



Citation for published version:

Clark, JW, Puttick, MN & Donoghue, PCJ 2019, 'Origin of horsetails and the role of whole-genome duplication in plant macroevolution', *Proceedings. Biological sciences*, vol. 286, no. 1914, 20191662, pp. 1-10.
<https://doi.org/10.1098/rspb.2019.1662>

DOI:

[10.1098/rspb.2019.1662](https://doi.org/10.1098/rspb.2019.1662)

Publication date:

2019

Document Version

Peer reviewed version

[Link to publication](#)

Copyright 2019 The Author(s). The final publication is available at Proceedings of Royal Society B via <https://doi.org/10.1098/rspb.2019.1662>

University of Bath

Alternative formats

If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

1 **Title:** Origin of horsetails and the role of whole genome duplication in plant macroevolution

2
3 James W. Clark^{1,2*}, Mark N. Puttick^{2,3} & Philip C.J. Donoghue²

4
5 ¹Department of Plant Sciences, University of Oxford, South Parks Road, OX1 3RB Oxford,
6 United Kingdom.

7 ²School of Earth Sciences University of Bristol, BS8 1TQ Bristol, United Kingdom

8 ³Milner Centre for Evolution, Department of Biology and Biochemistry, University of Bath,
9 BA2 7AY Bath, United Kingdom

10 *Author for correspondence

11
12 **Summary**

13 Whole Genome Duplication (WGD) has occurred commonly in land plant evolution and it is
14 often invoked as a causal agent in diversification, phenotypic and developmental innovation,
15 as well as conferring extinction resistance. The ancient and iconic lineage of *Equisetum* is no
16 exception, where WGD has been inferred to have occurred prior to the Cretaceous-Paleogene
17 (K-Pg) boundary, coincident with WGD events in angiosperms. In the absence of high
18 species diversity, WGD in *Equisetum* is interpreted to have facilitated the long-term survival
19 of the lineage. However, this characterisation remains uncertain as these analyses of the
20 *Equisetum* WGD event have not accounted for fossil diversity. Here we analyse additional
21 available transcriptomes and summarise the fossil record. Our results confirm support for at
22 least one WGD event shared among the majority of extant *Equisetum* species. Furthermore,
23 we use improved dating methods to constrain the age of gene duplication in geological time
24 and identify two successive *Equisetum* WGD events. The two WGD events occurred during
25 the Carboniferous and Triassic, respectively, rather than in association with the K-Pg
26 boundary. WGD events are believed to drive high rates of trait evolution and innovations, but
27 analysed trends of morphological evolution across the historical diversity of *Equisetum*
28 provide little evidence for further macroevolutionary consequences following WGD. WGD
29 events cannot have conferred extinction resistance to the *Equisetum* lineage through the K-Pg
30 boundary since the ploidy events occurred hundreds of millions of years before this mass
31 extinction and we find evidence of extinction among fossil polyploid *Equisetum* lineages.
32 Our findings precipitate the need for a review of the proposed roles of WGDs in biological
33 innovation and extinction survival in angiosperm and non-angiosperm lineages alike.

36 **1. Introduction**

37 The prevalence of Whole Genome Duplication (WGD) in land plants has contributed to the
38 widely held view that WGD is an agent of macroevolutionary change [1]. The most striking
39 pattern to have emerged is the apparent temporal clustering of WGD events about the
40 Cretaceous-Palaeogene (K-Pg) boundary interval [2-4]. Perhaps inevitably, this has led to
41 suggestions that WGD facilitated the survival and success of plant lineages in the wake of the
42 attendant ecological disturbance and mass extinction [5-7]. Further, polyploid formation at
43 mass extinction events is predicted to have been higher, as environmental disturbance and
44 stress led to the formation of unreduced gametes [8, 9]. However, the WGD-K-Pg hypothesis
45 is dependent on the accuracy and precision of estimates for the timing of WGD events.

46 Transcriptomics of *Equisetum giganteum* have revealed that, like many other land
47 plant lineages, *Equisetum* underwent at least one round of WGD [10]. The phylogenetic
48 position of *Equisetum* on a long depauperate branch makes direct molecular dating
49 challenging and hence previous studies have broad confidence intervals around estimated
50 ages. Nevertheless, age estimates from synonymous substitutions (*Ks*) between duplicate
51 gene pairs have been interpreted cautiously to reflect a duplication age overlapping the K-Pg
52 boundary [10].

53 WGD is often proposed as a driver of species diversification [11]. *Equisetum* seems to
54 be an exception, as with only 15 extant species the genus hardly evidences a link between
55 WGD and diversification. In lieu of high species diversity, Vanneste *et al.* [10] have
56 suggested that the WGD event may have contributed to the longevity of the lineage, despite
57 estimating a relatively recent *Equisetum* WGD. WGD is also generally proposed as a driver
58 of phenotypic innovation [12], however, few studies consider the diversity of extinct forms in
59 the context of WGD [13]. This is pertinent to *Equisetum* which exhibits a rich evolutionary
60 history that has been revealed by several recent palaeontological discoveries [14-17].

61 To test the association of *Equisetum* WGD and the K-Pg extinction event, we present
62 a thorough analysis of the timing of WGD within Equisetales and its putative
63 macroevolutionary consequences. We refine the phylogenetic position of putative WGD
64 events and use molecular clock methods to show that WGD occurred well before the K-Pg,
65 closer in age to the more ancient and profound Permian-Triassic extinction event. Further, we
66 show that the WGD is not responsible for the phenotypic distinctiveness of *Equisetum*. There
67 is no evidence that WGD conferred extinction resistance to Equisetales with many Mesozoic
68 lineages not making it through the K-Pg mass extinction.

69

70

71 **2. Materials and Methods**

72 **(a) Transcriptome Assembly**

73 Assembled transcriptomes were collected from the 1KP dataset for *Equisetum diffusum*,
74 *Equisetum hyemale*, *Culcita macrocarpa*, *Ophioglossum petiolatum*, *Tmesipteris parva*,
75 *Selaginella kraussiana*, *Danaea nodosa* and *Botrypus virginianus*, and an additional
76 transcriptome for *Equisetum giganteum* was obtained from [10].

77 Paired end short reads were downloaded from the SRA archive for *Equisetum*
78 *arvensense* (SRR4061754), *Equisetum telmateia* (SRR4061752) and *Equisetum ramosissimum*
79 (SRR5499399), and assembled following [18]. Reads were trimmed of adapter sequences
80 using Trimmomatic v.0.35 [19] using default settings. Assembly was performed using Trinity
81 [20] using default settings. Redundant transcripts were removed using CD-HIT with a cluster
82 value of 0.95 [21]. Each transcript was converted into the single best amino acid sequence
83 using TransDecoder [22]. The assembly of the *E. arvensense*, *E. ramosissimum* and *E. telmateia*
84 transcriptomes after clustering resulted in 24,187, 58,549 and 61,969 transcripts.

