. (1982). The material basis of evolution (Yale University Press).
107. Theißen, G. (2006). The proper place of hopeful monsters in evolutionary biology. Theory in Biosciences *124*, 349-369. 10.1016/j.thbio.2005.11.002.
108. Theißen, G. (2009). Saltational evolution: hopeful monsters are here to stay. Theory in Biosciences *128*, 43-51. 10.1007/s12064-009-0058-z.
109. Fawcett, J.A., Van de Peer, Y., and Maere, S. (2013). Significance and Biological Consequences of Polyploidization in Land Plant Evolution. In Plant Genome Diversity Volume 2: Physical Structure, Behaviour and Evolution of Plant Genomes, J. Greilhuber, J. Dolezel, and J.F. Wendel, eds. (Springer Vienna), pp. 277-293. 10.1007/978-3-7091-1160-4_17.
110. Mayrose, I., Zhan, S.H., Rothfels, C.J., Magnuson-Ford, K., Barker, M.S., Rieseberg, L.H., and Otto, S.P. (2011). Recently Formed Polyploid Plants Diversify at Lower Rates. Science *333*, 1257-1257. doi:10.1126/science.1207205.
111. Arrigo, N., and Barker, M.S. (2012). Rarely successful polyploids and their legacy in plant genomes. Curr. Opin. Plant Biol. *15*, 140-146. https://doi.org/10.1016/j.pbi.2012.03.010.
112. Hagen, E.R., and Beaulieu, J.M. (2024). New beginnings for dead ends: polyploidy, -SSE models and the dead-end hypothesis. Ann. Bot. *134*, 923-932. 10.1093/aob/mcae143.
113. Peters, S.E., and McClennen, M. (2016). The Paleobiology Database application programming interface. Paleobiology *42*, 1-7. 10.1017/pab.2015.39.
114. Kocsis, Á.T., Reddin, C.J., Alroy, J., and Kiessling, W. (2019). The r package divDyn for quantifying diversity dynamics using fossil sampling data. Methods Ecol. Evol. *10*, 735-743. https://doi.org/10.1111/2041-210X.13161.
115. Friedrich, O., Norris, R.D., and Erbacher, J. (2012). Evolution of middle to Late Cretaceous oceans—A 55 m.y. record of Earth's temperature and carbon cycle. Geology *40*, 107-110. 10.1130/G32701.1.

116. Song, H., Wignall, P.B., Song, H., Dai, X., and Chu, D. (2019). Seawater Temperature and Dissolved Oxygen over the Past 500 Million Years. Journal of Earth Science *30*, 236-243. 10.1007/s12583-018-1002-2.
117. Judd, E.J., Tierney, J.E., Lunt, D.J., Montañez, I.P., Huber, B.T., Wing, S.L., and Valdes, P.J. (2024). A 485-million-year history of Earth's surface temperature. Science *385*, eadk3705. doi:10.1126/science.adk3705.
118. Foster, G.L., Royer, D.L., and Lunt, D.J. (2017). Future climate forcing potentially without precedent in the last 420 million years. Nature Communications *8*, 14845. 10.1038/ncomms14845.
119. Potter, S.C., Luciani, A., Eddy, S.R., Park, Y., Lopez, R., and Finn, R.D. (2018). HMMER web server: 2018 update. Nucleic Acids Res. *46*, W200-W204. 10.1093/nar/gky448.
120. Katoh, K., and Standley, D.M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Molecular Biology and Evolution *30*, 772-780. 10.1093/molbev/mst010.
121. Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2014). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Molecular Biology and Evolution *32*, 268-274. 10.1093/molbev/msu300.
122. Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermiin, L.S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nature Methods *14*, 587-589. 10.1038/nmeth.4285.
123. Zhang, C., Scornavacca, C., Molloy, E.K., and Mirarab, S. (2020). ASTRAL-Pro: Quartet-Based Species-Tree Inference despite Paralogy. Mol. Biol. Evol. *37*, 3292-3307. 10.1093/molbev/msaa139.
124. Yang, Z. (2007). PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol. Biol. Evol. *24*, 1586-1591. 10.1093/molbev/msm088.
125. Proost, S., Fostier, J., De Witte, D., Dhoedt, B., Demeester, P., Van de Peer, Y., and Vandepoele, K. (2011). i-ADHoRe 3.0—fast and sensitive detection of genomic homology in extremely large data sets. Nucleic Acids Research *40*, e11-e11. 10.1093/nar/gkr955.
126. Chen, H., Zwaenepoel, A., and Van de Peer, Y. (2024). wgd v2: a suite of tools to uncover and date ancient polyploidy and whole-genome duplication. Bioinformatics *40*, btae272. 10.1093/bioinformatics/btae272.
127. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods *17*, 261-272. 10.1038/s41592-019-0686-2.
128. Li, L., Stoeckert, C.J., and Roos, D.S. (2003). OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. Genome Res. *13*, 2178-2189. 10.1101/gr.1224503.
129. Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. *20*, 238. 10.1186/s13059-019-1832-y.
130. De Smet, R., Adams, K.L., Vandepoele, K., Van Montagu, M.C.E., Maere, S., and Van de Peer, Y. (2013). Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. Proc. Natl. Acad. Sci. U.S.A. *110*, 2898-2903. doi:10.1073/pnas.1300127110.
131. Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Systematic Biology *59*, 307-321. 10.1093/sysbio/syq010.
132. Wickett, N.J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., Ayyampalayam, S., Barker, M.S., Burleigh, J.G., Gitzendanner, M.A., et al. (2014). Phylotranscriptomic analysis of the origin and early diversification of land plants. Proc. Natl. Acad. Sci. U.S.A. *111*, E4859-E4868. doi:10.1073/pnas.1323926111.

133. Sensalari, C., Maere, S., and Lohaus, R. (2022). ksrates: positioning whole-genome duplications relative to speciation events in KS distributions. Bioinformatics *38*, 530-532. 10.1093/bioinformatics/btab602.
134. Raup, D.M. (1975). Taxonomic diversity estimation using rarefaction. Paleobiology *1*, 333-342. 10.1017/s0094837300002633.
135. Alroy, J. (2010). The Shifting Balance of Diversity Among Major Marine Animal Groups. Science *329*, 1191-1194. doi:10.1126/science.1189910.
136. Foote, M. (1999). Morphological diversity in the evolutionary radiation of Paleozoic and post-Paleozoic crinoids. Paleobiology *25*, 1-115. 10.1017/s0094837300020236.
137. Alroy, J. (2008). Dynamics of origination and extinction in the marine fossil record. Proc. Natl. Acad. Sci. U.S.A. *105*, 11536-11542. doi:10.1073/pnas.0802597105.
138. Alroy, J. (2014). Accurate and precise estimates of origination and extinction rates. Paleobiology *40*, 374-397. 10.1666/13036.
139. Alroy, J. (2015). A more precise speciation and extinction rate estimator. Paleobiology *41*, 633-639. 10.1017/pab.2015.26.
140. Ogg, J.G., Ogg, G.M., and Gradstein, F. (2016). A concise geologic time scale: 2016 (Elsevier).
141. Ojala, M., and Garriga, G.C. (2009). Permutation Tests for Studying Classifier Performance. Proceedings of the 2009 Ninth IEEE International Conference on Data Mining (*IEEE Computer Society*), 908–913.