PROCEEDINGS B

royalsocietypublishing.org/journal/rspb

Research

Cite this article: Román-Palacios C, Molina-Henao YF, Barker MS. 2020 Polyploids increase overall diversity despite higher turnover than diploids in the Brassicaceae. *Proc. R. Soc. B* **287**: 20200962.
http://dx.doi.org/10.1098/rspb.2020.0962

Received: 28 April 2020
Accepted: 10 August 2020

Subject Category:
Evolution

Subject Areas:
evolution, plant science, ecology

Keywords:
diversification, evolution, ploidy, species richness, whole-genome duplications

Authors for correspondence:
Cristian Román-Palacios
e-mail: cromanpa94@email.arizona.edu
Y. Franchesco Molina-Henao
e-mail: yherson.molina@correounivalle.edu.co

†Both authors contributed equally to this work.

Electronic supplementary material is available online at https://dx.doi.org/10.6084/m9.figshare.c.5095302.

THE ROYAL SOCIETY PUBLISHING

Polyploids increase overall diversity despite higher turnover than diploids in the Brassicaceae

Cristian Román-Palacios[1,†], Y. Franchesco Molina-Henao[2,3,4,†] and Michael S. Barker[1]

[1]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA
[2]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA
[3]The Arnold Arboretum, Harvard University, Boston, MA 02131, USA
[4]Departamento de Biología, Universidad del Valle, Cali, Valle 760032, Colombia

CR-P, 0000-0003-1696-4886; YFM-H, 0000-0003-3375-6448; MSB, 0000-0001-7173-1319

Although polyploidy is widespread across the plant Tree of Life, its long-term evolutionary significance is still poorly understood. Here, we examine the effects of polyploidy in explaining the large-scale evolutionary patterns within angiosperms by focusing on a single family exhibiting extensive interspecific variation in chromosome numbers. We inferred ploidy from haploid chromosome numbers for 80% of species in the most comprehensive species-level chronogram for the Brassicaceae. After evaluating a total of 94 phylogenetic models of diversification, we found that ploidy influences diversification rates across the Brassicaceae. We also found that despite diversifying at a similar rate to diploids, polyploids have played a significant role in driving present-day differences in species richness among clades. Overall, in addition to highlighting the complexity in the evolutionary consequences of polyploidy, our results suggest that rare successful polyploids persist while significantly contributing to the long-term evolution of clades. Our findings further indicate that polyploidy has played a major role in driving the long-term evolution of the Brassicaceae and highlight the potential of polyploidy in shaping present-day diversity patterns across the plant Tree of Life.

1. Introduction

Although polyploidy—the heritable condition of carrying more than two complete sets of chromosomes—is widespread across the plant Tree of Life [1], its evolutionary significance is still debated [2–6]. Discussions on the evolutionary role of polyploidy date back to Stebbins [7,8] and Wagner [9], who considered polyploidy to have negligible effects on the long-term evolution of plants. This idea, later known as the 'dead end' hypothesis, was recently reframed as an expectation that polyploid species will undergo extinction more frequently than diploids [4,5,10–12]. However, polyploidy can influence the long-term evolution of lineages [1,13–17] regardless of whether polyploids are more likely to go extinct than diploids [4,5,10–12]. Both interpretations of the 'dead end' hypothesis are thus not equivalent to each other. The original hypothesis focused on the evolutionary role of polyploidy as a process [7–9], whereas the modern perspective compares macroevolutionary rates (i.e. speciation and extinction rates) between polyploids and diploids [4,5,10–12]. Discussions on the role of polyploidy in the evolution of plants have sometimes been obscured by the fact that there are at least two different interpretations of the 'dead end' hypothesis [11].

Although multiple studies have examined the association between extinction, speciation and diversification rates and ploidy [4,5,6,10–12], to our knowledge, whether or not polyploids influence differences in species richness among clades is still an open question [11]. Here, we evaluate the relationship between polyploidy, diversification rates and clade richness in plants [7–9]. Our analyses focused on the angiosperm family Brassicaceae, an ideal lineage for studying

the short- and long-term evolutionary significance of polyploidy in flowering plants. The systematics and evolutionary history of Brassicaceae have been extensively studied [18–20]. Furthermore, a comprehensive species-level phylogeny of the Brassicaceae including approximately 48% of the extant species (1667 out of approx. 3500 species) was recently published [21,22]. Haploid chromosome counts, which were used here to infer species ploidy, are available for approximately 55% of extant species within the family [21–24]. We used a phylogenetic model-based approach [25,26] to infer ploidy for 1333 species, representing more than 80% of the species in the tree and almost 40% of the total family richness. We then used the compiled ploidy database, along with the Brassicaceae phylogeny, to examine the role of ploidy in driving macroevolutionary rates within the family.