85

86 **(b) Ks analysis**

87 We compared rates of synonymous substitution between paralogous genes in *E. hyemale* and
88 *E. diffusum*, that represent the subgenera *Hippochaete* and *Equisetum*, respectively. Analyses
89 were performed using default parameters and the ‘phym1’ node-weighting method in the *wgd*
90 package [23-26]. *Ks* distributions were plotted based on node-averaged values as calculated
91 in the *wgd* package. Gaussian mixture models were fitted to the *Ks* distribution following the
92 *wgd* pipeline, with the optimal number of components assessed using the Bayesian
93 Information Criterion (BIC).

94

95 **(c) Gene family assignment**

96 Orthogroups from the transcriptomes were inferred using Orthofinder v.2.2.6 [27] under a
97 Diamond sequence search. The Orthofinder analysis initially produced 27,038 orthogroups.
98 An initial filtering step was performed to remove orthogroups that did not contain at least one
99 representative from 75% of species. Remaining orthogroups were aligned using MUSCLE
100 and trimmed using trimal [28]. A second filtering step removed all alignments shorter than
101 200 amino acids, resulting in 5,009 orthogroups. Phylogenetic inference was performed on
102 each remaining orthogroup under the best-fitting model and maximum likelihood criterion in
103 IQ-TREE [29], with 1000 ultra-fast bootstrap replicates [30].

104

105 **(d) Species Divergence Time Estimation**

106 Single copy orthogroups from the Orthofinder output formed the basis of a dating analysis.
107 An alignment of 45,977 amino acids was partitioned by gene for a topology search using the
108 edge-linked option (-spp) in IQ-TREE [29].

109 The topology formed the basis of a fixed-topology node-calibrated molecular clock
110 analysis in MCMCtree [24]. Node calibrations were specified with a uniform distribution
111 spanning the hard minimum and soft maximum constraints (with a 2.5% tail distribution)
112 established using *MCMCtreeR* in R (Table 1) [31]. Previous studies have placed the fossil
113 taxon *Equisetum fluviatoides* as sister to *E. diffusum* [17]. However, our analyses supported a
114 *E. fluviatoides* as sister to both *E. diffusum* and *E. arvense*, and so we established a
115 calibration for the divergence of the two subgenera (Supplementary Methods). The mean rate
116 was assigned a gamma prior, determined based on the mean number of substitutions along
117 the tree scaled by the approximate geological age, with a total of 0.12 substitutions per site
118 per million years. To ensure the model sampled from this distribution we fixed the shape
119 parameter to two and adjusted the scale parameter to 16 [32, 33]. The analysis was run
120 without sequence data to ensure that the effective time priors were compatible with the
121 palaeontological and phylogenetic constraints informing the specified node calibrations [34].
122 Using the approximate likelihood method [35], we ran two independent analyses, each for
123 5,000,000 generations, discarding the first 1,000,000 generations as burn-in. Convergence of
124 each run was assessed using Tracer [36].

125

126 **(e) Gene tree and species tree reconciliation**

127 Gene trees inferred from Orthofinder were reconciled with the dated species tree. Gene trees
128 were inferred under a DTL (Duplication, Transfer, Loss) model using a maximum likelihood
129 criterion in ALE (Amalgamated Likelihood Estimation) [37]. The reconciliations were
130 performed using 1000 ultrafast bootstrap replicates as tree samples. As there is no prior
131 hypothesis regarding an ancient hybridization (allopolyploidy) event in *Equisetum*, we set a
132 low prior rate of gene transfer (0.1). The total number of duplications was summed for each
133 branch in the phylogeny based on the number of inferred duplications across each of the 1000
134 sampled trees for each gene family.

135

136 **(f) Dating whole genome duplication**

137 Gene families inferred to have duplicated along the branch leading to *Equisetum* were
138 sampled from the ALE output (Supplementary Fig S1). To evaluate the hypothesis of a single
139 WGD event in *Equisetum*, we selected gene families that contained a single duplication along
140 this branch for a molecular clock analysis. Following [38], gene families were used if they:
141 (i) had a clear topological signal of the WGD event, represented by two paralogous copies
142 present in all *Equisetum* species forming two monophyletic groupings; (ii) had a topology
143 congruent with current understanding of tracheophyte phylogeny; and (iii) did not have a
144 signal of additional duplication events within *Equisetum*. We conducted a molecular clock
145 analysis for each gene family with the same settings as used for the species divergence
146 estimation. The 95% Highest Posterior Densities (HPDs) were combined between all gene
147 families. Peaks in this combined posterior distribution may represent duplication events
148 common to multiple gene families. To determine which gene families coincide with each
149 peak, the peaks in the combined posterior distribution were described using Gaussian mixture
150 models (GMMs) and the overlap between these peaks and the individual gene posterior
151 distributions were estimated using an overlapping coefficient [39]. Gene families with an
152 overlap > 0.8 for each respective peak were selected and concatenated. Molecular clock
153 analyses were performed for families corresponding to each peak, with the same set of fossil
154 calibrations employed as in the species divergence time estimation, with the exception that
155 the calibration within *Equisetum* was cross-calibrated on both sides of the duplication.
156 Analyses were performed as for the species divergence estimation.

157 To consider the possibility of multiple WGD events, we repeated the analysis with
158 gene families containing at least two duplications (four copies of each gene) in all extant
159 *Equisetum* species, allowing for simultaneous age estimation of two duplication nodes.

160

161 **(g) Dating of Fossils and Extant Taxa**

162 We used previously assembled phenotypic and molecular matrices of 77 binary and
163 multistate phenotypic characters and the *rbcL*, *atpA*, *atpB* and *matK* chloroplast genes [17].
164 The matrix contained 49 taxa, including 17 extant and 32 fossil taxa spanning the
165 Sphenophyllales + Equisetales as well as outgroup taxa *Hamatophyton verticillatum*,
166 *Rotafolia songziensis*, *Ophioglossum reticulatum* (Ophioglossales) and *Psilotum nudum*
167 (Psilotales).

