1 **Broad variation in rates of polyploidy and dysploidy across flowering plants is correlated**

2 **with lineage diversification**

3

4 Shing H. Zhan[1,2,*], Sarah P. Otto[1], Michael S. Barker[2]

5

6 [1]Department of Zoology & Biodiversity Research Centre, the University of British Columbia,

7 Vancouver, BC, Canada

8 [2]Department of Ecology and Evolutionary Biology, the University of Arizona, Tucson, AZ, USA

9

10 [*]Correspondence: Shing Hei Zhan. Address: Department of Zoology, the University of British

11 Columbia, 4200 - 6270 University Blvd., Vancouver, BC Canada V6T 1Z4. E-mail:

12 zhan@zoology.ubc.ca.

13

14 Running title: Chromosomal evolution in angiosperms.

15

ABSTRACT

Changes in chromosome number are considered an important driver of diversification in angiosperms. Single chromosome number changes caused by dysploidy may produce strong reproductive barriers leading to speciation. Polyploidy, or whole genome duplication, yields new species that are often reproductively isolated from progenitors and may exhibit novel morphology or ecology that may further facilitate diversification. Here, we examined the rates of polyploidy, dysploidy, and diversification across the angiosperms. Our analyses of nearly 30,000 taxa representing 46 orders and 147 families found that rates of polyploidy and dysploidy differed by two to three orders of magnitude. The rates of polyploidy and dysploidy were positively correlated with diversification rates, but relative importance analyses indicated that variation in polyploidy was better correlated with diversification rates than dysploidy. Our results provide an overview of angiosperm chromosomal evolution and a roadmap for future research on the complex relationships among polyploidy, dysploidy, and diversification.

INTRODUCTION

Chromosome number is a dynamic feature of plant evolution, especially among the flowering plants, or angiosperms [1]. Somatic chromosome counts range from $2n = 4$ in some angiosperms (e.g., *Colpodium versicolor*) [2] to $2n = 1,262$ in a fern, *Ophioglossum reticulatum* [3]. Changes in chromosome number occur via polyploidy (whole genome duplication) or dysploidy (single chromosome karyotypic changes). These mechanisms of chromosomal evolution have contributed to the rich diversity of angiosperm genome organization [4].

Polyploidy is proposed to underlie the evolution of key innovations [5], alter physiological and ecological adaptations [6–11], and influence lineage diversification in the angiosperms [12–14]. Indeed, polyploidy is pervasive throughout the history of the angiosperms [15]. Polyploidy has been correlated with some upticks in lineage diversification in the long term [13,14,16] but tends to be associated with lower diversification rates in the short term [17,18]. Given the potential deleterious effects of polyploidy (e.g., reduced hybrid fertility [19] and increased masking of alleles from selection [20–22]), it is still debated whether polyploidy is an evolutionary "dead-end" (*sensu* Stebbins [23] and Wagner [24] or *sensu* Mayrose et al. [17,25] and Arrigo and Barker [17,25]) or brings an occasional boon to angiosperm lineages, perhaps during moments of environmental upheaval [7,9]. Recent analyses indicate that polyploidy has a complex relationship with lineage diversification that is mediated by its interactions with other biological factors, such as selfing rate [16,26].

Compared to polyploidy, dysploidy has enjoyed far less attention. Dysploidy is a change in chromosome number (up or down, typically by one chromosome), caused by structural rearrangements. Increases in chromosome number occur typically by chromosome fission

53   events, whereas decreases occur by translocation events that combine two large chromosomes

54   and lead to a second smaller chromosome that is subsequently lost [27]. Dysploidy is common, as

55   indicated by chromosome numbers that deviate from euploid counts in numerous plant groups

56   [1,2,28,29]. However, the relative contribution of dysploidy to chromosomal evolution alongside

57   polyploidy is largely unexplored. Also, the effect of dysploidy on lineage diversification in the

58   angiosperms remains broadly untested. Recent studies have suggested that dysploidy may

59   elevate lineage diversification in focal plant clades, e.g., *Passiflora* [30,31], Pooideae [32], and

60   Cyperaceae [30,31]. However, a meta-analysis study [33] involving 15 plant clades (mostly genera or

61   infrageneric sections) did not find a correlation between dysploid changes and lineage

62   diversification.

63

64   Here, we explored variation in the rates of polyploidy and dysploidy across major angiosperm

65   clades (orders and families). We examined the relationships between polyploidy and dysploidy

66   with lineage diversification. To these ends, we combined chromosome numbers from the

67   Chromosome Count Database (CCDB) [2] and a time-calibrated mega-phylogeny [34,35] to assemble

68   clade-wise data sets for 46 orders and 147 families of angiosperms. For each clade, we estimated

69   the rates of polyploidy and dysploidy using ChromEvol [36,37] and estimated the rate of net

70   diversification independently of the rates of chromosome evolution using two methods:

71   Magallón and Sanderson [38,39] and Nee et al. [38,39]. We took this approach because of the lack of

72   enough resolved phylogenies for the angiosperm orders and families examined here, which

73   would have allowed us to relate chromosome evolution to speciation and extinction rates

74   explicitly across a phylogeny. By analyzing the relationships between the rates of chromosomal

75   evolution and the rate of net diversification across the major angiosperm groups, we provide

76   taxonomically broad quantitative support that both modes of chromosomal evolution —

77   polyploidy and dysploidy — likely play important roles in the formation of species and lineage

78   radiation in the angiosperms.

79

80   **RESULTS**

81   **Mode and pattern of chromosomal evolution across the angiosperms**

82   Here, we performed a large-scale phylogenetic analysis of chromosome numbers of 46 orders

83   and 147 families of angiosperms (**Supplementary Table 1**; **Supplementary Table 2**). In all, we

84   analyzed 29,770 tip taxa (55% of them had chromosome counts) across the orders as well as

85   28,000 tip taxa (56% of them had chromosome counts) across the families. We found

86   ChromEvol model support for polyploidization in all 46 angiosperm orders and 139 out of 147

87   families (**Supplementary Figure 1**; **Supplementary Table 3**; **Supplementary Table 4**). The

88   eight families with no support for polyploidization were Alstroemeriaceae, Fagaceae,

89   Grossulariaceae, Nepenthaceae, Nothofagaceae, Schisandraceae, Smilacaceae, and

90   Tamaricaceae. The most common best-fitting model among the angiosperm orders was the four-

91   parameter model "DEMI EST", which allows for both polyploidization and demi-

92   polyploidization, followed by the three-parameter model "DEMI", which sets the rates of

93   polyploidy and demi-polyploidization to be equal. Among the angiosperm families, the most

94   common best-fitting model was "DEMI" rather than "DEMI EST". Overall, in most clades, the

95   best-fitting model allowed for both polyploidization and demi-polyploidization to occur.

96

97   Using ChromEvol, we reconstructed the pattern of chromosomal evolution across the

98   angiosperm phylogeny under the best-fitting model. We found that the median gametic

99   chromosome number at internal nodes was $n = 11$ across the major groups (**Supplementary**

100   **Figure 2**), a number that was similar within each major group: in the asterids, $n = 11$; in the

101   rosids, $n = 11$; in other (non-asterid/non-rosid) eudicots, $n = 9$; in the commelinid monocots, $n =$

102   13; in other (non-commelinid) monocots, $n = 14$; in the magnoliids, $n = 12$; in Austrobaileyales

103   and Nymphaeales, $n = 14$. Moreover, the median ancestral chromosome number inferred at the

104   root of the phylogeny was $n = 7$ for angiosperm orders and $n = 9$ for angiosperm families.

105

106   **Among-lineage variation in the rates of chromosomal evolution**

107   The rates of polyploidy, as estimated under the four-parameter ChromEvol model, differed by

108   462 times and 447 times across the angiosperm orders and families, respectively (excluding eight

109   families with a rate of nearly 0 EMY, i.e., $< 1 \times 10^{-11}$ EMY) (**Figure 1**; **Figure 2**;

110   **Supplementary Figure 3**). The *order-wide* rate of polyploidy spanned from 0.00090 EMY in

111   Arecales to 0.42 EMY in Poales (median, 0.024 EMY) (**Figure 1**). The *family-wide* rate of

112   polyploidy ranged from nearly 0 EMY in eight families to 0.71 EMY in Poaceae (median, 0.025

113   EMY) (**Figure 2**; **Supplementary Figure 3**). We observed that the rosids underwent relatively

114   high rates of polyploidy (12 out of 15 rosid orders had a rate above the median rate), and the

115   non-commelinid monocots and the magnoliids experienced relatively low rates of polyploidy

116   (one out of five non-commelinid monocot orders and none of the magnoliid orders had a rate

117   above the median) (**Supplementary Table 5**; family-level results in **Supplementary Table 6**).

118

119   Rates of dysploidy (the rate of chromosome loss plus the rate of chromosome gain), as estimated

120   under the four-parameter ChromEvol model, were also remarkably variable. The rates of

121   dysploidy differed by 251 and 4,205 times across the orders and families, respectively (excluding

122   nine families with a rate of nearly 0 EMY) (**Figure 1**; **Figure 2**; **Supplementary Figure 3**). The

123   *order-wide* rates ranged from 0.0031 EMY in Austrobaileyales to 0.78 EMY in Poales (median,

124   0.029 EMY) (**Figure 1**). The *family-wide* rate of dysploidy was as low as 0 EMY in nine

125   families to as high as 5.50 EMY in Cyperaceae (median, 0.025 EMY) (**Figure 2**;

126   **Supplementary Figure 3**). We noticed that the asterids underwent relatively high rates of

127   dysploidy (eight out of 10 asterid orders had a rate above the median rate), and the non-

128   commelinid monocots and the magnoliids experienced relatively low rates of dysploidy (one out

129   of five non-commelinid monocot orders and none of the magnoliid orders had a rate above the

130   median rate) (**Supplementary Table 5**; family-level results in **Supplementary Table 6**).

131

132   We next examined the rates of chromosome loss and gain separately (**Figure 1**; **Figure 2**;

133   **Supplementary Figure 3**). The *order-wide* rates of loss ranged from nearly 0 EMY in three

134   orders to 0.61 EMY in Poales (median, 0.022 EMY), whereas the rates of gain ranged from

135   nearly 0 EMY in five orders to 0.17 EMY in Poales (median, 0.0087 EMY) (**Figure 1**). The

136   *family-wide* rates of loss range from nearly 0 EMY in 29 families to 2.76 EMY in Cyperaceae,

137   whereas the rates of gain range from nearly 0 EMY in 46 families to 2.73 EMY in Cyperaceae

138   (**Figure 2**; **Supplementary Figure 3**). The asterids underwent relatively high rates of loss (nine

139   of out 10 asterid orders had a rate above the median rate), and the non-commelinids and the

140   magnoliids had relatively low rates of loss (one out of five non-commelinid orders and none of

141   the magnoliid orders had a rate above the median rate) (**Supplementary Table 5**; family-level

142   results in **Supplementary Table 6**); these trends were not observed with the rates of gain, except

143   that the magnoliids experienced relatively low rates of gain (one out of five magnoliid orders had

144 a rate above the median rate). These results indicate that generally, the rates of loss are higher

145 than the rates of gain across the angiosperms.

146

147 Overall, the total rates of chromosomal evolution (the rate of dysploidy plus the rate of

148 polyploidy) differed by 212 and 4,324 fold across the angiosperm orders and families,

149 respectively (excluding two families with a total rate of nearly 0 EMY) (**Figure 3**; **Figure 4**;

150 **Supplementary Figure 4**). Chromosomal evolution proceeded most slowly in Austrobaileyales

151 (0.0056 EMY) and most rapidly in Poales (1.19 EMY; see below) (median, 0.053 EMY) (**Figure**

152 **3**). Generally, the rates of chromosomal evolution were lower in the monocots, the magnoliids,

153 Austrobaileyales, and Nymphaeales (median, 0.016 EMY) and higher in the eudicots (median,

154 0.026 EMY) (p = 0.032, Wilcoxon's test).

155

156 We observed broad variation in the relative contribution of polyploidy and dysploidy to

157 chromosomal evolution across the angiosperm clades. The rate of dysploidy was greater than the

158 rate of polyploidy in 27 out of 46 (59%) angiosperm orders and in 69 out of 147 (47%) families

159 (**Figure 3**; **Figure 4**; **Supplementary Figure 4**). This result suggests that dysploidy may be as

160 pervasive as polyploidy throughout the evolutionary past of the angiosperms. Additionally, the

161 rate of loss was greater than the rate of gain in 38 out of 46 (83%) angiosperm orders and in 98

162 out of 147 (67%) families (**Figure 3**; **Figure 4**; **Supplementary Figure 4**). This result suggests

163 that chromosome loss is the predominant mode of dysploidy.

164

165 **Highest rates of chromosomal evolution in Poales**

166   Poales displayed the most striking rates of dysploidy and polyploidy among the angiosperm

167   orders (**Figure 1**) and was often an outlier in subsequent analyses. Two speciose families within

168   Poales, the Cyperaceae and Poaceae, are known to experience frequent chromosomal changes

169   [23,30]. Indeed, Cyperaceae experienced the highest total rate of chromosomal evolution (5.64

170   EMY) (median, 0.055 EMY), and Poaceae had the fifth highest total rate among the angiosperm

171   families examined (**Figure 4**). Cyperaceae is well known for rampant dysploidy, especially in its

172   largest genus, *Carex* [40–42]. However, this family has a relatively low frequency of genome

173   doubling except in non-*Carex* lineages [43]. The exceptional rate of dysploidy in Cyperaceae is

174   attributed to the holocentric chromosomes in many of its species [43]. Indeed, our results found

175   that dysploidy events were frequent in the Cyperaceae (5.5 EMY; 218 times the median rate of

176   families) and polyploidy was estimated to occur at a lower rate than dysploidy (0.16 EMY; six

177   times the median rate of families) (**Figure 2**; **Figure 4**). Conversely, we found that Poaceae

178   experienced the highest rate of polyploidy among the angiosperm families (0.71 EMY; 28 times

179   the median rate of families) but a lower rate of dysploidy (0.28 EMY; 11 times the median rate

180   of families) (**Figure 2**); Poaceae still experienced a high total rate of chromosomal evolution

181   (**Figure 4**). This result is consistent with previous observations that polyploidy is common in the

182   grasses [23,28,44].

183

184   **Net diversification rate is more correlated with polyploidy rate than dysploidy rate**

185   It has long been posited that chromosomal rearrangements lead to reproductive isolation, thereby

186   contributing to species formation and lineage diversification in the angiosperms [45,46]. To test for

187   evidence of a relationship, we investigated whether the rates of chromosomal evolution

188   (polyploidy and dysploidy) were correlated with the rate of net diversification. We found that

189   the rates of net diversification estimated using MS method (assuming a relative extinction rate of

190   0.00, 0.50, or 0.90) and using the Nee method were highly positively correlated with each other

191   (**Supplementary Figure 5**), and so we took the MS rate under an extinction rate of zero for

192   correlation analysis with the rates of chromosomal evolution estimated under the four-parameter

193   ChromEvol model. Simple correlation analyses indicated that the rates of polyploidy, dysploidy,

194   and chromosomal evolution (polyploidy plus dysploidy) were positively correlated with the MS

195   net diversification rates across angiosperm orders (**Supplementary Figure 6**; see

196   **Supplementary Figure 7** for the results without Poales; **Supplementary Table 7**) and families

197   (**Supplementary Figure 8**; see **Supplementary Figure 9** for the results without Cyperaceae and

198   Poaceae; **Supplementary Table 8**). In the order-level analysis, the rate of chromosome loss –

199   but not the rate of gain – was positively correlated with the MS net diversification rate

200   (**Supplementary Figure 6**; **Supplementary Figure 7**); however, in the family-level analysis,

201   both the rates of loss and gain were positively correlated with the MS net diversification rate

202   (**Supplementary Figure 8**; **Supplementary Figure 9**). Patterns were obscured when Nee's

203   method was used to estimate diversification rates, likely because of uncertainty in extinction

204   rates when using incompletely sampled phylogenies (**Supplementary Figures 6-9**).