Using the Brassicaceae, we tested if polyploid species have significantly contributed to the short- (i.e. differences in diversification rates among species) and long-term (i.e. differences in species richness among clades) evolution of plants, despite previous evidence that polyploids are more likely to go extinct than diploids [4,5,10–12]. Although these previous studies compared the average net diversification rates of diploid and polyploid species, they did not examine the combined impacts of clade age and diversification at both ploidal levels on clade richness. We used recently developed phylogenetic methods to test the relationship between diversification rates and ploidy. Specifically, we analysed the fit of 26 different trait-dependent and trait-independent models of diversification that accounted for the potential effects of unassessed traits (i.e. hidden states; Hidden State Speciation and Extinction models [27]) or not (Binary State Speciation and Extinction models [28,29]). Then, we examined the role of polyploidy in driving long-term evolution across clades by testing if polyploidy has significantly contributed to the present-day differences in species richness among clades (i.e. genera). Using phylogenetic path regression analyses [30] of 65% of the total genera in the family (243 out of 372), we tested the contribution of polyploid richness on differences in species richness among clades in the Brassicaceae. We compared 68 models that assumed clade richness to be affected only by diversification rates or clade age to alternative models assuming ploidy influencing clade richness through net rates of diversification. With these species- and clade-level analyses in a large and well-sampled phylogeny of Brassicaceae, we provide one of the most comprehensive tests of the relationship between polyploidy and plant diversity.

2. Material and methods

(a) Phylogeny and ploidy inference

Our analyses are based on the Brassicaceae subtree from the GBMB Angiosperm phylogeny from Smith and Brown [22]. The Brassicaceae phylogeny was extracted using the extract.clade function in the ape R package version 5.2 [31] (the resulting tree is provided in electronic supplementary material, data S1). Additional details on this phylogeny are provided in electronic supplementary material, text S2.1. Chromosome numbers for each species in the tree were retrieved from the Chromosome Count Database [23] using the chromer R package v. 0.1 [32]. Raw haploid chromosome counts are provided in electronic supplementary material, table S1.

Chromosome counts were then used to infer ploidy for most species in the tree. Species were coded as being either diploid or polyploid (e.g. triploids, tetraploids) based on two different approaches. First, species ploidy was inferred with a maximum-likelihood approach using the Ploidy Inference Pipeline (PIP) implemented in ChromEvol v. 2.0 [25,26]. Alternatively, we used a more conservative approach [10,33] for estimating ploidy from haploid chromosome counts. Results for this alternative non-model-based method were congruent with those based on ChromEvol and are presented in the supplemental results (see electronic supplementary material, text S1).

Model-based ploidy inference analyses in ChromEvol failed to run on the full Brassicaceae phylogeny consisting of 1667 species, and with chromosome counts ranging between 4 and 156 (electronic supplementary material, table S1). To resolve this issue, we used an alternative approach in ChromEvol to infer ploidy for the maximum number of species included in our tree (1667 species). First, we estimated the median chromosome number across all the available counts for each species (see also [34]). Second, we extracted all possible subtrees from the Brassicaceae phylogeny using the subtrees function implemented in the ape R package. We then selected subtrees with (i) sizes between 25 and 550 species, and (ii) a maximum of 70% of missing data (i.e. species in tree lacking chromosome counts). This produced 342 subtrees that we used to run the Ploidy Inference Pipeline (PIP) and infer ploidy. The newly developed R function used to partition the Brassicaceae phylogeny into optimal subtrees for ChromEvol analyses is provided in electronic supplementary material, data S3. All ChromEvol-related files are provided in electronic supplementary material, data S2. We used the default parameters for each ChromEvol run (input files are provided in electronic supplementary material, data S2).