168 We estimated divergence times using the estimates obtained from the molecular
169 species divergence analysis as priors on nodes present in this dataset. Fossil tip ages were
170 based on a uniform distribution across their occurrence ranges (Supplementary Table 1) and a

171 uniform distribution was placed on the root between 451-384 million years [33]. A stepping
172 stone analysis was used to test for the best-fitting clock model in MrBayes v.3.2.6 [40, 41];
173 this showed significant support for the correlated model [42] over the Independent Gamma
174 Rates [43] and strict clock models. A correlated rates clock model [42] was implemented
175 with the clock rate prior set as a lognormal distribution; the mean of the lognormal
176 distribution was estimated from a topological analysis to estimate the tree height scaled by
177 the approximate geological age of the root (0.02 substitutions site⁻¹ million years⁻¹) [44].
178 Finally, we set a uniform birth-death prior across the tree [41]. The phenotypic data and each
179 gene were partitioned separately, with molecular data analysed under the GTR+ Γ model and
180 the phenotypic data under the MKv+ Γ model [45]. Four independent chains were run for
181 20,000,000 generations. Convergence between the chains was assessed based on the average
182 standard deviation of split frequencies (< 0.01), Effective Sample Size (target > 200) and by
183 examining the parameters of the chain in Tracer [36].

184

185 **(h) Rates of Phenotypic Evolution**

186 To examine the rates of phenotypic evolution across the tree, we performed a morphological
187 clock analysis using only the phenotypic dataset with the tree constrained to the topology
188 resolved by the combined analysis. A relaxed clock model was used, allowing rates to vary
189 between branches.

190 The rate of phenotypic evolution was estimated by sampling the effective branch
191 lengths from 1000 points of the posterior distribution; the mean rates were estimated from
192 these samples. Only branches from the majority-rule consensus topology were considered for
193 further analyses; from the 1000 posterior samples, rates were summarised for branches on the
194 posterior tree that matched branches on the majority-rule consensus tree.

195

196 **(i) Phenotypic Disparity**

197 The phenotype matrix was recoded following [46], such that non-applicable (NA) states were
198 coded as '0' and missing data as '?', to distinguish the two types of 'missing data' [47]. The
199 distance between taxa was calculated using Gower's dissimilarity metric [48]. The distances
200 were projected into two-dimensional space using Non-metric Multi-Dimensional Scaling
201 (NMDS). We plotted a phylomorphospace using the majority-rule (50%) consensus tree from
202 the total evidence analysis [49]. The most likely ancestral state was reconstructed along the
203 tree by summarising states across 1000 stochastic character maps [50]; the estimated states
204 were used to position the nodes within the morphospace.

205 We calculated mean disparity as Sum Of Variances from the distance matrix [51]
206 using *dispRity* in R [52]. Disparity through time was estimated using the time-slicing
207 approach using 10 bins and the ‘gradual split’ model as implemented in *dispRity*, with the
208 probability of a character state being that of either the descendent or the ancestor dependent
209 on the length of the branch [52].

210

211 (j) Genome Size Analysis

212 Genome size estimates (1C-values) were downloaded from the c-value database [53]. The
213 1C-values were estimated for fossil taxa by Franks *et al.* [54] who derived a linear regression
214 model for the relationship between 1C-value and stomata guard cell length. They estimated
215 1C-value for members of Sphenophyllales (*Sphenophyllum*) and Calamitaceae
216 (*Calamocladus*) as well as *Equisetum haukeanum*. For this analysis we took the values for
217 Sphenophyllales and Calamitaceae to be representative of each lineage. We used the linear
218 model ($y = 1.83x + -5.46$) to convert the logged guard cell widths of other fossil *Equisetum*
219 and to a logged 1C-value [14-16, 54, 55]. In total, 21 1C-values were obtained
220 (Supplementary Table 1) and were analysed as continuous characters in BayesTraits v.3 [56]
221 using a homogeneous continuous random walk model and the ancestral 1C-values were
222 estimated at internal nodes. The MCMC was run for 15,000,000 generations, with the first
223 10,000,000 generations discarded as burn-in.

224

225 3. Results

226 (a) Transcriptomic Analyses Reveal Triassic and Carboniferous WGD Events

227 The distribution of *Ks* values in *E. hyemale* and *E. diffusum* exhibit at least 3 conspicuous
228 peaks: one close to 0.1 representing recent duplicates, another with a mean close to 1, and
229 third more ancient peak close to 2 (Fig 1). Mixture modelling supported 4 components, but
230 the fourth component had a low mean weight (Fig 1, Supplementary Fig S1). Coincidence of
231 these peaks suggests that the WGD event initially identified in *E. giganteum* is shared
232 between both subgenera, though *Ks* values >2 are increasingly unreliable predictors of WGD
233 [57].

234 ALE analysis revealed rates of duplication that were generally higher on terminal
235 branches (likely due to recent local duplication events) and some of the long branches
236 included in the study. Among all branches, however, ALE provided strong support for a
237 duplication event on the branch leading to *Equisetum* (Supplementary Fig S2). 240 gene
238 families were selected from the ALE output that showed a clear signal of the duplication

239 event. Molecular clock analyses of these gene families supported two clear clusters of ages
240 (Fig 2). For each cluster, we found 52 and 51 corresponding gene families that were
241 concatenated to form alignments of 21,894 and 19,360 amino acids. These analyses
242 suggested a first duplication within the interval 329-307 Ma (Serpukhovian-Moscovian: mid-
243 late Carboniferous) and a second within 253-233 Ma (Changhsingian-Carnian: latest Permian
244 to Late Triassic) (Fig 3).

245 We identified a further 14 gene families with a clear signal of two successive
246 duplications with all 4 paralogs retained. The two successive duplications were estimated to
247 360-322 Ma (Fammenian-Bashkirian: latest Devonian to mid Carboniferous) and 261-211
248 Ma (Capitanian-Norian: late Permian to Late Triassic; Supplementary Fig S3).

249

250 **(b) An Evolutionary Framework: Triassic-Jurassic origin of total-group *Equisetum***

251 Analysis of the combined molecular and morphological dataset partially resolved the
252 backbone phylogeny of Equisetales (Fig 4). Monophyly of Equisetales is strongly supported,
253 with Neocalamitaceae as sister to all remaining Equisetaceae, but there is only weak support
254 for Neocalamitaceae. As with [17], we resolve *Equisetites arenaceus* and *Spaciinodum*
255 *collinsonii* as sister to the total group *Equisetum*.

256 Relationships within *Equisetum* are poorly resolved; the two subgenera (*Equisetum*
257 and *Hippochaete*) are well supported, as are the positions of *E. clarnoi* and *E. fluviatoides*
258 within each, respectively. The relationships of the outgroups are also poorly resolved,
259 including the order of divergence of Archaeocalamitaceae and Calamitaceae, although as we
260 confirm that Equisetaceae did not originate from within Calamitaceae.