205

206   Although the correlation analyses suggested a relationship between chromosomal evolution and

207   diversification, the analyses were potentially confounded by shared error in a common variable.

208   The rates of chromosomal evolution and net diversification both depend on the root age (in

209   millions of years), which is estimated with error, like all phylogenetic age estimates. If the root

210   age was under- or overestimated, that would elevate or diminish, respectively, both the rates of

211   chromosome evolution and the rates of diversification, thereby artifactually leading to a positive

212   correlation between the two. As an alternative approach, we examined the relative association

213   between polyploidy versus dysploidy (or chromosome loss and gain) and net diversification,

214   which did not suffer from the problem because the rates being compared (polyploidy versus

215   dysploidy) were affected by the root age in the same way. We performed a regression analysis in

216   conjunction with a relative importance analysis (RIA) to estimate the relative proportion of

217   variance in the rates of net diversification explained by polyploidy and dysploidy. Our analysis

218   indicated that the rate of polyploidy was more correlated with the MS net diversification rate

219   than the rate of dysploidy in the angiosperm orders and the families (**Figure 5**). This trend was

220   more apparent when analyzing the angiosperm families (polyploidy, 75% versus dysploidy,

221   25%) than the orders (polyploidy, 65% versus dysploidy, 35%) (**Figure 5**). Furthermore, when

222   analyzing the rates of chromosome loss and gain separately (instead of a combined rate of

223   dysploidy), an RIA revealed that the rate of polyploidy was more correlated with the MS net

224   diversification rate than the rate of loss and the rate of gain in angiosperm orders (polyploidy,

225   56% versus loss, 35% versus gain, 10%) and families (polyploidy, 57% versus loss, 30% versus

226   gain, 13%) (**Figure 5**). These qualitative results generally hold when the outlier clades (Poales,

227   Cyperaceae, and Poaceae) were excluded (**Supplementary Figure 10**). Moreover, when the

228   Nee net diversification rates were analyzed instead, the results of the RIA were generally

229   reflected but were weaker, except in the order-level RIA considering the loss rate and gain rate

230   separately (**Supplementary Figure 11**; **Supplementary Figure 12**).

231

232   **DISCUSSION**

233   Our analyses of nearly 30,000 species found that the rates and modes of chromosomal evolution

234   vary dramatically across the angiosperms. Consistent with previous research, we found that

235 polyploidy [47,48] and dysploidy are widespread and common [33,49]. Other recent studies of plant

236 chromosomal evolution at this scale have capped chromosome numbers to keep computational

237 time reasonable [50]. In our analyses, we chose an alternative approach and analyzed individual

238 families and orders without excluding species with higher chromosome numbers. This allowed

239 us to compare and rank variation among angiosperm clades. We found that the rates of

240 chromosomal evolution were generally lower in the monocots, the magnoliids, Austrobaileyales,

241 and Nymphaeales (median, 0.016 EMY) and higher in the eudicots (median, 0.026 EMY) (p =

242 0.032, Wilcoxon's test). Although many features may explain this trend, growth form (i.e.,

243 herbaceous versus woody lineages) likely underlies some of the variation in the rate of

244 chromosomal evolution [51]. The magnoliid orders Laurales, Canellales, and Chloranthales exhibit

245 some of the lowest rates of polyploidy and dysploidy, along with Austrobaileyales. These four

246 orders are predominantly woody (i.e., trees and shrubs). Intriguingly, the remaining magnoliid

247 order Piperales contains an assortment of herbaceous and woody plants and has a higher rate of

248 dysploidy. On the other end of the spectrum, the monocots Poales and Commelinales and the

249 dicots Asterales, Brassicales, and Bruniales are herbaceous plants with the most dynamic

250 karyotypes among the orders examined, undergoing the highest rates of polyploidy and

251 dysploidy. These results are consistent with the finding of a recent study [52] showing the

252 association between polyploidy and growth forms (higher frequency of polyploidy in herbaceous

253 versus woody plants), a hypothesis proposed early on by Stebbins [53]. Moreover, it is quite

254 possible that polyploidy is associated with other traits yet to be examined jointly with polyploidy

255 [16,26,54]. For example, Stebbins [53] observed that polyploidization occurs more frequently in annual

256 and biennial lineages than in perennial lineages. Combined analyses of traits and chromosomal

257   evolution are needed to better understand what underlies the impressive variation in the rates of

258   polyploidy and dysploidy across the angiosperms.

259

260   Looking more closely at polyploidy, we found that polyploidization rates varied over 400-fold

261   across families and orders of angiosperms. Although previous research has indicated that the

262   frequency and incidence of polyploidy varies among angiosperm clades [17,47,48], prior studies have

263   not comprehensively quantified this dramatic difference in the rates of polyploidy. Polyploidy

264   was most frequent in the Poaceae with an estimated polyploid speciation event every 0.71

265   million years, consistent with expectations of the frequency of polyploidy in the family [1,55]. The

266   rate of polyploidization in the Poaceae was nearly three times the rate of the next two families,

267   the Montiaceae and the Brassicaceae. Overall, the estimated rates of polyploidization were

268   similar to estimates of the rate of plant speciation from other phylogenetic analyses [56]. Perhaps

269   most surprising given the association of polyploidy with flowering plant evolution is that we

270   detected no polyploidy in eight families. To provide some external validation of our analyses

271   with different data and confirm if polyploidy is rare or absent in these families, we compared our

272   results to inferences of ancient WGDs in the 1KP data set [15,57]. We may expect that if recent

273   polyploidy is generally rare in these families they may not have any signs of ancient polyploidy

274   either. Six of the families were represented in the 1KP study (Fagaceae, Grossulariaceae,

275   Nepenthaceae, Nothofagaceae, Schisandraceae, and Smilacaceae). There was no evidence for

276   ancient polyploidization in three (Fagaceae, Grossulariaceae, and Nothofagaceae), but there is

277   evidence for WGD in the ancestry of Nepenthaceae (NEPNα), Schisandraceae (ILFLα), and

278   Smilacaceae (SMBOα) [15,57]. The sampled contemporary chromosome numbers in Nepenthaceae

279   and Schisandraceae did not indicate recent polyploidization. However, because ChromEvol

13

280 models only chromosome counts — a gross feature of genome organization — it may not be able

281 to detect ancient WGDs in lineages that have fully diploidized. In Smilacaceae, there is one

282 possible instance of recent polyploidization ($n = 48$; $n = 13$ to 16 in other taxa), but it was not

283 enough for model support for polyploidization. Beyond these exceptional families, polyploidy

284 was a major feature of nearly all families and occurred frequently throughout the history of

285 flowering plants.

286

287 Rates of dysploidy varied more than 4,000-fold across flowering plant families, an order of

288 magnitude more than polyploidy. Perhaps unsurprisingly, this variation was driven by rapid

289 rates of dysploidy in the holocentric Cyperaceae [30]. We estimated that nearly 5.5 dysploid

290 changes occur million years in Cyperaceae, more than three times higher than the next fastest

291 family, the Gentianaceae. Holocentric chromosomes are generally thought to drive high rates of

292 chromosomal evolution [58], but a recent study in insects found that holocentric lineages did not

293 have higher rates of chromosomal evolution than monocentric lineages [59]. Similarly, besides

294 Cyperaceae, other families with at least some holocentric taxa [58] did not have exceptional rates of

295 dysploidy in our analyses. More detailed analyses are needed to assess if the holocentric

296 lineages in these families do have higher rates of chromosomal evolution, and rigorously test if

297 holocentricity is a major driver of dysploidy in plants.

298

299 Overall, we found that chromosome loss was the predominant form of dysploidy, as found in

300 other recent research [50]. Rates of loss were higher than the rates of gain across most clades of

301 angiosperms (**Figure 3**). The predominance of descending dysploidy is important because

302 chromosome numbers across flowering plants occur in a relatively tight range centered around n

303    $= 7\text{--}12$ [2,60] despite multiple rounds of polyploidy [4,15,57]. Most flowering plants are estimated to

304   have experienced 4 to 6 rounds of WGD in their ancestry [15]. Descending dysploidy following

305   polyploidy must have repeatedly brought these numbers back down to this range. In our

306   estimates, the rates of dysploidy and polyploidy were similar in many families and orders.

307   However, there is evidence that chromosome numbers may be reduced rapidly following

308   polyploidy [61–66]. This could be problematic for estimating rates in ChromEvol because the

309   models assume that the rate is equally distributed across branches. Implementing models that

310   allow additional parameters or approaches to account for the likely heterogeneity of dysploidy,

311   especially in concert with WGDs, may provide better rate estimates. Regardless, the root

312   numbers estimated in our analyses were consistent with the analysis of Carta et al. [50] despite

313   different approaches and scales of data, and this number, $n = 7$, has long been suggested as a

314   base number for flowering plants [1,28].

315

316   The changes driving these genome dynamics are also thought to drive the diversity of flowering

317   plants. Our analyses uncovered evidence for a positive relationship between chromosomal

318   evolution and diversification. Net diversification rates were positively correlated with the rates

319   of polyploidy and dysploidy in the angiosperms. Combining both types of chromosome changes,

320   Levin and Wilson [51] observed a similar positive correlation in the seed plants (rate of overall

321   karyotypic change versus rate of speciation), despite that their pioneering analysis used different

322   methods and data. In a meta-analysis, Escudero et al. [33] found a lack of correlation between

323   dysploid changes and lineage diversification in 15 plant genera or infra-generic groups.

324   However, there is a large difference in the evolutionary time scales considered. Escudero et al. [33]

325   tested the relationship at relatively short evolutionary time scales (since the divergence of genera

326   or groups nested therein). Our study, on the other hand, examined the long-term rates and

327   consequences of dysploid changes on lineage diversification (since the divergence of orders and

328   families). Consistent with our results, Márquez-Corro et al. [30] reported that dysploidy is

329   associated with increased net diversification in Cyperaceae. Similarly, considerable differences

330   in the evolutionary time scales may help explain the seemingly contradictory results that

331   polyploid lineages are correlated with lower net diversification rates within angiosperm genera [17]

332   versus higher diversification across the deep evolutionary past of the angiosperms [13,14]. Many

333   polyploid lineages may be doomed to extinction over the short term, for example, but those that

334   do persist may be a biased subset that can persist and eventually benefit from the duplicated

335   genes.

336

337   Using RIA, we teased apart the relative contributions of the rates of polyploidy and dysploidy to

338   the rate of net diversification. We found that the rate of polyploidy explained more of the

339   variation in net diversification than the rate of dysploidy in the angiosperms. It is hypothesized

340   that dysploidy may have a lesser effect on lineage diversification than polyploidy [33,67], because

341   dysploidy generally neither alters the DNA content nor disrupts gene dosage balance, whereas

342   polyploidy tends to exert noticeable effects on an organism's phenotype. Indeed, polyploids

343   have been associated with physiological changes (e.g., heterosis, [68]; changes in secondary

344   metabolite content [69–71]; drought tolerance, [72,73]; reviewed in [8]) and niche differentiation and

345   ecological changes [6,8,11,74,75], although nascent polyploids often suffer from negative effects (such

346   as reduced hybrid fertility [19]) and inefficient selection when genes are masked by multiple copies

347   [20–22]. Despite the observation that many polyploid lineages may be evolutionary "dead-ends" in

348   the short term [17], they enjoy occasional moments of success [9,14,15,25]. Furthermore, polyploidy

349 has been found to interact with other traits that shape the outcome of polyploid success [16,26] and

350 may explain the complicated relationship between polyploidy and diversification.

351

352 There are two reasons that polyploidy may be correlated with higher net diversification rate,

353 which cannot be distinguished with the correlation analyses used here. One explanation is that

354 polyploid or dysploid species diversify at a higher rate than other species, but a second

355 explanation is that diploids that readily undergo polyploidization have access to a faster route to

356 speciation than diploids that rarely undergo polyploidization, leading to a correlation between

357 groups with more species and groups with more polyploids. The same applies to dysploid routes

358 to speciation. Previous work [16,17,76] has found stronger evidence for the second explanation than

359 the former. Future analyses, such as the phylogenetic path analyses in Roman-Palacios et al. [16],

360 are needed to tease apart these mechanisms.

361

362 Our results suggest that dysploid changes may nevertheless have effects on cladogenesis and

363 lineage diversification in the angiosperms that have been largely overlooked. Others have

364 posited that post-polyploidization diploidization (via chromosome number reduction alongside

365 genome downsizing) could lead to lineage radiation [62,64,65,77]. In support of this hypothesis, our

366 results indicate that the rate of chromosomal loss is better correlated with the rate of net

367 diversification than the rate of gain. Also, an intriguing, but complicating, possibility may be

368 that the effects of dysploid changes on diversification are delayed [62], in a vein similar to the "lag-

369 time" hypothesis proposed for ancient polyploidization events in plants [13,78]. Detailed studies of

370 particular groups (e.g., [63,64]) would be insightful to disentangle the impact of polyploidy and

371 dysploidy on diversification in the angiosperms.

17

372

373   Dysploid changes arise from mechanistic errors during meiosis, which produce gametes with

374   unequal sets of chromosomes. The connection between dysploid changes and diversification

375   have been investigated in lineages with holocentric chromosomes, particularly Cyperaceae (e.g.,

376   Márquez-Corro et al. [30]). Márquez-Corro et al. [79] reviewed existing data for holocentric lineages

377   and their monocentric kins across the eukaryotes. The workers found that differences in

378   chromosome pairing behavior during meiosis may be associated with differences in lineage

379   diversification. Data for holocentricity across the angiosperm lineages are scarce, however.

380   Only a few plant clades are known to have holocentric species [79]. Nevertheless, it would be

381   fruitful for future research to disentangle the intricate relationship between holocentricity,

382   meiotic chromosome pairing behavior, and diversification in other angiosperm groups, such as

383   Marantaceae and Gentianaceae, that also experience frequent dysploidy.

384

385   The contribution of polyploidy to speciation and lineage diversification in the angiosperms has

386   been widely appreciated, but dysploidy has received far less attention. Our results indicate that

387   variation in the rates of polyploidy and dysploidy are positively correlated with the rate of net

388   diversification in the angiosperms. In turn, the among-lineage disparity in the rates of

389   chromosomal evolution may be attributed to differences in centromere structure [30], variation in

390   the mechanisms of meiotic chromosome pairing [66], growth form and life history [51–53], mating

391   system [23,26,54], and other unrecognized traits [16]. A profitable direction of future studies may be

392   more sophisticated joint phylogenetic modelling of chromosome number change, trait evolution

393   (e.g., growth form and life history), and diversification dynamics (see [16,80,81]). This interplay of

394   karyotypic, genomic, and phenotypic features is undoubtedly complex; how this interplay sways

395   the fates of angiosperm lineages remains a rich area for future research.

396

397 **MATERIALS AND METHODS**

398 **Data curation**

399 We collated (1) the phylogenetic trees, (2) chromosome numbers, and (3) estimates of species

400 richness of 46 angiosperm orders and 147 families by combining data from several sources.