We acknowledge that statistical power might have compromised our ChromEvol analyses [35]. The ploidy inference under ChromEvol run on trees with a median tip number of 98.5 (mean = 118 tips, range = 30–302 tips; 148 subtrees or 43% of 342 subtrees). We note, however, that species-level sampling among subtrees partially overlapped. Specifically, 142 of the 148 subtrees shared at least one species. Thus, we examined congruence in ploidy inference under ChromEvol by assessing whether species shared among different subtrees were consistently inferred under the same ploidy state. We found that ploidy was ambiguous for only 1.9% of species that overlapped in at least two subtrees (25 of 1333 taxa). For these 25 species, different ploidy states were inferred in >60% of the subtrees where the species were sampled. In short, our ploidy inference approach under ChromEvol based on overlapping subtrees extracted from the Brassicaceae phylogeny indicates that ploidy was consistently inferred for approximately 98% of the species regardless of the analysed subtree.

Alternatively, we inferred species ploidy relative to their generic base following Stebbins [31] and others [10,36,37]—an estimate sometimes called the 'Stebbins fraction'. For this, we coded species as polyploid when the haploid count was greater than or equal to 3.5 times the lowest haploid (n) count of the corresponding genus (electronic supplementary material, figure S1 and table S2). Additional details are provided in electronic supplementary material, text S2.2.

(b) Testing for phylogenetic signal of polyploidy

Ploidy was coded as a binary trait with species in the tree being either diploid or polyploid (e.g. including triploid, tetraploid and so on). We analysed the phylogenetic signal of ploidal level variation across the Brassicaceae phylogeny using Pagel's lambda [38] and the D-statistics [39]. The first index measures the fit of the data to a Brownian motion model in which trait evolution matches the phylogeny. Next, we used the phylosig function in Phytools version 0.6–60 [40] to estimate the lambda value of ploidal variation across the tree. We also used the D-statistics to measure the phylogenetic signal of binary traits (such as polyploid versus diploid). We estimated the phylogenetic signal

under the *D*-statistics using the phylo.d function in the Caper R package version 1.0.1 [41]. For both analyses, the ploidy database is provided in electronic supplementary material, table S2 and the phylogeny in electronic supplementary material, data S4.

(c) Species-level diversification analyses

We used Binary State Speciation and Extinction models (BiSSE [28,29]) and Hidden State Speciation and Extinction models (HiSSE [27]) to examine the importance of ploidy changes on diversification of the Brassicaceae. BiSSE models were fitted using the diversitree R package version 0.9–10 [42]. HiSSE models were fitted using the hisse R package version 1.8.9 [27,43]. Species were coded as being either diploid (0) or polyploid (1) based both on the Stebbins fraction and ChromEvol-inferred ploidy states.

We accounted for incomplete state-specific sampling in all the fitted SSE models using three different estimates of the frequency of polyploidy within the Brassicaceae. First, we used our newly generated datasets to estimate the frequency of polyploidy within the Brassicaceae. Specifically, we assumed 49% of taxa within the family to be polyploids under the ChromEvol dataset and 26% under the Stebbins fraction dataset. As an alternative, we used the frequency of 33.58% polyploids estimated by Wood *et al*. [10]. Nevertheless, because results were congruent among datasets and sampling fractions, we only focus on those based on the ChromEvol dataset with state-frequency corrections from the same dataset.

For each dataset (ChromEvol and Stebbins fraction) and sampling fraction estimate (Wood *et al*. [10] and those from our study), we fitted a total of 13 different trait-dependent and independent diversification models (both BiSSE and HiSSE). Details on the parameters set for each of these models are provided in electronic supplementary material, table S3. We compared the fit among the analysed models using Akaike information criterion on (AIC) values [44] (electronic supplementary material, table S3). Finally, we summarized rates of speciation, extinction, turnover, and net diversification at the tips of the tree using the utilhisse R package [45]. The species-level phylogeny analysed in HiSSE and BiSSE is provided in electronic supplementary material, data S4. Similarly, the ploidy dataset is provided in electronic supplementary material, table S2. Results based on the Stebbins fraction dataset are presented in electronic supplementary material, tables S3 and S4, and text S1. Additional details are provided in electronic supplementary material, text S2.3.