261 We estimate a Devonian origin of both sphenopsids and ferns. Sphenophyllales and
262 Equisetales diverged during the Carboniferous along with most of the extinct lineages of
263 Equisetales, including the Archaeocalamitaceae and Calamitaceae. Equisetaceae and
264 Neocalamitaceae diverged during the Permian. We report a Triassic-Jurassic origin of total
265 group *Equisetum*, but a Cretaceous origin of the crown-group, with both extant subgenera
266 originating during the Palaeogene (Supplementary Fig S4).

267

268 **(c) High Rates of Phenotypic Evolution at The Origin of Major Clades**

269 Rates of phenotypic evolution are heterogeneous across the tree (Fig 4). The origin of major
270 lineages is marked by the fastest rates of phenotypic evolution, including Equisetales,
271 Equisetaceae and *Hippochaete* (Fig 4). Generally, phenotypic evolution is much greater
272 between higher-order lineages than within them, with slow rates observed within

273 Equiseteceae and most lineages within Calamitaceae, except the branch leading to
274 *Cruciaetheca*.

275 High rates of phenotypic evolution correspond to large distances in morphospace (Fig
276 5a). Major lineages cluster tightly within morphospace across both axes, though on the
277 individual axes there is considerable overlap. The proportion of total disparity represented by
278 extant taxa is low (Fig 5b) and disparity through time analyses show that modern levels of
279 disparity are a small fraction of a Carboniferous acme (Fig 5c). Mean disparity, measured as
280 the average Euclidean pairwise distance between taxa, is lower in Equisetaceae (0.195) than
281 Calamitaceae (0.381), but they do occupy a novel region of morphospace.

282

283 (d) Genome Duplication and Genome Size

284 Reconstruction of ancestral genome size within Sphenopsida reveals that the largest genome
285 sizes are found within extant *Equisetum* (mean ancestral 1C-value = 17.09pg), in particular
286 the subgenus *Hippochaete* (ancestral 1C-value = 20.9pg) (Fig 6). Across nodes, we observed
287 three large increases in genome size: from the base of *Equisetum* to *Hippochaete* (17.6pg to
288 20.9pg), from the base of Equisetales to total group *Equisetum* (3.9pg to 11.01pg), and from
289 total group to crown group *Equisetum* (11.01 to 17.6pg) (Fig 6).

290

291 4. Discussion

292 (a) Duplication and Evolution in *Equisetum*

293 The WGD shared by extant *Equisetum* was previously proposed as one of several WGD
294 events that coincide with the K-Pg boundary [2, 10]. The significance of this clustering of
295 events has been explored from various angles: that WGD confers an ‘extinction resistance’,
296 that WGD may have provided a means of rapid adaptation amidst ecological disturbance, that
297 WGD may be a response to environmental stresses, and that WGD itself might just be a non-
298 selective consequence [58] of a switch to vegetative reproduction often associated with
299 polyploidy [2, 59, 60]. The new age estimates presented here render these hypotheses
300 unlikely given that the WGDs predate the K-Pg mass extinction by hundreds of millions of
301 years. Indeed, we find no evidence of beneficial evolutionary consequences of WGD in
302 *Equisetum*, suggesting that these events do not universally precipitate changes on the
303 macroevolutionary scale across the tree of life.

304 Our analyses supported multiple bursts of gene duplication throughout the evolution
305 of the *Equisetum* lineage. Their interpretation as WGD events can be difficult [61], yet their
306 clustering within time and the repeated history of WGD across land plants suggests that there

307 is a high probability that they represent WGD events. Though congruent with the findings of
308 Vanneste *et al.* [10], we have better resolved the phylogenetic position of these putative
309 WGD events and find that they are likely shared by both subgenera of *Equisetum* (Fig 1).
310 However, the WGD event proposed by Vanneste *et al.* [10] to have occurred in *E. giganteum*
311 was known only from a single transcriptome and the geological age was difficult to constrain
312 using both phylogenomic and *Ks* methods. Indeed, ages inferred directly from *Ks*
313 distributions can be inaccurate due to sequence saturation and the assumption of a strict clock
314 [57, 62].

315 Using phylogenomic and molecular clock methods, we estimated both events to have
316 occurred long before the K-Pg boundary. Rather, these WGD events are among the most
317 ancient detected in land plants, occurring within the latest Devonian-mid Carboniferous and
318 late Permian-Late Triassic, respectively (Fig 3). This estimate is comparable in precision to
319 recent estimates for other WGD events associated with the K-Pg boundary [63] and serves to
320 highlight the power of these methods to constrain the timing of the event to within 20 million
321 years, along one of the most isolated branches within living land plants. The discrepancy in
322 age for the *Equisetum* WGD events reported here and by Vanneste *et al.* [10] may be due to
323 the initial paucity of transcriptomic data representative of the lineage and highlights the
324 benefits of increased taxonomic sampling and the value of concatenation in estimating the
325 timing of WGD events [1].

326 We reconstructed the evolutionary history of Equisetales using a combination of
327 molecular and phenotype data in a Bayesian framework (Fig 4). Broadly, the relationships
328 resolved are congruent with previous parsimony-based results [17], though the species
329 relationships are less well resolved. The lack of resolution in the phylogeny here may be the
330 consequence of the previously-used parsimony methods producing more highly-resolved, but
331 less accurate trees compared to Bayesian analyses of morphological data [64, 65].
332 Nevertheless, our results corroborate the distinction between the Calamitaceae and
333 Equisetaceae and the hypothesis that both lineages have evolved independently since the
334 Carboniferous (Fig 4).

335 Crucially, these analyses provide a framework in which WGD can be considered in
336 light of both extant and extinct diversity. We have shown that the more ancient WGD event
337 took place prior to the divergence of Equisetaceae and Neocalamitaceae, and the more recent
338 WGD event appears to coincide with the origin of Equisetaceae, either prior to, or after the
339 divergence of *Spaciinodum*. As well as a establishing a more precise estimate for the timing

340 of WGD, our analyses place WGD within the context of the gross historical diversity of the
341 lineage, rather than merely the net diversity that has survived to the present. This represents a
342 novel approach to understanding the role of WGD in land plant evolution that is likely to be
343 key to more thoroughly testing existing hypotheses, such as the proposed link between WGD
344 events and the K-Pg mass extinction event in angiosperm evolution.