401

402 First, we prepared the phylogenetic trees of major angiosperm clades. Subtrees corresponding to

403 angiosperm orders and families were extracted from the time-calibrated mega-phylogeny of Qian

404 and Jin [35] (obtained from their online Supplementary Data), using the node labels made by the

405 authors as a guide. The mega-phylogeny of Qian and Jin [35] is a corrected and expanded version

406 of the mega-phylogeny originally reconstructed by Zanne et al. [34]. We checked whether there

407 were any polytomies in the subtrees that might interfere with our analyses. We encountered one

408 polytomy in the Brassicales tree; it was randomly resolved once, and the length of the new

409 branch was set arbitrarily to 10^{-8}. The mega-phylogeny and its subtrees were processed using the

410 R package *ape* version 4.1 [82].

411

412 Next, we obtained the chromosome numbers of the angiosperm clades for which phylogenies

413 were available. Chromosome numbers were downloaded from the CCDB version 1.45 [2]

414 (accessed on Oct. 12, 2017) For many species, the CCDB provides multiple entries of gametic

415 or somatic chromosome counts. To summarize the chromosome numbers for a single species,

416 the mode gametic chromosome number (n) was taken as the representative count. The mode

417 count most likely captures the putative base number and diploid cytotype, which is the most

418 common cytotype [47]; it also minimizes the influence of unusual individual karyotypes. Somatic

419 counts ($2n$) were halved and then rounded down before they were pooled together with the

420 gametic counts to determine the mode count. Also, in instances where multiple mode counts

421 occur for a species, the lowest mode count was retained, because it more likely reflects the

422 diploid cytotype of the species (e.g., the mode count is $n = 4$ in this hypothetical series $n = 4, 4,$

423 8, 8, 16). The representative species mode counts were matched to the tips of the phylogenetic

424 trees according to their binomial names. Before that, synonyms and spelling errors in the taxon

425 names of the phylogenies were resolved using the same approach taken by Rice et al. [2]. Briefly,

426 the taxon names were matched to the accepted species names in The Plant List version 1.1 [83]

427 using Taxonome version 1.5 [84]. This step facilitated matching of the chromosome numbers with

428 the tips of the phylogenetic trees, as the species names in the Qian and Jin [35] mega-phylogeny

429 were also standardized to The Plant List. Only clades having more than 10 tip taxa with matched

430 chromosome counts were retained for analysis.

431

432 Lastly, we recorded the number of extant species of each angiosperm clade from the Angiosperm

433 Phylogeny Website (APW) version 14 [85] (accessed on Oct. 4, 2017). For one order and five

434 families, a range of species numbers was available; in these instances, the midpoint was taken as

435 the species number of the clade. For example, the APW indicated the species number of

436 Amaranthaceae ranges from 2,050 to 2,500; we took 2,275 to be the species richness of this

437 family. Also, the main estimate of species richness was taken whenever there were alternative

438 estimates.

439

440 The above data curation procedure resulted in 46 order-level and 147 family-level angiosperm

441 data sets. There were 29,770 tip taxa in total in the angiosperm order phylogenies, and 16,331

442 (55%) of these taxa had chromosome counts. On average, the percentage of taxon sampling

21

443   (number of tip taxa divided by the species count from APW) in the order phylogenies was 18%

444   (range, 3% to 75%), and the percentage of the tip taxa with chromosome counts in those

445   phylogenies was 52% (range, 22% to 87%) (**Supplementary Table 1**). Also, there were 28,000

446   tip taxa in total in the angiosperm family phylogenies, and 15,809 (56%) of these taxa had

447   chromosome counts. On average, the percentage of taxon sampling in the family phylogenies

448   was 21% (range, 2% to 100%), and the percentage of the tip taxa with chromosome counts in the

449   phylogenies was 57% (range, 10% to 97%) (**Supplementary Table 2**).

450

451   **Estimation of clade-wide rates of chromosomal evolution**

452   Using ChromEvol [36,37], we estimated the clade-wide rates of polyploidy and dysploidy for each

453   angiosperm order and family. The ChromEvol models describe chromosome number change as

454   a continuous-time Markov process along a phylogeny with observed chromosome numbers

455   assigned to the tip taxa. ChromEvol jointly estimates the rates of chromosome gain (i.e., single

456   increment), loss (i.e., single decrement), polyploidy (i.e., doubling), and demi-polyploidy (i.e.,

457   multiplication by 1.5 or half-doubling, e.g., transition between tetraploids and hexaploids) in a

458   maximum likelihood framework.

459

460   Here, we applied four basic ChromEvol models that assume that the rate of chromosome gain

461   (λ), the rate of chromosome loss (δ), the rate of polyploidy (ρ), and the rate of demi-polyploidy

462   (μ) are constant throughout a given phylogeny. In the "CONST RATE" model, μ is fixed to zero

463   while λ, δ, and ρ are allowed to vary; in "DEMI", ρ is set to equal to μ while λ and δ are allowed

464   to vary; in "DEMI EST", all four rate parameters are allowed to vary; and in "NO DUPL", ρ and

465   μ are both set to zero while λ and δ are allowed to vary (therefore, no polyploidization occurs).

466   We infer that there is no evidence for polyploidy if no model is favoured over the "NO DUPL"

467   model.

468

469   ChromEvol was run on the default optimization scheme, while setting the minimum allowed

470   chromosome number to one and the maximum allowed chromosome number to be twice the

471   highest chromosome number observed in a given dataset (e.g., in Malvaceae, the highest species

472   gametic mode count is $n = 135$, and therefore the maximum allowed chromosome number was

473   set to $n = 270$). The upper limit was applied to ensure that the transition rate matrix is reasonably

474   large to capture a plausible range of chromosome numbers (i.e., allowing doubling for the

475   species with the highest chromosome number), while maintaining computational tractability. In

476   one instance, the gametic count of *Voanioala gerardii* (Arecaceae, Arecales) was exceptionally

477   large ($n = 298$); the ChromEvol runs for Arecaceae and Arecales hardly proceeded even after a

478   month, and so we reran ChromEvol by treating the chromosome count of *V. gerardii* as missing.

479   For each clade, we determined the best-fitting ChromEvol model to be the one with the lowest

480   Akaike Information Criterion (AIC) [86].

481

482   We considered the clade-wide rates of polyploidy (defined as the sum of the rate of polyploidy

483   and the rate of demi-polyploidy) and dysploidy (defined as the sum of the rate of chromosome

484   loss and the rate of gain) under each fitted ChromEvol model. We found that the rates of

485   polyploidy and the rates of dysploidy were highly similar when estimated under the four-

486   parameter model (i.e., "DEMI EST") and the best-fitting model, which might not necessarily be

487   the four-parameter model. We used the results obtained under the best-fitting models to assess

488   ChromEvol model support across the data sets (**Supplementary Figure 1**) and to reconstruct

489   ancestral chromosome numbers in the phylogenies (**Supplementary Figure 2**). When we

490   analyzed the rate estimates, we took the results obtained under the four-parameter model to avoid

491   biases in the rate estimates that could be introduced by forcing the rate of polyploidy and/or the

492   rate of demi-polyploidization to be zero. The results of the analyses using the rates estimated

493   under the four-parameter model are presented in **Figures 1-5** and **Supplementary Figures 3, 4,**

494   **6-12**.

495

496   **Estimation of clade-wide rates of net diversification**

497   To estimate the clade-wide net diversification rate (speciation rate minus extinction rate) of each

498   angiosperm order and family, we employed two approaches: Magallón and Sanderson [38,39] and

499   Nee et al. [38,39].

500

501   Magallón and Sanderson [38,39], or MS, proposed a simple method to estimate the absolute

502   diversification rate of a clade. The MS method assumes a pure birth model, leading to

503   exponential growth, so that the rate of lineage diversification of a clade is $\log(n)/t$, where n is its

504   number of extant species in the clade and t is the age of the clade. For each clade, we set n to be

505   the species richness obtained from the APW and t to be the crown age taken from the mega-

506   phylogeny of Qian and Jin [35]. Magallon and Sanderson [38] also suggested a means to correct for

507   the number of unobserved taxa due to extinction. This involves assuming a relative extinction

508   fraction, or the extinction rate divided by the speciation rate. The standard practice is to set the

509   relative extinction fraction to different values (e.g., 0.00, 0.50, and 0.90 in Scholl and Wiens [87])

510   and then to estimate the net diversification rate under each value. We observed that the different

511   values (0, 0.50, and 0.90) gave estimates that were highly correlated (r > 0.98; Pearson's

512   correlation). Hence, we took the net diversification rates computed assuming an extinction rate

513   of zero in later correlative analyses. The rates were calculated using the implementation of the

514   MS method in the R package *geiger* version 2.0.6 [88].

515

516   Next, we estimated net diversification rates using the constant-rate birth-death model of Nee et

517   al. [39]. The Nee model describes the rates of speciation and extinction as functions of the

518   branching time distribution of a phylogeny. The rates of speciation and extinction were inferred

519   using a maximum likelihood approach, which was implemented using the R package *diversitree*

520   version 0.9-10 [89]. We accounted for random, incomplete taxon sampling by providing an

521   estimate of sampling fraction (i.e., the number of tip taxa divided by the species count collected

522   from the APW). Due to numerical imprecision from rounding, the subtrees appeared to be non-

523   ultrametric; that is, the variation among root-to-tip distances within each tree exceeded the

524   default machine tolerance used by *diversitree*. Therefore, we made minuscule corrections to the

525   branch lengths of the subtrees so that the subtrees were ultrametric. We made these corrections

526   using the non-negative least-squares method, as implemented in the R package *phangorn* version

527   2.4.0 [90] via *phytools* version 0.6-60 [91].

528

529   We conducted a maximum likelihood (ML) procedure as follows. To reduce the possibility of

530   the ML search getting stuck at local optima, we ran the search ten times using ten starting points

531   for the rates of speciation and extinction, which were randomly drawn from exponential

532   distributions (whose means were equal to the rates of speciation and extinction heuristically

533   estimated under the Nee model). The ML search was conducted using the subplex optimization

534 algorithm [92] implemented in the R package *subplex* version 1.5-4 [93]. For each clade, we took the

535 net diversification rate that had the highest likelihood score to be the ML estimate.

536

537 The above analyses yielded estimates of the net diversification rate (events per million years, or

538 EMY) obtained by two methods: (1) MS assuming an extinction rate of zero and (2) Nee under

539 ML. These rates of net diversification are positively correlated with each other (**Supplementary**

540 **Figure 3**). In the main results, we focus on the rates estimated under the simpler method, that is,

541 MS assuming an extinction rate of zero. Moreover, the MS estimator is useful for poorly

542 sampled clades (in our dataset, 26% and 27% of the angiosperm orders and families had a

543 sampling fraction of at most 10%, respectively), because it does not require a well sampled

544 species-level phylogeny, which is needed by approaches that require branch length information

545 such as that of Nee et al. [39].

546

547 **Robust regression combined with relative importance analysis**

548 A standard multiple linear regression analysis would not be meaningful here for two reasons.

549 First, there are clearly influential outliers (Poales, Cyperaceae, and Poaceae), which experience

550 exceptionally high rates of chromosomal evolution (**Figure 1**; **Figure 2**). Because those outliers

551 are genuine biological anomalies, we wished to include them in our analyses, downweighting

552 their importance as described below (see **Supplementary Figures 5, 7, 8, 10** for analyses where

553 these clades were excluded). Second, the rates of polyploidy and dysploidy exhibit collinearity.

554 Indeed, Pearson's correlation tests indicated that the rate of polyploidy and the rate of dysploidy

555 are positively correlated (orders, $r = 0.91$, $p = 2.2 \times 10^{-16}$; orders except Poales, $r = 0.60$, $p = 1.3$

556 $\times 10^{-5}$; families, $r = 0.23$, $p = 0.005$; families except Cyperaceae and Poaceae, $r = 0.33$, $p = 6.1 \times$

557   10^{-5}). To address these two issues, we performed robust regression coupled with a relative

558   importance analysis.

559

560   Robust regression handles outliers by down-weighting them as per their influence. Here, we

561   applied the method of M-estimation with Huber weighting [94], as implemented in the R package

562   *MASS* version 7.3-50 [95]. Next, we conducted a relative importance analysis, taking as input the

563   model fit by M-estimation, to estimate the proportion of variance in the rate of net diversification

564   explained by differences in rates of polyploidy and dysploidy. We obtained the relative

565   contributions of polyploidy and dysploidy using the recommended method of Lindeman et al. [96].

566   Briefly, the Lindeman et al. [96] method computes the relative proportions of the explained

567   variance (i.e., the decomposed R^2 of each predictor variable divided by the total R^2) averaged

568   over all possible sequential orderings of the independent variables, rather than basing the relative

569   proportions on a single arbitrarily chosen ordering. We performed bootstrapping (1,000

570   replicates) to obtain 95% confidence intervals around the relative contributions of the rates of

571   polyploidy and dysploidy to the rate of net diversification. We repeated this analysis for the MS

572   rate (assuming an extinction rate of zero) and the Nee rate. The relative importance analyses

573   were done using the R package *relaimpo* version 2.2-3 [97,98].

574

ACKNOWLEDGMENTS

AUTHOR CONTRIBUTIONS

SHZ designed and conducted the study, wrote the original draft of the manuscript, and reviewed and edited the manuscript. SPO provided supervision and reviewed and edited the manuscript. MSB conceived of and designed study, provided supervision, and reviewed and edited the manuscript.

COMPETING INTERESTS

The authors have no competing interests to declare.
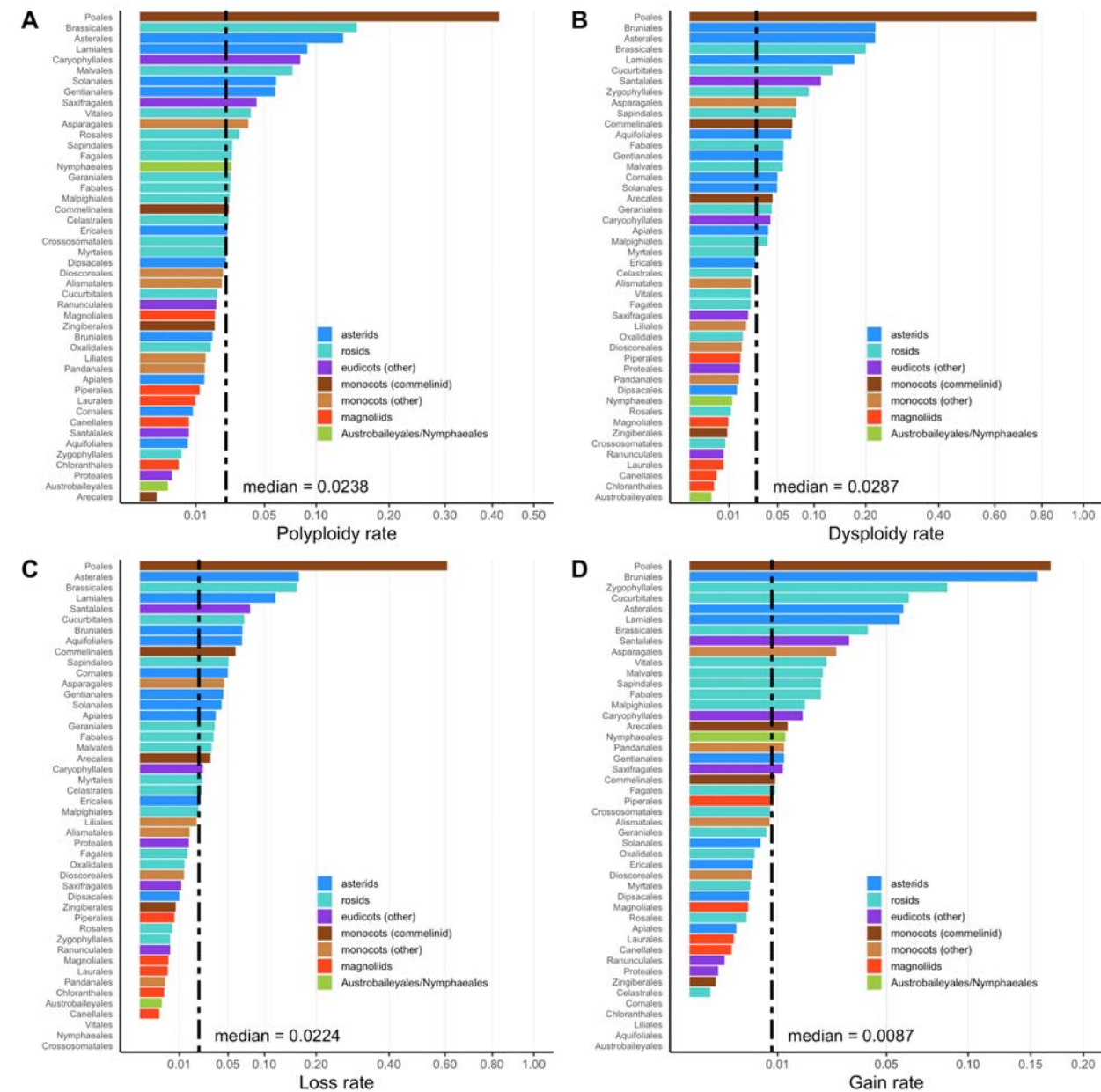
28