345

346 **(b) Evolutionary consequences of WGD in a non-angiosperm lineage**

347 The ancient timing of the *Equisetum* WGD events could be interpreted to strengthen the
348 hypothesis that WGD has facilitated the longevity of the lineage [10]. The tentative
349 hypothesis that the *Equisetum* WGD event conferred extinction resistance across the K-Pg
350 seems unlikely given our estimates for the timing of the WGD events, and current hypotheses
351 linking WGD to success emphasize only short-term advantages. Furthermore, our analyses
352 have shown that many polyploid taxa descended from the WGD events are now extinct.

353 WGD events have also been implicated as drivers of phenotypic variance within the
354 plant kingdom. Multiple models and a few examples demonstrate how novel traits have
355 arisen in the wake of WGD that have been maintained and diversified on a
356 macroevolutionary scale [12, 66]. The precise estimates that we have obtained for the timing
357 of the WGD events allow us to constrain them within tight bounds on the species phylogeny
358 and to consider their impact within the context of subsequent phenotypic evolution. The
359 evolution of Equisetales is generally associated with relative stability and few character state
360 changes, yet the first WGD event coincides with higher rates of phenotypic evolution (Fig 4)
361 and each WGD event also coincides topologically with a movement into a novel area of
362 morphospace (Fig 5a).

363 However, extant *Equisetum* and the fossil taxa that descended from the WGD event
364 represent only a fraction of the phenotypic diversity of Equisetales (Fig 5b). In addition, both
365 Equisetales and Calamitaceae exhibit fast early rates of phenotypic evolution (Fig 4);
366 Calamitaceae also achieved greater disparity (Fig 3a). Indeed, while WGD may have played a
367 role in promoting phenotypic novelty, it has not been sufficient to sustain disparity over time
368 (Fig 3c). Based on previously identified synapomorphies [17], the first WGD event coincides
369 with the evolution of lacunae (vallecular canals), the loss of internode differentiation,
370 alternating sporangiophore shields, an increase in sporangium numbers and, possibly, the
371 expression of all three reproductive regulatory modules [17]. The second WGD also
372 coincides with a number of synapomorphies, including alternating ribs, leaf tips, and a
373 reduction in the length of reproductive structures [17]. Throughout the evolutionary history of

374 Equisetales, the accumulation and transformation of characters associated with the extant
375 taxa is gradual and many of the distinguishing features, including a compacted strobilus and
376 small size, have evolved slowly and in a mosaic pattern over several nodes [17, 67, 68]. This
377 suggests that while WGD may have had a role in promoting the diversity of the Equisetaceae,
378 it was not a prerequisite to the evolution of disparity within Equisetales.

379

380 **(c) Genome size correlates with WGD in *Equisetum***

381 Genome size evolution within Equisetales shows that the inferred WGD events may
382 also correlate with an increase in ancestral genome size (Fig 6). This is in some ways
383 surprising since the signal of genome duplication in genome size estimates rapidly erodes
384 across most plant genomes [69, 70]. However, there is also a more recent shift towards much
385 larger genomes that does not appear to be associated with a WGD event (Fig 6). As there are
386 no extant members of Calamitaceae it is not possible to rule out the possibility that they may
387 have undergone their own independent WGD event. However, the small genome size inferred
388 for Calamitaceae [54] and relative stasis of fern genome evolution means that we may
389 speculate that there may have been no further WGD events in this lineage [71]. Multiple
390 WGD events may in part explain the fixed high chromosome numbers shared among extant
391 species of *Equisetum* [71], yet does not appear to explain the distribution of genome sizes
392 between the two extant subgenera.

393 Clearly, to elucidate a macroevolutionary role for WGD in land plant evolution, it is
394 insufficient to consider only extant taxa. *Equisetum* is a good example, since its extant
395 diversity is a poor representation of the taxonomic and phenotypic diversity that existed
396 historically within Sphenopsida. Here, we suggest that a combination of palaeontological and
397 genomic approaches provides additional power and greater insight when considering the
398 impact of ancient or ‘palaeo’-polyploidy.

399

400 **5. Conclusions**

401 It is generally accepted that WGD events are agents of macroevolutionary change. Here, we
402 have shown that a combination of macroevolutionary and comparative genomic approaches
403 can be used to improve estimates of the timing and characterise outcomes of WGD. In
404 *Equisetum*, WGD did not coincide with the K-Pg boundary, nor does it appear to have
405 facilitated greater resistance to extinction. Rather, while WGD in *Equisetum* appears to
406 correlate with the occupation of novel regions of morphospace, it has not led to significant
407 morphological diversification. The formative role of WGD in the evolutionary history of

408 many angiosperm lineages is generally accepted, yet its role remains to be explored in many
409 other plant lineages where rates of WGD are expected to be high. It is possible that differing
410 genome dynamics may determine equally different roles for WGD in macroevolution.

411

412 **Acknowledgements**

413 The authors thank members of the Bristol Palaeobiology Group, Jill Harrison, and Andrew
414 Leitch for helpful discussion. JC was funded by a BBSRC SWBIO DTP studentship, MNP
415 by an 1851 Research Fellowship from the Royal Commission for the Exhibition of 1851,
416 PCJD by NERC (NE/N002067/1) and BBSRC (BB/N000919/1).

417

418

419 **References**

420

- 421 1. Clark J.W., Donoghue P.C.J. 2018 Whole-Genome Duplication and Plant
422 Macroevolution. *Trends in Plant Science*. (doi:10.1016/j.tplants.2018.07.006).
- 423 2. Lohaus R., Van de Peer Y. 2016 Of dups and dinos: evolution at the K/Pg boundary.
424 *Current Opinion in Plant Biology* **30**, 62-69. (doi:10.1016/j.pbi.2016.01.006).
- 425 3. Vanneste K., Maere S., Van de Peer Y. 2014 Tangled up in two: a burst of genome
426 duplications at the end of the Cretaceous and the consequences for plant evolution.
427 *Philosophical Transactions of the Royal Society B: Biological Sciences* **369**(1648).
428 (doi:10.1098/rstb.2013.0353).
- 429 4. Koenen E.J., Ojeda D.I., Steeves R., Migliore J., Bakker F.T., Wieringa J.J., Kidner
430 C., Hardy O., Pennington R.T., Herendeen P.S. 2019 The Origin and Early Evolution of the
431 Legumes are a Complex Paleopolyploid Phylogenomic Tangle closely associated with the
432 Cretaceous-Paleogene (K-Pg) Boundary. *bioRxiv*, 577957.
- 433 5. Renne P.R., Sprain C.J., Richards M.A., Self S., Vanderkluysen L., Pande K. 2015
434 State shift in Deccan volcanism at the Cretaceous-Paleogene boundary, possibly induced by
435 impact. *Science* **350**(6256), 76-78.
- 436 6. Wilf P., Johnson K.R. 2004 Land plant extinction at the end of the Cretaceous: a
437 quantitative analysis of the North Dakota megafloral record. *Paleobiology* **30**(3), 347-368.
- 438 7. Sessa E.B. 2019 Polyploidy as a mechanism for surviving global change. *New*
439 *Phytologist* **221**(1), 5-6.
- 440 8. Kurschner W.M., Batenburg S.J., Mander L. 2013 Aberrant Classopollis pollen
441 reveals evidence for unreduced (2n) pollen in the conifer family Cheirolepidiaceae during the

442 Triassic-Jurassic transition. *Proc Biol Sci* **280**(1768), 20131708.
443 (doi:10.1098/rspb.2013.1708).

444 9. Visscher H., Looy C.V., Collinson M.E., Brinkhuis H., van Konijnenburg-van Cittert
445 J.H., Kurschner W.M., Sephton M.A. 2004 Environmental mutagenesis during the end-
446 Permian ecological crisis. *Proc Natl Acad Sci U S A* **101**(35), 12952-12956.
447 (doi:10.1073/pnas.0404472101).

448 10. Vanneste K., Sterck L., Myburg A.A., Van de Peer Y., Mizrachi E. 2015 Horsetails
449 Are Ancient Polyploids: Evidence from *Equisetum giganteum*. *Plant Cell* **27**(6), 1567-1578.
450 (doi:10.1105/tpc.15.00157).

451 11. Landis J.B., Soltis D.E., Li Z., Marx H.E., Barker M.S., Tank D.C., Soltis P.S. 2018
452 Impact of whole-genome duplication events on diversification rates in angiosperms.
453 *American Journal of Botany* **105**(3), 348-363. (doi:10.1002/ajb2.1060).

454 12. Moriyama Y., Koshiha-Takeuchi K. 2018 Significance of whole-genome duplications
455 on the emergence of evolutionary novelties. *Briefings in functional genomics*.

456 13. Donoghue P.C., Purnell M.A. 2005 Genome duplication, extinction and vertebrate
457 evolution. *Trends in Ecology and Evolution* **20**(6), 312-319. (doi:10.1016/j.tree.2005.04.008).

458 14. Stanich N.A., Rothwell G.W., Stockey R.A. 2009 Phylogenetic diversification of
459 *Equisetum* (Equisetales) as inferred from Lower Cretaceous species of British Columbia,
460 Canada. *American Journal of Botany* **96**(7), 1289-1299. (doi:10.3732/ajb.0800381).

461 15. Channing A., Zamuner A., Edwards D., Guido D. 2011 *Equisetum thermale* sp. nov.
462 (Equisetales) from the Jurassic San Agustín hot spring deposit, Patagonia: Anatomy,
463 paleoecology, and inferred paleoecophysiology. *American Journal of Botany* **98**(4), 680-697.
464 (doi:10.3732/ajb.1000211).

465 16. Elgorriaga A., Escapa I.H., Bomfleur B., Cúneo R., Ottone E.G. 2015 Reconstruction
466 and Phylogenetic Significance of a New *Equisetum* Linnaeus Species from the Lower
467 Jurassic of Cerro Bayo (Chubut Province, Argentina). *Ameghiniana* **52**(1), 135-152.
468 (doi:10.5710/AMGH.15.09.2014.2758).

469 17. Elgorriaga A., Escapa I.H., Rothwell G.W., Tomescu A.M.F., Rubén Cúneo N. 2018
470 Origin of *Equisetum*: Evolution of horsetails (Equisetales) within the major euphyllophyte
471 clade Sphenopsida. *American Journal of Botany* **105**(8), 1286-1303. (doi:10.1002/ajb2.1125).

472 18. Carruthers M., Yurchenko A.A., Augley J.J., Adams C.E., Herzyk P., Elmer K.R.
473 2018 De novo transcriptome assembly, annotation and comparison of four ecological and
474 evolutionary model salmonid fish species. *BMC Genomics* **19**(1), 32. (doi:10.1186/s12864-
475 017-4379-x).

- 476 19. Bolger A.M., Lohse M., Usadel B. 2014 Trimmomatic: a flexible trimmer for
477 Illumina sequence data. *Bioinformatics* **30**(15), 2114-2120.
- 478 20. Grabherr M.G., Haas B.J., Yassour M., Levin J.Z., Thompson D.A., Amit I., Adiconis
479 X., Fan L., Raychowdhury R., Zeng Q. 2011 Full-length transcriptome assembly from RNA-
480 Seq data without a reference genome. *Nature Biotechnology* **29**. (doi:10.1038/nbt.1883).
- 481 21. Fu L., Niu B., Zhu Z., Wu S., Li W. 2012 CD-HIT: accelerated for clustering the
482 next-generation sequencing data. *Bioinformatics* **28**(23), 3150-3152.
- 483 22. Haas B., Papanicolaou A. 2012 Transdecoder. (
- 484 23. Zwaenepoel A., Van de Peer Y. 2019 wgd: simple command line tools for the
485 analysis of ancient whole genome duplications. *Bioinformatics*.
- 486 24. Yang Z. 2007 PAML 4: phylogenetic analysis by maximum likelihood. *Molecular*
487 *Biology and Evolution* **24**(8), 1586-1591. (doi:10.1093/molbev/msm088).
- 488 25. Edgar R.C. 2004 MUSCLE: multiple sequence alignment with high accuracy and
489 high throughput. *Nucleic Acids Research* **32**. (doi:10.1093/nar/gkh340).
- 490 26. Guindon S., Dufayard J.-F., Lefort V., Anisimova M., Hordijk W., Gascuel O. 2010
491 New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the
492 performance of PhyML 3.0. *Systematic biology* **59**(3), 307-321.
- 493 27. Emms D.M., Kelly S. 2015 OrthoFinder: solving fundamental biases in whole
494 genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*
495 **16**(1), 157. (doi:10.1186/s13059-015-0721-2).
- 496 28. Capella-Gutierrez S., Silla-Martinez J.M., Gabaldon T. 2009 trimAl: a tool for
497 automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**(15),
498 1972-1973. (doi:10.1093/bioinformatics/btp348).
- 499 29. Nguyen L.-T., Schmidt H.A., von Haeseler A., Minh B.Q. 2015 IQ-TREE: A Fast and
500 Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular*
501 *Biology and Evolution* **32**(1), 268-274. (doi:10.1093/molbev/msu300).
- 502 30. Hoang D.T., Chernomor O., von Haeseler A., Minh B.Q., Vinh L.S. 2018 UFBoot2:
503 Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution* **35**(2),
504 518-522. (doi:10.1093/molbev/msx281).
- 505 31. Puttick M.N. 2019 MCMCTreeR: functions to prepare MCMCtree analyses and
506 visualise posterior ages on trees. *Bioinformatics*. (doi:10.1093/bioinformatics/btz554).
- 507 32. dos Reis M., Donoghue P.C.J., Yang Z. 2016 Bayesian molecular clock dating of
508 species divergences in the genomics era. *Nature Reviews Genetics* **17**(2), 71-80.
509 (doi:10.1038/nrg.2015.8).