FIGURE LEGENDS

596

Figure 1. Angiosperm *orders* ranked by the rate of polyploidy (A), the rate of dysploidy (B), the

rate of chromosome loss (C), and rate of chromosome gain (D) estimated under the four-

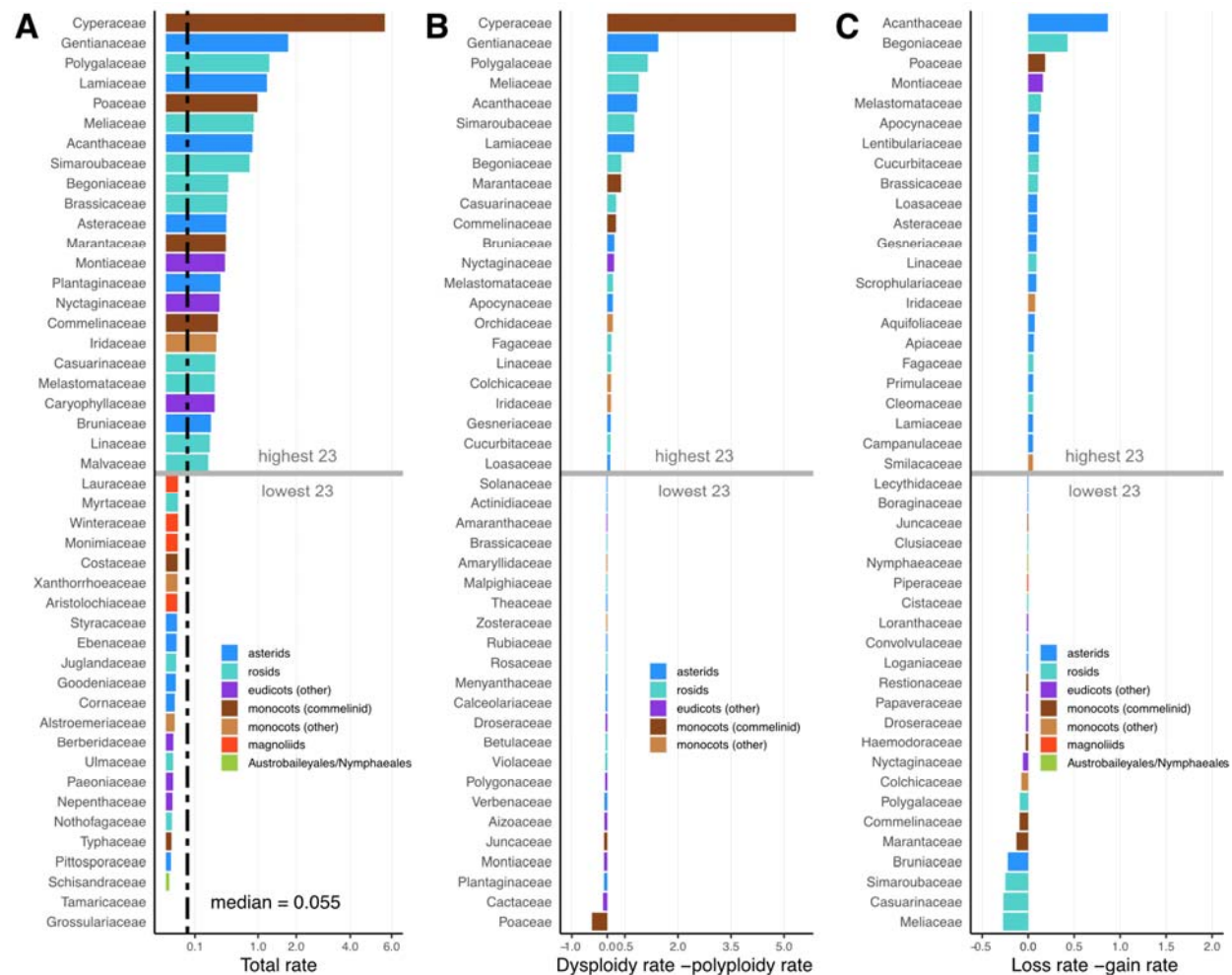parameter ChromEvol model. The x-axis scale is transformed by square root.

600
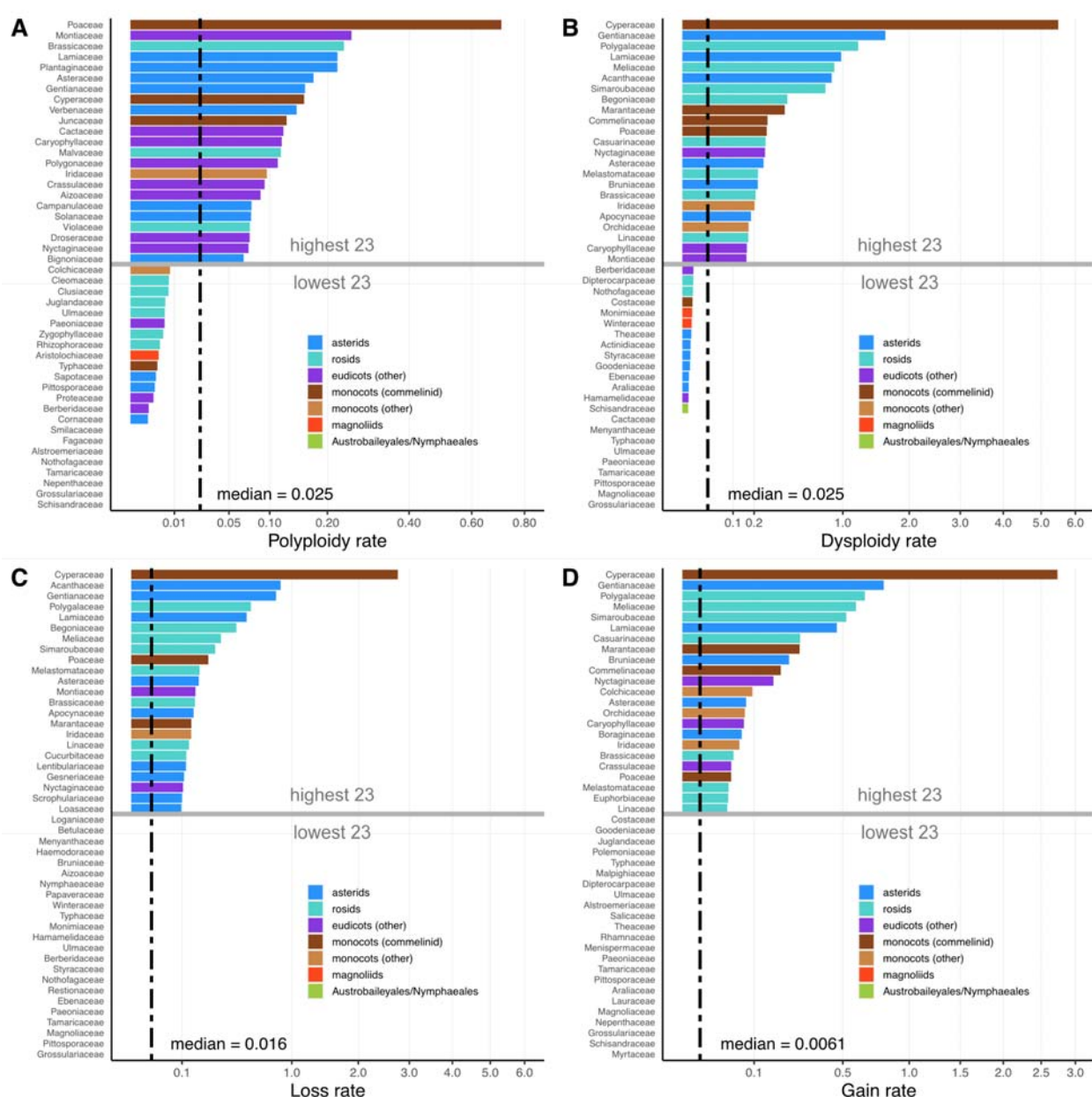
Figure 2. Angiosperm *families* ranked by the rate of polyploidy (A), the rate of dysploidy (B), the rate of chromosome loss (C), and the rate of chromosome gain (D) estimated under the four-parameter ChromEvol model. The x-axis scale is transformed by square root. Only the 23 clades with the highest rates and the 23 clades with the lowest rates for each type of rate are shown here due to space constraint (see **Supplementary Figure 3** for the full ranking).

608

609
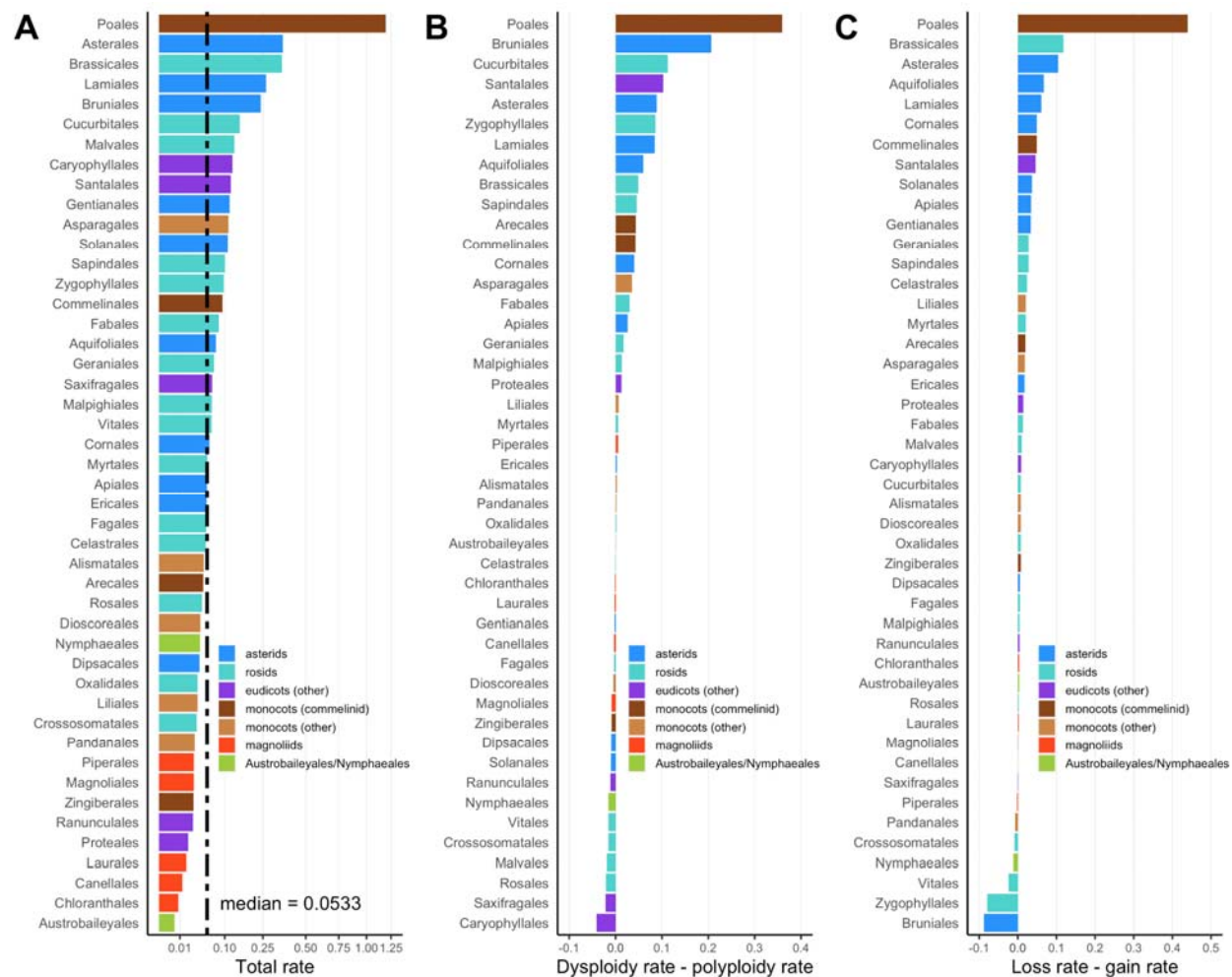
Figure 3. Angiosperm *orders* ranked by the total rate of chromosomal evolution (A), the

difference between the rate of dysploidy and the rate of polyploidy (B), and the difference

between the rate of chromosome loss and the rate of chromosome gain (C). The rates were

estimated under the four-parameter ChromEvol model. The x-axis scale is transformed by

square root.

615

Figure 4. Angiosperm *families* ranked by the total rate of chromosomal evolution (A), the difference between the rate of dysploidy and the rate of polyploidy (B), and the difference between the rate of chromosome loss and the rate of chromosome gain (C). The rates were estimated under the four-parameter ChromEvol model. The x-axis scale is transformed by square root. Only the 23 clades with the highest rates and the 23 clades with the lowest rates for each type of rate are shown here due to space constraint (see **Supplementary Figure 4** for the full ranking).
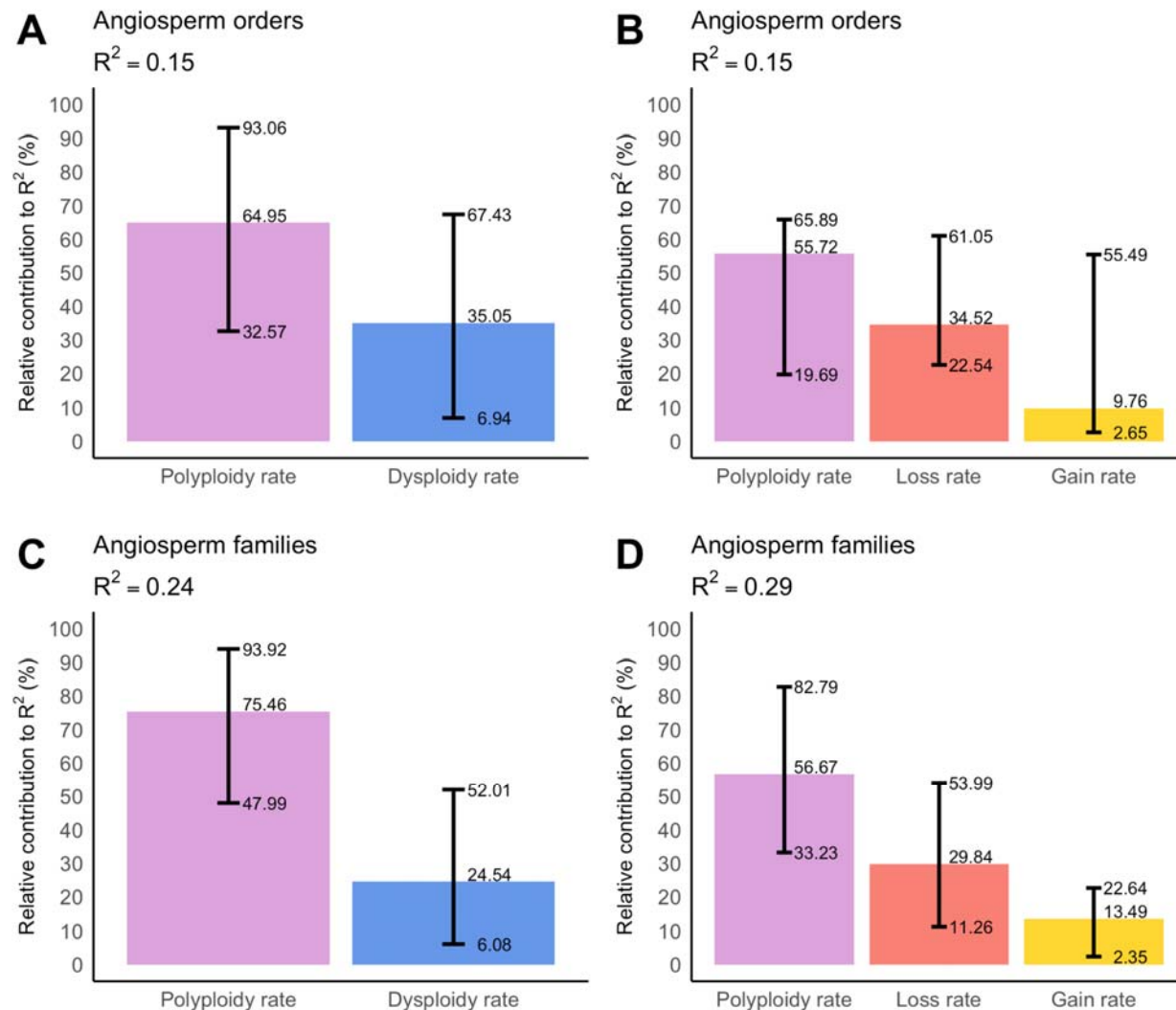
Figure 5. Relative association of the rates of chromosomal evolution to the rate of net

diversification in angiosperm orders (A, B) and families (C, D). The net diversification rate was

estimated as per Magallon and Sanderson (2011), setting the extinction rate to zero. Robust

regression followed by a relative importance analysis was conducted using the method of

Lindeman et al. (1980) with bootstrapping (1,000 replicates). The rates of chromosomal

evolution were estimated under the four-parameter ChromEvol model. We considered two

regression models; first we considered the rate of polyploidy and the rate of dysploidy (A, C),

and second we considered the rate of polyploidy and the rates of chromosome loss and gain

634   separately (B, D). The proportion of variance explained by each model is shown in the panel

635   title. We estimated the relative contribution to R^2 by the rates of polyploidy and dysploidy (A,

636   C) and by the rates of polyploidy, loss, and gain (B, D). The bars represent 95% confidence

637   intervals around the relative contributions.

638

639   **SUPPLEMENTARY MATERIALS**

640   **Supplementary Table 1.** Information about the angiosperm *orders* examined in this study. The

641   order-level phylogenies were extracted from the Qian and Jin (2016) mega-phylogeny. The

642   crown ages of the orders (in millions of years, or MY) were taken from these phylogenies and

643   the species counts of the orders were taken from the Angiosperm Phylogeny Website (APW).

644   For each clade, the number of tip taxa in the phylogeny, the number of tip taxa having

645   chromosome counts, the taxon sampling fraction (number of tip taxa in the phylogeny divided by

646   the species count), and the chromosome count sampling fraction (number of tip taxa with

647   chromosome count divided by the number of tip taxa in the phylogeny) are presented.

648

649   **Supplementary Table 2.** Information about the angiosperm *families* examined in this study.

650   See the explanation in **Supplementary Table 1**.

651

652   **Supplementary Table 3.** Results of the ChromEvol analysis for the angiosperm *order*-level data

653   sets. The best-fitting model (out of "NO DUPL", "CONST RATE", "DEMI", and "DEMI EST";

654   see **Materials and Methods** for a description of the models) was inferred based on the lowest

655   AIC score. The rates of chromosomal evolution (in events per a million years, or EMY)

656   estimated under the best-fitting model and the full four-parameter model ("DEMI EST") are

657   provided. The rate of dysploidy was defined as the sum of the rate of chromosome loss (-1) and

658   the rate of gain (+1). The rate of polyploidy was defined as the sum of the rate of duplication

659   (2x) and the rate of demi-duplication (1.5x). The ancestral gametic chromosome number was

660   inferred at the root of each clade under the best-fitting and four-parameter model.

661

662   **Supplementary Table 4.** Results of the ChromEvol analysis for the angiosperm *family*-level

663   data sets. See the explanation for **Supplementary Table 3**.

664

665   **Supplementary Table 5.** Number of angiosperm *orders* that have rates of chromosome

666   evolution above the median rate in each major angiosperm group. Total rate was defined as the

667   sum of the rate of polyploidy and the rate of dysploidy, and the dysploidy rate was defined as the

668   sum of the rate of loss and the rate of gain. The bottom row indicates the number of clades in

669   each major angiosperm group. For example, there were 7 asterid orders that had a total rate

670   above the median rate of the 10 angiosperm orders examined here.

671

672   **Supplementary Table 6.** Number of angiosperm *families* which have rates of chromosome

673   evolution above the median rate in each major angiosperm group. See the explanation for

674   **Supplementary Table 5**.

675

676   **Supplementary Table 7.** Results of the diversification analysis for the angiosperm *order*-level

677   data sets. Absolute diversification rates were estimated using the method of Magallon and

678   Sanderson (2001) assuming a relative extinction fraction of 0.00, 0.50, or 0.90. Additionally, net

679   diversification rates were estimated using the method of Nee et al. (1994) under maximum

680   likelihood, which infers the speciation rate and extinction rate of a given phylogeny.

681

682   **Supplementary Table 8.** Results of diversification analysis for the angiosperm *family*-level data

683   sets. See the explanation for **Supplementary Table 7**.

684
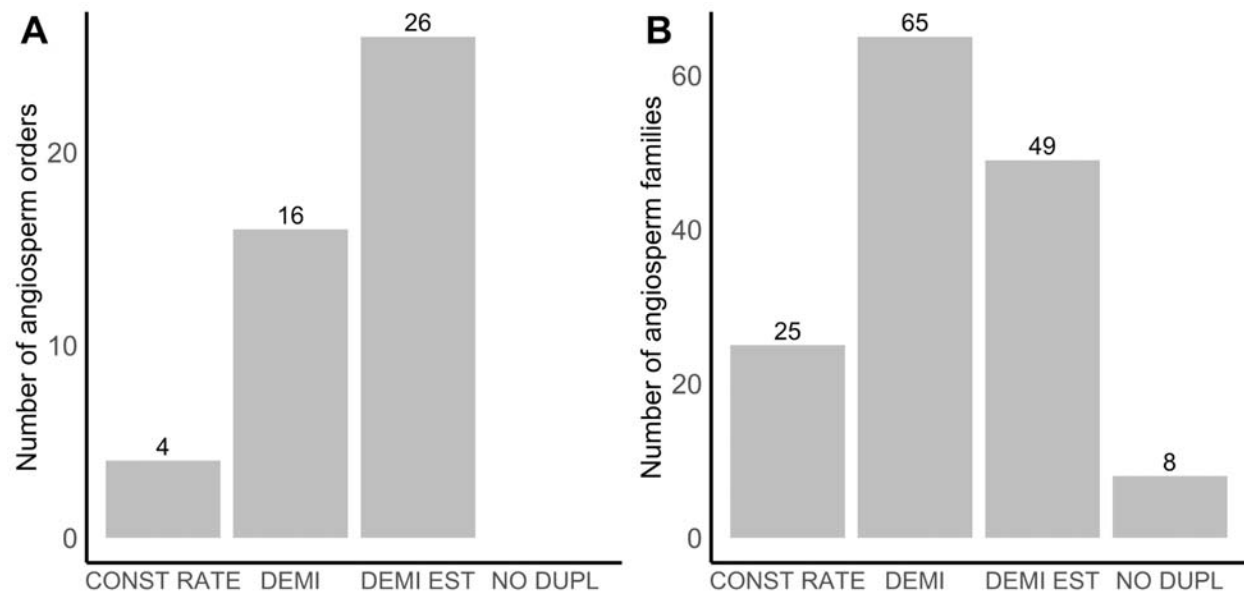
Supplementary Figure 1. Best-fitting ChromEvol models for the angiosperm *orders* (A) and *families* (B) examined in this study. The models have four rate parameters: (1) chromosome gain or ascending dysploidy, λ; (2) chromosome loss or descending dysploidy, δ; (3) doubling or polyploidization, ρ; and (4) half-doubling or demi-polyploidization, μ. In the "CONST RATE" model, μ is fixed to zero; in "DEMI", ρ is set to equal to μ; in "DEMI EST", all four rates are allowed to vary; and in "NO DUPL", ρ and μ are both set to zero (therefore, no polyploidization is allowed). In all the angiosperm orders and most of the families (except eight cases), there was model support for polyploidization.

693
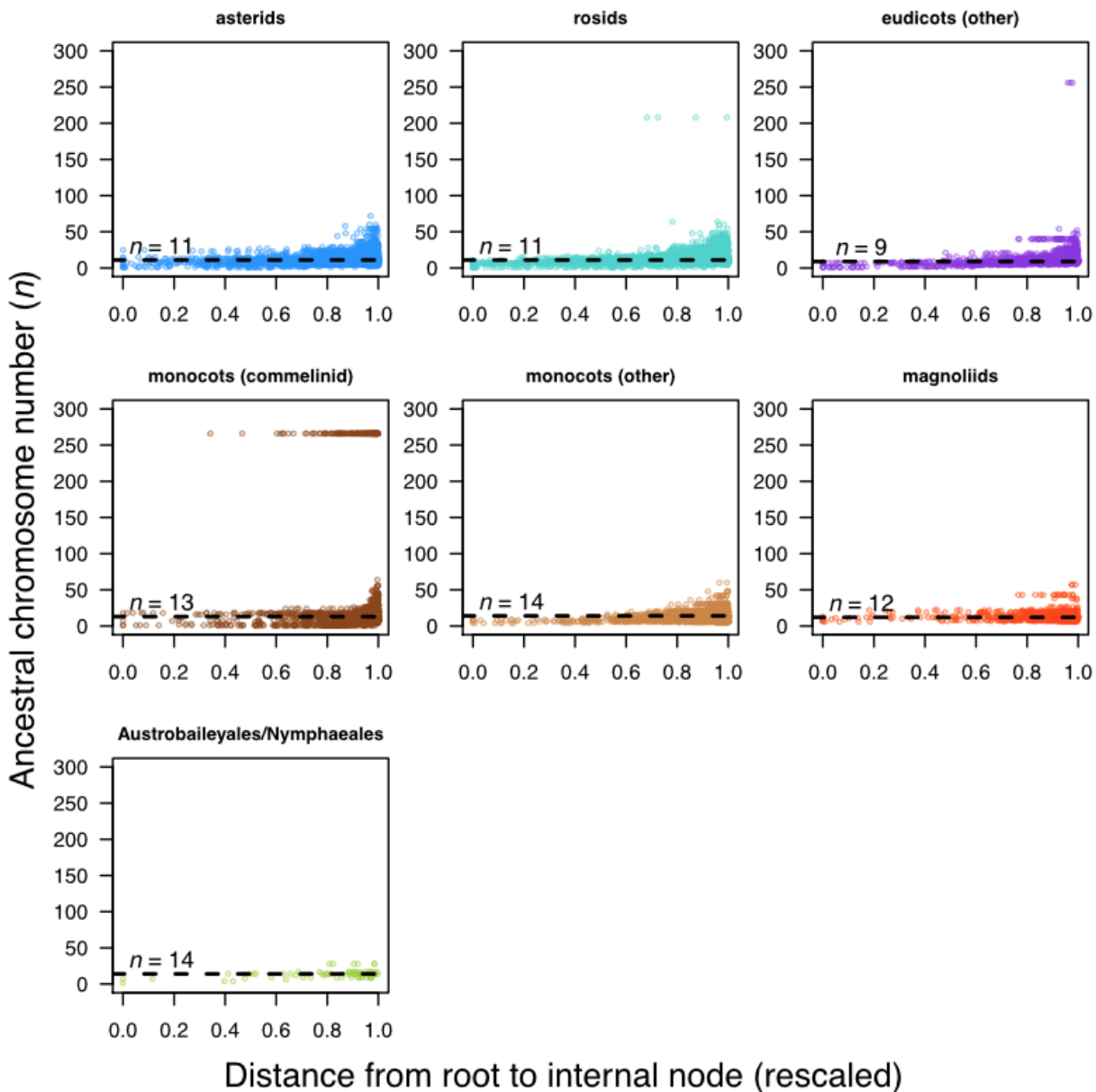
Supplementary Figure 2. Temporal pattern of chromosome number evolution across major angiosperm groups. The ancestral chromosome numbers were inferred, under the best-fitting ChromEvol model by maximum likelihood, at each internal node in the ***order***-level subtrees extracted from the Qian and Jin (2016) mega-phylogeny. The inferred gametic chromosome numbers at internal nodes are plotted against the age of the node (rescaled such that in each

38

700   subtree the root depth was one). The dashed horizontal line indicates the median chromosome

701   number across the internal nodes in all the subtrees belonging to a major group.