- 510 33. Morris J.L., Puttick M.N., Clark J.W., Edwards D., Kenrick P., Pressel S., Wellman
511 C.H., Yang Z., Schneider H., Donoghue P.C. 2018 The timescale of early land plant
512 evolution. *Proceedings of the National Academy of Sciences* **115**(10), E2274-E2283.
- 513 34. Inoue J.G., Donoghue P.C.J., Yang Z. 2010 The impact of the representation of fossil
514 calibrations on bayesian estimation of species divergence times. *Systematic Biology* **59**(1),
515 74-89.
- 516 35. dos Reis M., Yang Z. 2011 Approximate likelihood calculation on a phylogeny for
517 Bayesian estimation of divergence times. *Molecular Biology and Evolution* **28**(7), 2161-
518 2172. (doi:10.1093/molbev/msr045).
- 519 36. Rambaut A., Suchard M., Drummond A.J. 2014 Tracer v1. 6. (
- 520 37. Szöllősi G.J., Boussau B., Abby S.S., Tannier E., Daubin V. 2012 Phylogenetic
521 modeling of lateral gene transfer reconstructs the pattern and relative timing of speciations.
522 *Proceedings of the National Academy of Sciences* **109**(43), 17513-17518.
523 (doi:10.1073/pnas.1202997109).
- 524 38. Clark J.W., Donoghue P.C. 2017 Constraining the timing of whole genome
525 duplication in plant evolutionary history. *Proceedings of the Royal Society B: Biological*
526 *Sciences* **284**(1858), 20170912.
- 527 39. Inman H.F., Bradley E.L. 1989 The overlapping coefficient as a measure of
528 agreement between probability distributions and point estimation of the overlap of two
529 normal densities. *Communications in Statistics - Theory and Methods* **18**(10), 3851-3874.
530 (doi:10.1080/03610928908830127).
- 531 40. Ronquist F., Huelsenbeck J.P. 2003 MrBayes 3: Bayesian phylogenetic inference
532 under mixed models. *Bioinformatics* **19**. (doi:10.1093/bioinformatics/btg180).
- 533 41. Ronquist F., Klopfstein S., Vilhelmsen L., Schulmeister S., Murray D.L., Rasnitsyn
534 A.P. 2012 A total-evidence approach to dating with fossils, applied to the early radiation of
535 the Hymenoptera. *Systematic Biology* **61**(6), 973-999.
- 536 42. Thorne J.L., Kishino H. 2002 Divergence time and evolutionary rate estimation with
537 multilocus data. *Systematic Biology* **51**, 689-702.
- 538 43. Lepage T., Bryant D., Philippe H., Lartillot N. 2007 A general comparison of relaxed
539 molecular clock models. *Molecular Biology and Evolution* **24**(12), 2669-2680.
- 540 44. Dos Reis M., Zhu T., Yang Z. 2014 The impact of the rate prior on Bayesian
541 estimation of divergence times with multiple loci. *Systematic biology* **63**(4), 555-565.
- 542 45. Lewis P.O. 2001 A likelihood approach to estimating phylogeny from discrete
543 morphological character data. *Systematic biology* **50**(6), 913-925.

- 544 46. Deline B., Greenwood J.M., Clark J.W., Puttick M.N., Peterson K.J., Donoghue
545 P.C.J. 2018 Evolution of metazoan morphological disparity. *Proceedings of the National*
546 *Academy of Sciences*.
- 547 47. Deline B. 2009 The effects of rarity and abundance distributions on measurements of
548 local morphological disparity. *Paleobiology* **35**(2), 175-189. (doi:10.1666/08028.1).
- 549 48. Gower J.C. 1971 A General Coefficient of Similarity and Some of Its Properties.
550 *Biometrics* **27**(4), 857-871. (doi:10.2307/2528823).
- 551 49. Revell L.J. 2012 phytools: An R package for phylogenetic comparative biology (and
552 other things). *Methods in Ecology and Evolution* **3**, 217-223.
- 553 50. Huelsenbeck J.P., Nielsen R., Bollback J.P. 2003 Stochastic mapping of
554 morphological characters. *Systematic Biology* **52**(2), 131-158.
- 555 51. Chartier M., Löfstrand S., von Balthazar M., Gerber S., Jabbour F., Sauquet H.,
556 Schönenberger J. 2017 How (much) do flowers vary? Unbalanced disparity among flower
557 functional modules and a mosaic pattern of morphospace occupation in the order Ericales.
558 *Proceedings of the Royal Society B: Biological Sciences* **284**(1852), 20170066.
559 (doi:10.1098/rspb.2017.0066).
- 560 52. Guillerme T., Cooper N., Smith A. 2018 Time for a rethink: time sub-sampling
561 methods in disparity-through-time analyses. *Palaeontology* **0**(0). (doi:10.1111/pala.12364).
- 562 53. Bennett M., Leitch I. 2012 Plant DNA C-values database. (Royal Botanic Gardens
563 Kew.
- 564 54. Franks P.J., Freckleton R.P., Beaulieu J.M., Leitch I.J., Beerling D.J. 2012
565 Megacycles of atmospheric carbon dioxide concentration correlate with fossil plant genome
566 size. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**(1588), 556.
- 567 55. Gould R.E. 1968 Morphology of *Equisetum laterale* Phillips, 1829, and *E. bryanii* sp.
568 nov. from the Mesozoic of South-Eastern Queensland. *Australian Journal of Botany* **16**(1),
569 153-176.
- 570 56. Pagel M. 1999 Inferring the historical patterns of biological evolution. *Nature* **401**,
571 877. (doi:10.1038/44766).
- 572 57. Vanneste K., Van de Peer Y., Maere S. 2013 Inference of genome duplications from
573 age distributions revisited. *Molecular Biology and Evolution* **30**(1), 177-190.
574 (doi:10.1093/molbev/mss214).
- 575 58. Gould S.J., Lewontin R.C. 1979 The spandrels of San Marco and the Panglossian
576 paradigm: a critique of the adaptationist programme. *Proc R Soc Lond B* **205**(1161), 581-598.