702

703

Supplementary Figure 3. Full version of **Figure 2**.
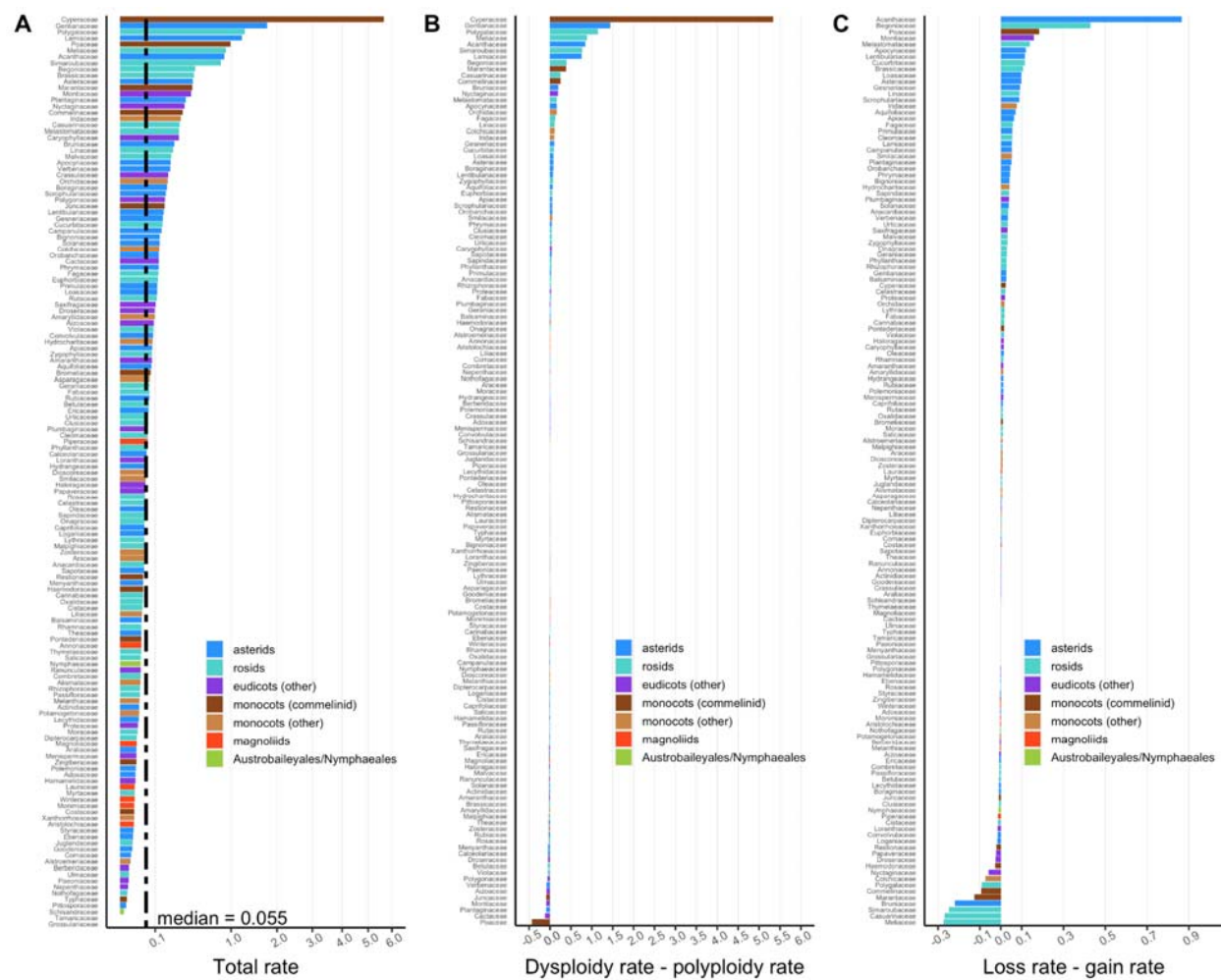
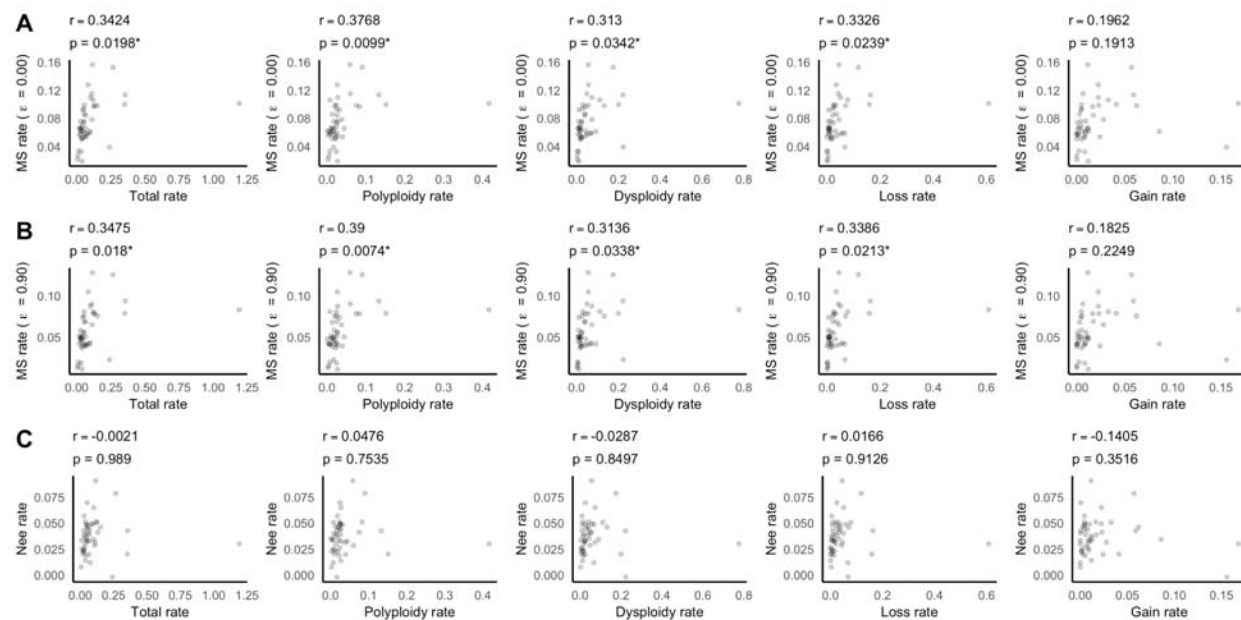**Supplementary Figure 4.** Full version of **Figure 4**.
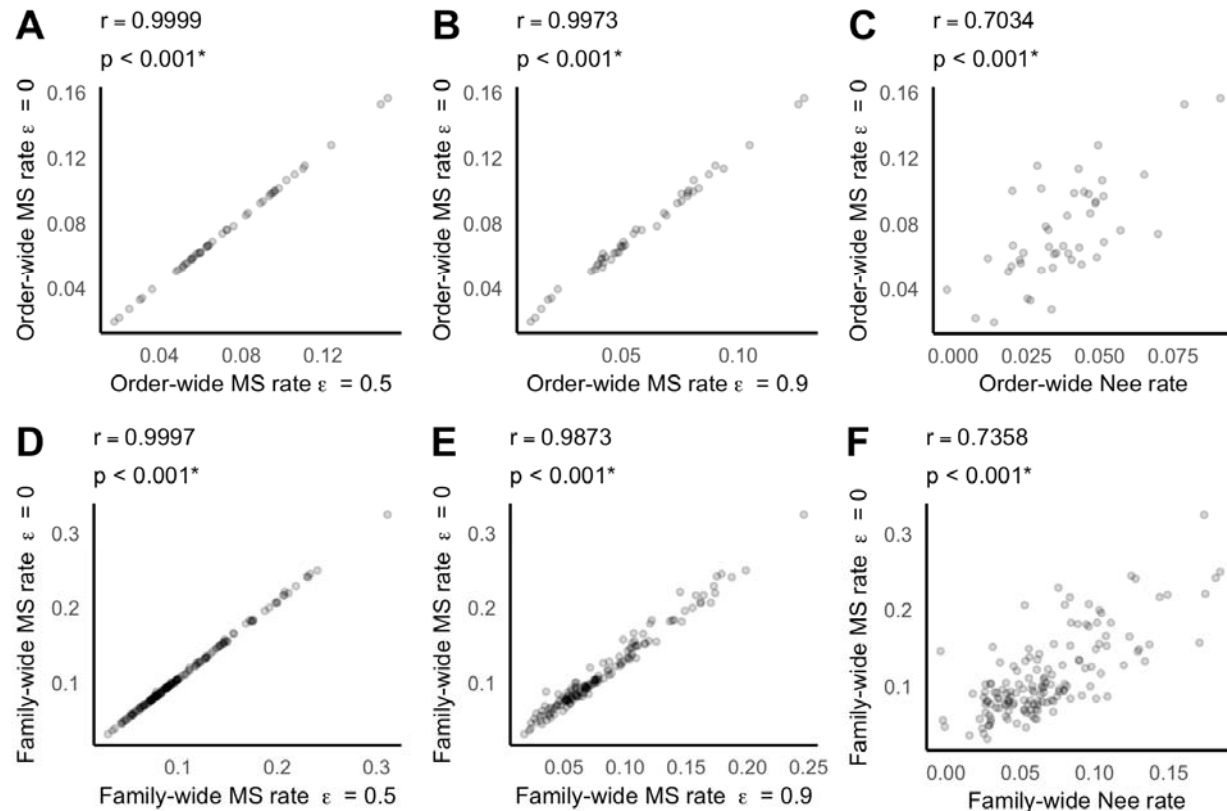
Supplementary Figure 5. Pairwise correlations between the rates of net diversification estimated using different methods for the angiosperm *orders* (A, B, C) and *families* (D, E, F) . The Magallon and Sanderson (2001) rates were estimated assuming three different values of the relative extinction fraction (ε = 0.00, 0.50, or 0.90). The Nee et al. (1994) rates were estimated under maximum likelihood. Pearson's correlation coefficients and p-values obtained using the Pearson's test are shown in the panel titles. An asterisk indicates a p-value of at most 0.05.
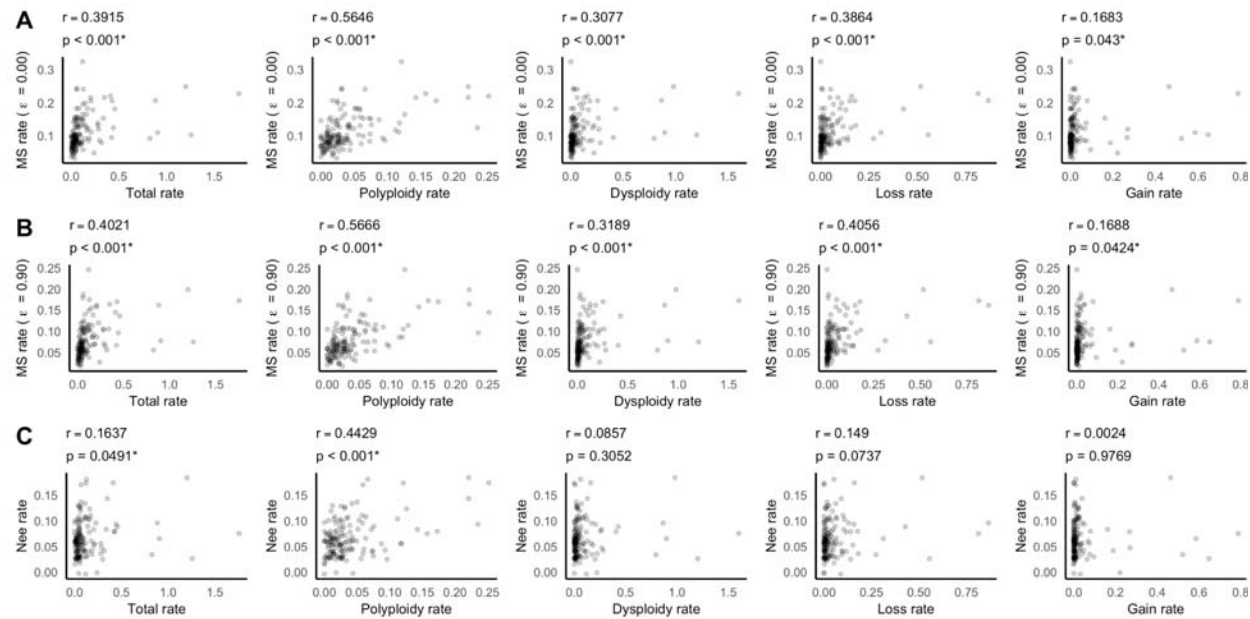
Supplementary Figure 6. Pairwise correlations between the *order*-wide rates of net diversification and the rates of chromosome evolution. The Magallon and Sanderson (2001) rates were estimated assuming a relative extinction fraction (ε) of 0.00 (A) or 0.90 (B), and the Nee et al. (1994) rates were estimated under maximum likelihood (C). The rates of chromosomal evolution were estimated under the four-parameter ChromEvol model. The dysploidy rate was computed as the sum of the chromosome gain rate and loss rate. The total rate was computed as the sum of the polyploidy rate and dysploidy rate. Pearson's correlation coefficients and p-values obtained using the Pearson's test are shown. An asterisk indicates a p-value of at most 0.05. See **Supplementary Figure 7** for similar plots without Poales (the right most point in each panel).
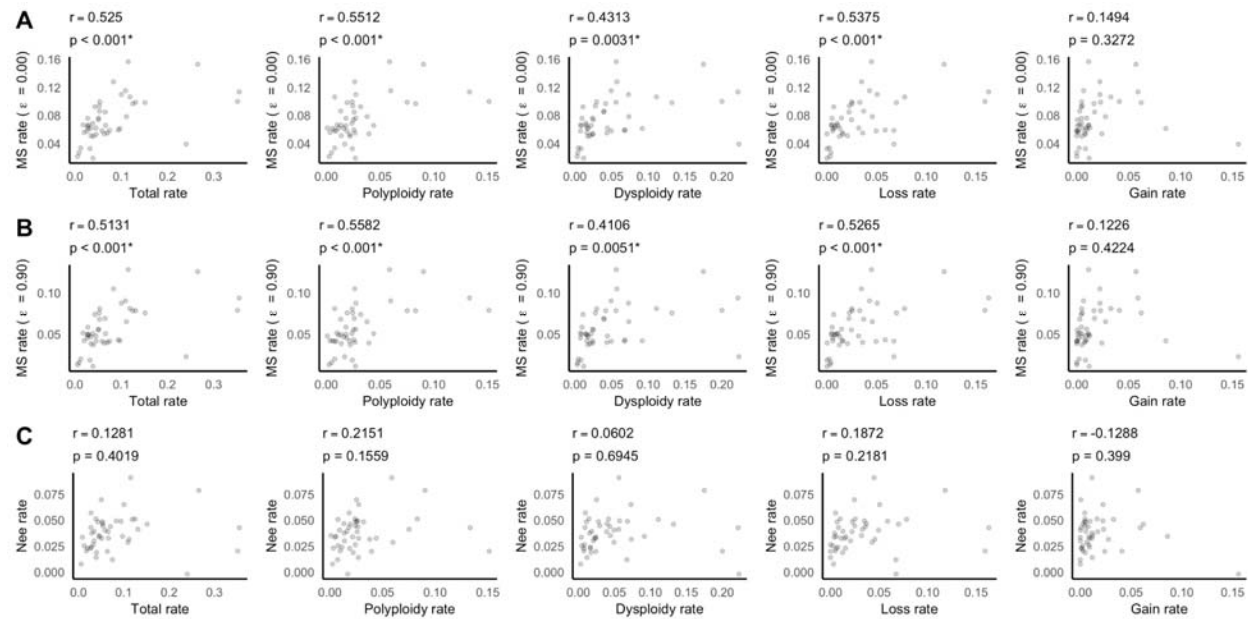
Supplementary Figure 7. Pairwise correlations between the *order*-wide rates of net diversification and the rates of chromosomal evolution, *excluding* Poales. See **Supplementary Figure 6** for further explanation.

733
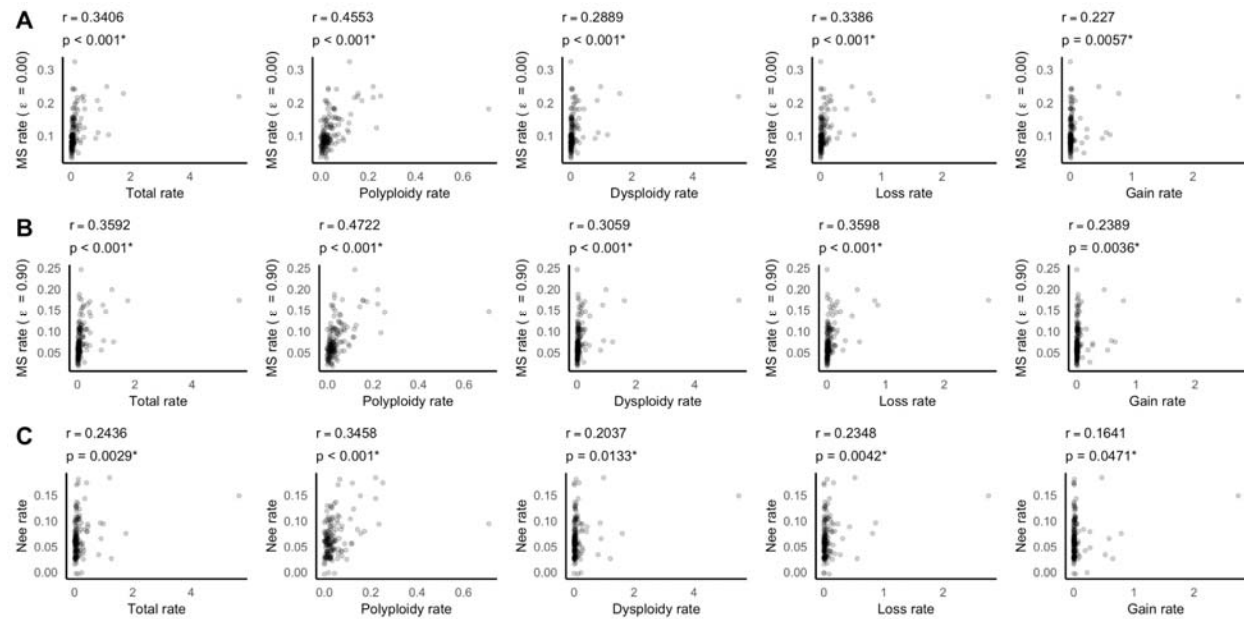
734 **Supplementary Figure 8.** Pairwise correlations between the *family*-wide rates of net

735 diversification and the rates of chromosomal evolution. See **Supplementary Figure 6** for

736 further explanation and **Supplementary Figure 9** for similar plots without Cyperaceae (the right

737 most points in each panel) and Poaceae, which both have atypically high rates of karyotype
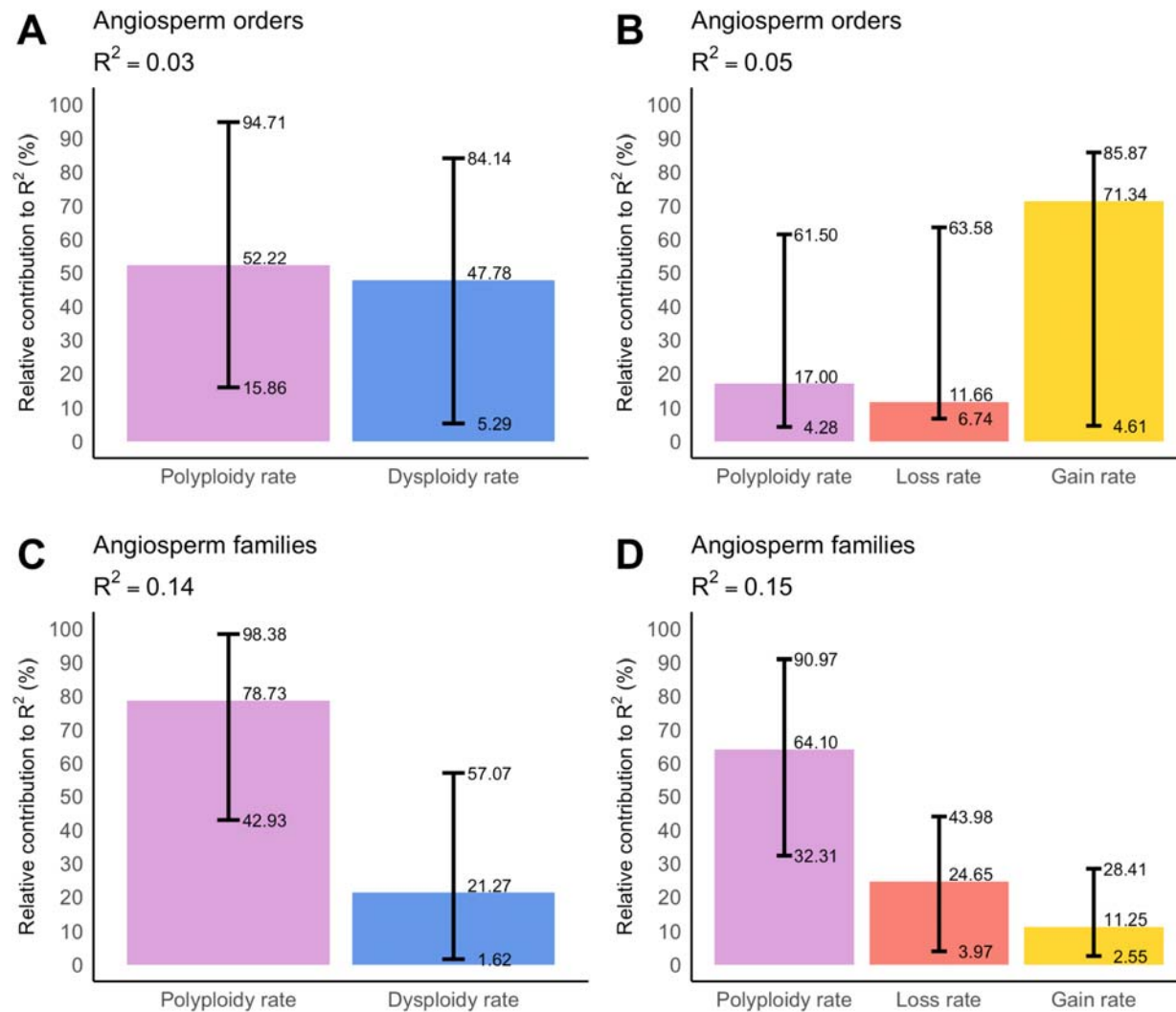
738 evolution.

739

Supplementary Figure 9. Pairwise correlations between the *family*-wide rates of net diversification and the rates of chromosomal evolution, *excluding* Cyperaceae and Poaceae. See **Supplementary Figure 6** for further explanation.

745
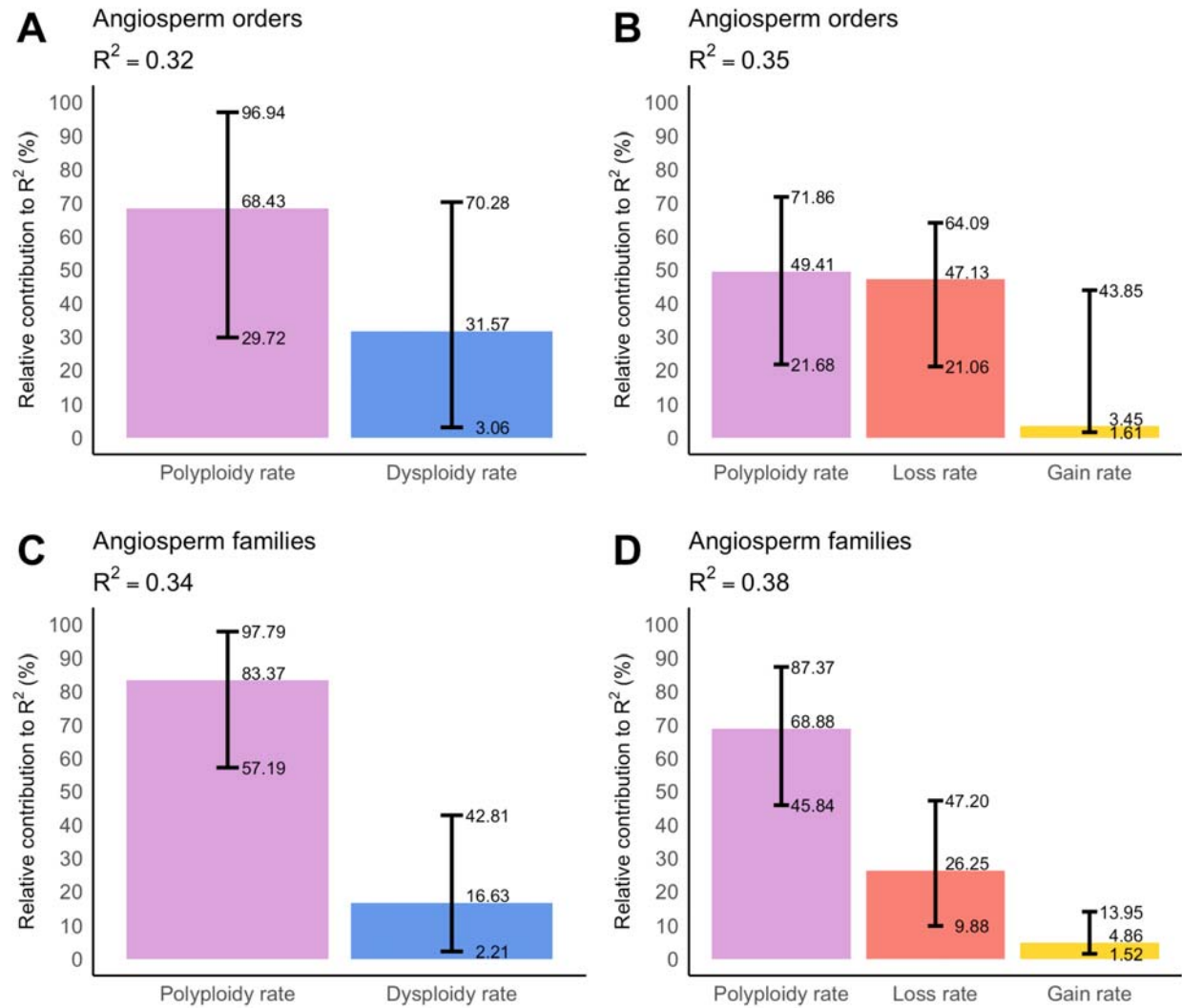
Supplementary Figure 10. Relative association between the rate of net diversification estimated

using the method of Magallon and Sanderson (2001) rate (assuming an extinction rate of zero)

and the rates of chromosomal evolution in angiosperm *orders* (A, B) and *families* (C, D),

excluding Poales, Cyperaceae, and Poaceae. See **Figure 5** for further explanation.

750

Supplementary Figure 11. Relative association between the rate of net diversification estimated using the Nee et al. (1994) method and the rates of chromosomal evolution in angiosperm *orders* (A, B) and *families* (C, D). See **Figure 5** for further explanation.
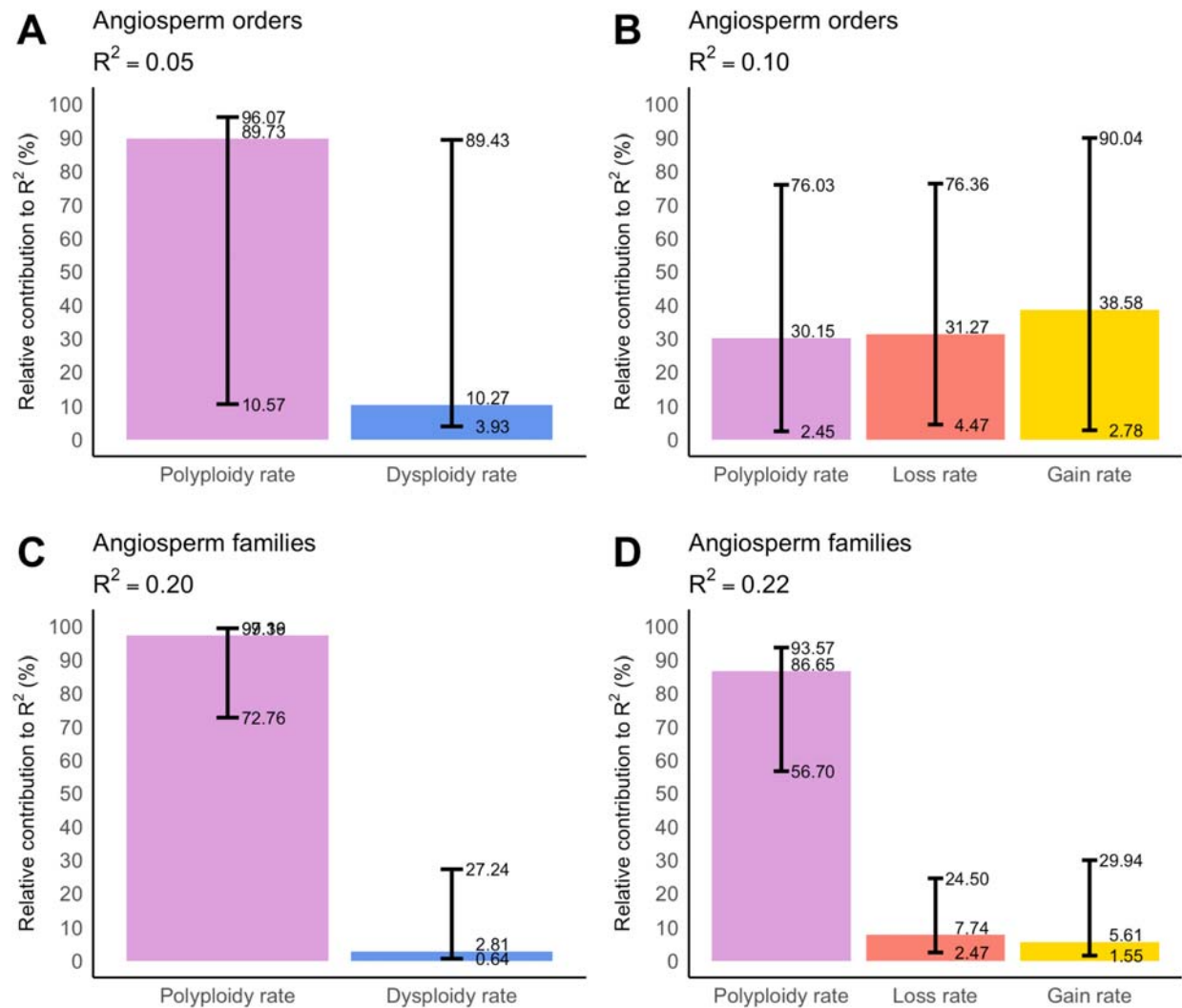
Supplementary Figure 12. Relative association between the rate of net diversification estimated using the Nee et al. (1994) method and the rates of chromosomal evolution in angiosperm *orders* (A, B) and *families* (C, D), *excluding* Poales, Cyperaceae, and Poaceae. See **Figure 5** for further explanation.

REFERENCES

1. Stebbins, G. L. *Chromosomal Evolution in Higher Plants*. (Edward Arnold, 1971).

2. Rice, A. *et al.* The Chromosome Counts Database (CCDB) – a community resource of plant chromosome numbers. *New Phytol.* **206**, 19–26 (2015).

3. Abraham, A. & Ninan, C. A. The chromosomes of Ophioglossum reticulatum L. *Curr. Sci.* **23**, 213–214 (1954).

4. Escudero, M. & Wendel, J. F. The grand sweep of chromosomal evolution in angiosperms. *New Phytol.* **228**, 805–808 (2020).

5. Soltis, P. S. & Soltis, D. E. Ancient WGD events as drivers of key innovations in angiosperms. *Curr. Opin. Plant Biol.* **30**, 159–165 (2016).

6. Otto, S. P. The evolutionary consequences of polyploidy. *Cell*. **131**, 452–462 (2007).

7. Van de Peer, Y., Maere, S. & Meyer, A. The evolutionary significance of ancient genome duplications. *Nat. Rev. Genet.* **10**, 725–732 (2009).

8. te Beest, M. *et al.* The more the better? The role of polyploidy in facilitating plant invasions. *Ann. Bot.* **109**, 19–45 (2012).

9. Van de Peer, Y., Mizrachi, E. & Marchal, K. The evolutionary significance of polyploidy. *Nat. Rev. Genet.* **18**, 411–424 (2017).

10. Wei, N., Cronn, R., Liston, A. & Ashman, T.-L. Functional trait divergence and trait plasticity confer polyploid advantage in heterogeneous environments. *New Phytol.* **221**, 2286–2297 (2019).

11. Baniaga, A. E., Marx, H. E., Arrigo, N. & Barker, M. S. Polyploid plants have faster rates of multivariate niche differentiation than their diploid relatives. *Ecol. Lett.* **23**, 68–78 (2020).

785   12.   Soltis, D. E. *et al.* Polyploidy and angiosperm diversification. *Am J Bot.* **96**, 336–348

786   (2009).