- 577 59. Freeling M. 2017 The distribution of ancient polyploidies in the plant phylogenetic
578 tree is a spandrel of occasional sex. *The Plant Cell*.
- 579 60. Levin D.A., Soltis D.E. 2018 Factors promoting polyploid persistence and
580 diversification and limiting diploid speciation during the K–Pg interlude. *Current opinion in*
581 *plant biology* **42**, 1-7.
- 582 61. Nakatani Y., McLysaght A. 2019 Macrosynteny analysis shows the absence of
583 ancient whole-genome duplication in lepidopteran insects. *Proceedings of the National*
584 *Academy of Sciences* **116**(6), 1816-1818.
- 585 62. Doyle J.J., Egan A.N. 2010 Dating the origins of polyploidy events. *New Phytologist*
586 **186**(1), 73-85. (doi:10.1111/j.1469-8137.2009.03118.x).
- 587 63. Schwager E.E., Sharma P.P., Clarke T., Leite D.J., Wierschin T., Pechmann M.,
588 Akiyama-Oda Y., Esposito L., Bechsgaard J., Bilde T., et al. 2017 The house spider genome
589 reveals an ancient whole-genome duplication during arachnid evolution. *BMC Biology* **15**(1),
590 62. (doi:10.1186/s12915-017-0399-x).
- 591 64. O'Reilly J.E., Puttick M.N., Pisani D., Donoghue P.C.J. 2017 Probabilistic methods
592 surpass parsimony when assessing clade support in phylogenetic analyses of discrete
593 morphological data. *Palaeontology* **61**(1), 105-118. (doi:10.1111/pala.12330).
- 594 65. Puttick M.N., Reilly J.E., Tanner A.R., Fleming J.F., Clark J., Holloway L., Lozano-
595 Fernandez J., Parry L.A., Tarver J.E., Pisani D., et al. 2017 Uncertain-tree: discriminating
596 among competing approaches to the phylogenetic analysis of phenotype data. *Proceedings of*
597 *the Royal Society B: Biological Sciences* **284**(1846).
- 598 66. Edger P.P., Heidel-Fischer H.M., Bekaert M., Rota J., Glöckner G., Platts A.E.,
599 Heckel D.G., Der J.P., Wafula E.K., Tang M., et al. 2015 The butterfly plant arms-race
600 escalated by gene and genome duplications. *Proceedings of the National Academy of*
601 *Sciences* **112**(27), 8362-8366. (doi:10.1073/pnas.1503926112).
- 602 67. Taylor E.L., Taylor T.N., Krings M. 2009 *Paleobotany: the biology and evolution of*
603 *fossil plants*, Academic Press.
- 604 68. Stewart W.N., Stewart W.N., Rothwell G.W. 1993 *Paleobotany and the evolution of*
605 *plants*, Cambridge University Press.
- 606 69. Puttick M.N., Clark J., Donoghue P.C.J. 2015 Size is not everything: rates of genome
607 size evolution, not C-value, correlate with speciation in angiosperms. *Proceedings of the*
608 *Royal Society B: Biological Sciences* **282**(1820).

609 70. Leitch I.J., Bennett M.D. 2004 Genome downsizing in polyploid plants. *Biological*
610 *Journal of the Linnean Society* **82**(4), 651-663. (doi:DOI 10.1111/j.1095-
611 8312.2004.00349.x).

612 71. Clark J., Hidalgo O., Pellicer J., Liu H., Marquardt J., Robert Y., Christenhusz M.,
613 Zhang S., Gibby M., Leitch I.J., et al. 2016 Genome evolution of ferns: evidence for relative
614 stasis of genome size across the fern phylogeny. *New Phytologist* **210**(3), 1072-1082.
615 (doi:10.1111/nph.13833).

616

617

618

619

620

621

622

623

624

625

626

627

628

629

630

631

632 **Figure 1.** Node-averaged rates of synonymous substitution (K_s) between paralogous pairs for
633 A) *Equisetum diffusum* and B) *Equisetum hyemale*. Components among the distributions
634 were fitted using the function *gmm()* in the *wgd* pipeline.

635

636 **Figure 2.** A histogram showing the combined posterior distribution of ages for the
637 duplication node among 240 gene families containing the signal of a gene duplication event
638 in *Equisetum*. Two clusters are defined using mixture models.

639

640 **Figure 3.** Inferred age of the whole genome duplication (WGD) event in *Equisetum*. Multi-
641 copy gene families were concatenated to inform a molecular clock analysis for each putative

642 WGD event. The 95% HPD is shown for each speciation node in blue, with the duplication
643 events in red.

644

645 **Figure 4.** Total evidence phylogeny of extinct and extant Equisetales. The tree was
646 constructed using Bayesian analysis of phenotypic and molecular data with the ages of the
647 fossils as tip calibrations and nodes calibrated using estimates from the molecular analysis.
648 Rates of phenotypic evolution (low rates in blue, high rates in red) are from the mean
649 effective branch rates from a posterior sample of 1000 trees estimated morphological data
650 alone. High rates are shown in text next to branches. The position of each putative WGD is
651 shown on the tree.

652

653 **Figure 5.** Phenotypic evolution within the Equisetales. A) An empirical phylomorphospace
654 showing the distribution of disparity within the order. The distances between taxa were
655 calculated using Gower's index and ordinated using non-metric multidimensional scaling
656 (NMDS). Character states for all ancestral nodes were reconstructed and were projected into
657 the morphospace with the tree. Convex hulls were fitted around each lineage. Colours
658 correspond to different lineages. B) The comparative morphospace occupation of extant and
659 fossil Equisetales. C) The evolution of disparity (Sum Of Variances) through time estimated
660 from the distance matrix.

661

662 **Figure 6.** The reconstruction of ancestral genome size across the Equisetales. The genome
663 size was reconstructed based on both extant and fossil 1C-value estimates. The reconstructed
664 size is shown at each node, with the width of the circle proportional to the 1C-value. The
665 middle circle represents the mean estimate, while the small and large circles represent the
666 lower and upper 95% HPD values, respectively. Branches are coloured to show the evolution
667 of large (red) and small (blue) genome sizes.

668

669

670

671

672