787   13.   Tank, D. C. *et al.* Nested radiations and the pulse of angiosperm diversification: increased

788   diversification rates often follow whole genome duplications. *New Phytol.* **207**, 454–467

789   (2015).

790   14.   Landis, J. B. *et al.* Impact of whole-genome duplication events on diversification rates in

791   angiosperms. *Am. J. Bot.* **105**, 348–363 (2018).

792   15.   One Thousand Plant Transcriptomes Initiative. One thousand plant transcriptomes and the

793   phylogenomics of green plants. *Nature* **574**, 679–685 (2019).

794   16.   Román-Palacios, C., Molina-Henao, Y. F. & Barker, M. S. Polyploids increase overall

795   diversity despite higher turnover than diploids in the Brassicaceae. *Proc. Biol. Sci.* **287**,

796   20200962 (2020).

797   17.   Mayrose, I. *et al.* Recently formed polyploid plants diversify at lower rates. *Science* **333**,

798   1257 (2011).

799   18.   Mayrose, I. *et al.* Methods for studying polyploid diversification and the dead end

800   hypothesis: a reply to Soltis et al. (2014). *New Phytol.* **206**, 27–35 (2015).

801   19.   Ramsey, J. & Schemske, D. W. Neopolyploidy in flowering plants. *Annu. Rev. Ecol. Syst.*

802   **33**, 589–639 (2002).

803   20.   Fisher, R. A. The sheltering of lethals. *Am. Nat.* **69**, 446–455 (1935).

804   21.   Wright, S. *Evolution and the Genetics of Populations, Volume 2: Theory of Gene*

805   *Frequencies*. (University of Chicago Press, 1984).

806   22.   Haldane, J. B. *The Causes of Evolution*. (Princeton University Press, 1990).

807   23.   Stebbins, G. L. *Variation and Evolution in Plants*. (Columbia University Press, 1950).

808   24. Wagner, W. H. Biosystematics and evolutionary noise. *Taxon* **19**, 146–151 (1970).

809   25. Arrigo, N. & Barker, M. S. Rarely successful polyploids and their legacy in plant genomes.

810   *Curr. Opin. Plant Biol.* **15**, 140–146 (2012).

811   26. Zenil☐Ferguson, R. *et al.* Interaction among ploidy, breeding system and lineage

812   diversification. *New Phytol.* **224**, 1252–1265 (2019).

813   27. Schubert, I. & Lysak, M. A. Interpretation of karyotype evolution should consider

814   chromosome structural constraints. *Trends Genet.* **27**, 207–216 (2011).

815   28. Grant, V. *Plant Speciation*. (Columbia University Press, 1981).

816   29. Otto, S. P. & Whitton, J. Polyploid incidence and evolution. *Annu. Rev. Genet.* **34**, 401–437

817   (2000).

818   30. Márquez-Corro, J. I., Martín-Bravo, S., Spalink, D., Luceño, M. & Escudero, M. Inferring

819   hypothesis-based transitions in clade-specific models of chromosome number evolution in

820   sedges (Cyperaceae). *Mol. Phylogenet. Evol*. **135**, 203–209 (2019).

821   31. Sader, M. A., Amorim, B. S., Costa, L., Souza, G. & Pedrosa-Harand, A. The role of

822   chromosome changes in the diversification of Passiflora L. (Passifloraceae). *Syst. Biodivers.*

823   **17**, 7–21 (2019).

824   32. Pimentel, M., Escudero, M., Sahuquillo, E., Minaya, M. Á. & Catalán, P. Are

825   diversification rates and chromosome evolution in the temperate grasses (Pooideae)

826   associated with major environmental changes in the Oligocene-Miocene? *PeerJ* **5**, e3815

827   (2017).

828   33. Escudero, M. *et al.* Karyotypic changes through dysploidy persist longer over evolutionary

829   time than polyploid changes. *PLoS ONE* **9**, e85266 (2014).

830   34. Zanne, A. E. *et al.* Three keys to the radiation of angiosperms into freezing environments.

831   *Nature* **506**, 89–92 (2014).

832   35. Qian, H. & Jin, Y. An updated megaphylogeny of plants, a tool for generating plant

833   phylogenies and an analysis of phylogenetic community structure. *J. Plant Ecol.* **9**, 233–239

834   (2016).

835   36. Mayrose, I., Barker, M. S. & Otto, S. P. Probabilistic models of chromosome number

836   evolution and the inference of polyploidy. *Syst. Biol.* **59**, 132–144 (2010).

837   37. Glick, L. & Mayrose, I. ChromEvol: assessing the pattern of chromosome number evolution

838   and the inference of polyploidy along a phylogeny. *Mol. Biol. Evol.* **31**, 1914–1922 (2014).

839   38. Magallón, S. & Sanderson, M. J. Absolute diversification rates in angiosperm clades.

840   *Evolution* **55**, 1762–1780 (2001).

841   39. Nee, S., May, R. M. & Harvey, P. H. The reconstructed evolutionary process. *Philos. Trans.*

842   *R. Soc. Lond. B Biol. Sci.* **344**, 305–311 (1994).

843   40. Heilborn, O. Chromosome numbers and dimensions, species-formation and phylogeny in

844   the genus Carex. *Hereditas* **5**, 129–216 (2010).

845   41. Davies, E. W. Cytology, evolution and origin of the aneuploid series in the genus Carex.

846   *Hereditas* **42**, 349–365 (1956).

847   42. Hipp, A. L., Rothrock, P. E. & Roalson, E. H. The evolution of chromosome arrangements

848   in Carex (Cyperaceae). *Bot. Rev.* **75**, 96–109 (2009).

849   43. Escudero, M., Hipp, A. L., Waterway, M. J. & Valente, L. M. Diversification rates and

850   chromosome evolution in the most diverse angiosperm genus of the temperate zone (Carex,

851   Cyperaceae). *Mol. Phylogenet. Evol.* **63**, 650–655 (2012).

852   44. Levy, A. A. & Feldman, M. The impact of polyploidy on grass genome evolution. *Plant*

853   *Physiol.* **130**, 1587–1593 (2002).

854   45. Rieseberg, L. H. Chromosomal rearrangements and speciation. *Trends in Ecology &*

855   *Evolution* **16**, 351–358 (2001).

856   46. Owens, G. L. & Rieseberg, L. H. Hybrid incompatibility is acquired faster in annual than in

857   perennial species of sunflower and tarweed. *Evolution* **68**, 893–900 (2014).

858   47. Wood, T. E. *et al.* The frequency of polyploid speciation in vascular plants. *Proc. Natl.*

859   *Acad. Sci. U. S. A.* **106**, 13875–13879 (2009).

860   48. Barker, M. S., Arrigo, N., Baniaga, A. E., Li, Z. & Levin, D. A. On the relative abundance

861   of autopolyploids and allopolyploids. *New Phytol.* **210**, 391–398 (2016).

862   49. Brochmann, C. *et al.* Polyploidy in arctic plants. *Biol. J. Linn. Soc.* **82**, 521–536 (2004).

863   50. Carta, A., Bedini, G. & Peruzzi, L. A deep dive into the ancestral chromosome number and

864   genome size of flowering plants. *New Phytol.* **228**, 1097–1106 (2020).

865   51. Levin, D. A. & Wilson, A. C. Rates of evolution in seed plants: Net increase in diversity of

866   chromosome numbers and species numbers through time. *Proc. Natl. Acad. Sci. U. S. A.* **73**,

867   2086–2090 (1976).

868   52. Zenil-Ferguson, R., Ponciano, J. M. & Burleigh, J. G. Testing the association of phenotypes

869   with polyploidy: An example using herbaceous and woody eudicots. *Evolution* **71**, 1138–

870   1148 (2017).

871   53. Stebbins, G. L. Cytological characteristics associated with the different growth habits in the

872   dicotyledons. *Am. J. Bot.* **25**, 189 (1938).

873   54. Robertson, K., Goldberg, E. E. & Igić, B. Comparative evidence for the correlated evolution

874   of polyploidy and self-compatibility in Solanaceae. *Evolution* **65**, 139–155 (2011).

875   55. Estep, M. C. *et al.* Allopolyploidy, diversification, and the Miocene grassland expansion.

876   *Proc. Natl. Acad. Sci. U. S. A.* **111**, 15149–15154 (2014).

877   56.   Thomas, C. D. Rapid acceleration of plant speciation during the Anthropocene. *Trends*

878   *Ecol. Evol.* **30**, 448–455 (2015).

879   57.   Li, Z. & Barker, M. S. Inferring putative ancient whole-genome duplications in the 1000

880   Plants (1KP) initiative: access to gene family phylogenies and age distributions.

881   *Gigascience* **9**, (2020).

882   58.   Mandrioli, M. & Manicardi, G. C. Holocentric chromosomes. *PLoS Genet.* **16**, e1008918

883   (2020).

884   59.   Ruckman, S. N., Jonika, M. M., Casola, C. & Blackmon, H. Chromosome number evolves

885   at equal rates in holocentric and monocentric clades. *PLoS Genet.* **16**, e1009076 (2020).

886   60.   Román-Palacios, C., Medina, C. A., Zhan, S. H. & Barker, M. S. Animal chromosome

887   counts reveal a similar range of chromosome numbers but with less polyploidy in animals

888   compared to flowering plants. *Cold Spring Harbor Laboratory* 2020.10.10.334722 (2020)

889   doi:10.1101/2020.10.10.334722.

890   61.   Wendel, J. F. The wondrous cycles of polyploidy in plants. *Am. J. Bot.* **102**, 1753–1756

891   (2015).

892   62.   Dodsworth, S., Chase, M. W. & Leitch, A. R. Is post-polyploidization diploidization the key

893   to the evolutionary success of angiosperms? *Bot. J. Linn. Soc.* **180**, 1–5 (2015).

894   63.   Mandáková, T., Li, Z., Barker, M. S & Lysak, M. A. Diverse genome organization

895   following 13 independent mesopolyploid events in Brassicaceae contrasts with convergent

896   patterns of gene retention. *Plant J.* **91**, 3–21 (2017).

897   64.   Mandáková, T. *et al.* Multispeed genome diploidization and diversification after an ancient

898   allopolyploidization. *Mol. Ecol.* **26**, 6445–6462 (2017).

899   65.   Mandáková, T. & Lysak, M. A. Post-polyploid diploidization and diversification through

900   dysploid changes. *Curr. Opin. Plant Biol.* **42**, 55–65 (2018).

901   66.   Li, Z. *et al.* Patterns and processes of diploidization in land plants. *Annu. Rev. Plant Biol.*

902   (2021) doi:10.1146/annurev-arplant-050718-100344.

903   67.   Fawcett, J. A., Maere, S. & Van de Peer, Y. Plants with double genomes might have had a

904   better chance to survive the Cretaceous-Tertiary extinction event. *Proc. Natl. Acad. Sci. U.*

905   *S. A.* **106**, 5737–5742 (2009).

906   68.   Chen, Z. J. Molecular mechanisms of polyploidy and hybrid vigor. *Trends Plant Sci.* **15**,

907   57–71 (2010).

908   69.   Dijkstra, H. & Speckmann, G. J. Autotetraploidy in caraway (Carum carvi L.) for the

909   increase of the aetheric oil content of the seed. *Euphytica* **29**, 89–96 (1980).

910   70.   Lavania, U. C. & Srivastava, S. Enhanced productivity of tropane alkaloids and fertility in

911   artificial autotetraploids of *Hyoscyamus niger* L. *Euphytica* **52**, 73–77 (1991).

912   71.   Hull☐Sanders, H. M. & Johnson, R. H. Effects of polyploidy on secondary chemistry,

913   physiology, and performance of native and invasive genotypes of Solidago gigantea

914   (Asteraceae). *Am. J. Bot.* **96**, 762–770. (2009).

915   72.   Li, W.-L., Berlyn, G. P. & Ashton, P. M. S. Polyploids and their structural and

916   physiological characteristics relative to water deficit in Betula papyrifera (Betulaceae). *Am.*

917   *J. Bot.* **83**, 15–20 (1996).

918   73.   Maherali, H., Walden, A. E. & Husband, B. C. Genome duplication and the evolution of

919   physiological responses to water stress. *New Phyt.* **184**, 721–731 (2009).

920   74.   Fawcett, J. A. & Van de Peer, Y. Angiosperm polyploids and their road to evolutionary

921   success. *Trends Evol. Biol.* **1**, e3 (2010).

922   75.   Novikova, P. Y. *et al.* Polyploidy breaks speciation barriers in Australian burrowing frogs

923   *Neobatrachus*. *PLoS Genet.* **16**, e1008769 (2020).

924   76. Scarpino, S. V., Levin, D. A. & Meyers, L. A. Polyploid formation shapes flowering plant

925   diversity. *Am. Nat.* **184**, 456–465 (2014).

926   77. Meudt, H. M. *et al.* Is genome downsizing associated with diversification in polyploid

927   lineages ofVeronica? *Bot. J. Linn. Soc.* **178**, 243–266 (2015).

928   78. Schranz, M. E., Eric Schranz, M., Mohammadin, S. & Edger, P. P. Ancient whole genome

929   duplications, novelty and diversification: the WGD Radiation Lag-Time Model. *Curr. Opin.*

930   *Plant Biol.* **15**, 147–153 (2012).

931   79. Márquez-Corro, J. I., Escudero, M. & Luceño, M. Do holocentric chromosomes represent

932   an evolutionary advantage? A study of paired analyses of diversification rates of lineages

933   with holocentric chromosomes and their monocentric closest relatives. *Chromosome Res.*

934   **26**, 139–152 (2018).

935   80. Freyman, W. A. & Höhna, S. Cladogenetic and anagenetic models of chromosome number

936   evolution: a Bayesian model averaging approach. *Syst. Biol.* **67**, 195–215 (2018).

937   81. Zenil-Ferguson, R. *et al.* Interaction among ploidy, breeding system and lineage

938   diversification. *New Phytol.* **224**, 1252–1265 (2019).

939   82. Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R

940   language. *Bioinformatics* **20**, 289–290 (2004).

941   83. The Plant List. *The Plant List* http://www.theplantlist.org/.

942   84. Kluyver, T. A. & Osborne, C. P. Taxonome: a software package for linking biological

943   species data. *Ecol. Evol.* **3**, 1262–1265 (2013).

944   85. Angiosperm Phylogeny Website. http://www.mobot.org/MOBOT/research/APweb/.

945   86. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.*

946 **19**, 716–723 (1974).

87. Scholl, J. P. & Wiens, J. J. Diversification rates and species richness across the Tree of Life. *Proc. Biol. Sci.* **283**, (2016).

88. Pennell, M. W. *et al.* geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. *Bioinformatics* **30**, 2216–2218 (2014).

89. FitzJohn, R. G. Diversitree : comparative phylogenetic analyses of diversification in R. *Methods Ecol. Evol.* **3**, 1084–1092 (2012).

90. Schliep, K. P. phangorn: phylogenetic analysis in R. *Bioinformatics* **27**, 592–593 (2011).

91. Revell, L. J. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).

92. Rowan, T. H. Functional stability analysis of numerical algorithms. (1991).

93. King, A. A. *subplex*. (Github).

94. Huber, P. J. *Robust Statistics*. (John Wiley & Sons, Inc., 1981).

95. Venables, W. N. & Ripley, B. D. *Modern Applied Statistics with S*. (Springer-Verlag New York, 1994).

96. Lindeman, R. H., Merenda, P. F. & Gold, R. Z. *Introduction to Bivariate and Multivariate Analysis*. (Scott, Foresman, 1980).

97. Grömping, U. Relative importance for linear regression in R: the package *relaimpo*. *J. Stat. Soft.* **17** (2006).

98. Grömping, U. Estimators of relative importance in linear regression based on variance decomposition. *Am. Stat.* **61**, 139–147 (2007).

967