(d) Clade-level diversification analyses

We analysed the role of polyploid species in influencing species richness among Brassicaceae genera using phylogenetic regressions [41] and phylogenetic path analyses [30,46]. We first constructed a generic-level phylogeny for the family by pruning from each genus in tree, all species except for one. This step was conducted using the drop.tip function in the ape R package. The resulting generic-level tree (one species per genus) is provided in electronic supplementary material, data S5.

Next, we constructed a database summarizing clade-level diversification rates (two different estimators; see below), species richness, clade age (stem and crown) and the proportion of polyploid species within clades. Net diversification was estimated for the crown and stem groups after assuming three different relative extinction fractions (e) using the Method-of-Moments estimator [47] (MS hereafter; $e = 0$, $e = 0.5$, and $e = 0.9$). MS rates were estimated using the bd.ms function implemented in the geiger R package version 2.0.6.1 [48]. We note 51% of Brassicaceae genera are excluded from crown-based analyses. Nevertheless, stem-based analyses are based on all the 243 genera in a tree. We also estimated species-specific rates of diversification (potentially reflecting speciation rates [49]) based on the DR statistic [50]. We estimated species-specific rates based on the full species-level phylogeny of the Brassicaceae (1667 species). Species-specific DR rates were then used to estimate the mean DR within genera ($n = 243$). We summarized the richness of polyploids within each clade using three different indexes (see details in electronic supplementary material, text S2.4). The analysed database including clade age, richness, net diversification rates and proportion of polyploids is provided in electronic supplementary material, table S6.

Finally, we tested whether diversification rate differences among genera are influenced by polyploidy using phylogenetic regressions. Regression models were fitted using the caper R package [40]. We allowed lambda to be estimated from the dataset. Additionally, phylogenetic path analyses were used to test for the indirect effect of polyploidy in explaining differences in species richness among clades. Details on the analysed models are provided in electronic supplementary material, text S2.4. Phylogenetic path models were compared using a modified version of the AIC that was developed for phylogenetic path analyses. This index, known as the C Statistic Information Criterion (i.e. CIC statistic [30]), is also calculated in the phylopath R package. Additional details are provided in electronic supplementary material, text S2.4. Finally, we examined whether the sign of the relationship between polyploid richness and diversification was influenced by confounding factors in our phylogenetic path analysis (Simpson's paradox [51,52]). For this, we simply fitted ordinary least square (OLS hereafter) regression models between polyploid richness and diversification rates using the lm function in R [53]. If alternative variables such as diploid richness and clade age did not influence the relationship between polyploid richness and diversification, we expected to find the same sign in the slope between polyploid richness and diversification in PGLS (coefficient phylogenetic paths) and OLS models [52].

3. Results

We collected haploid chromosome counts for 816 Brassicaceae species representing 49% of taxa sampled in the phylogeny [22] and 24% of the total family richness (electronic supplementary material, data S1 and table S1 [21]). We used this database to infer ploidy for 1333 species in the tree (80% of the species in the tree and 38% of the total family richness; figure 1; electronic supplementary material, table S2) using the likelihood-based approach implemented in ChromEvol [24,25] (electronic supplementary material, data S2). Overall, we inferred that nearly half of the analysed Brassicaceae species were polyploids ($n = 654$) and that these lineages were phylogenetically clustered in particular branches of the Brassicaceae phylogeny (Pagel's lambda = 0.816, $p < 0.001$; *D*-statistic = 0.006, $p < 0.001$; figure 1). Alternative analyses of polyploid richness based on ploidy estimates from multiples of putative base numbers (Stebbins fraction [10,33]) yielded largely similar results (see electronic supplementary material, text S1).

Using ploidal inferences, we found that changes in ploidy were important drivers of diversification rates across the Brassicaceae. We compared the fit of 11 different character-dependent models of diversification under HiSSE ($n = 7$; electronic supplementary material, table S3) and BiSSE models ($n = 4$; electronic supplementary material, table S3) against two null models that assumed diversification rates to be independent of ploidy changes. Results were largely robust to the analysed ploidy dataset (ChromEvol and Stebbins fraction) as well as the selected sampling fraction for ploidy used to account for the incomplete sampling in our dataset (Wood *et al*. [10] and estimates from this study). Main results are shown for the ChromEvol dataset

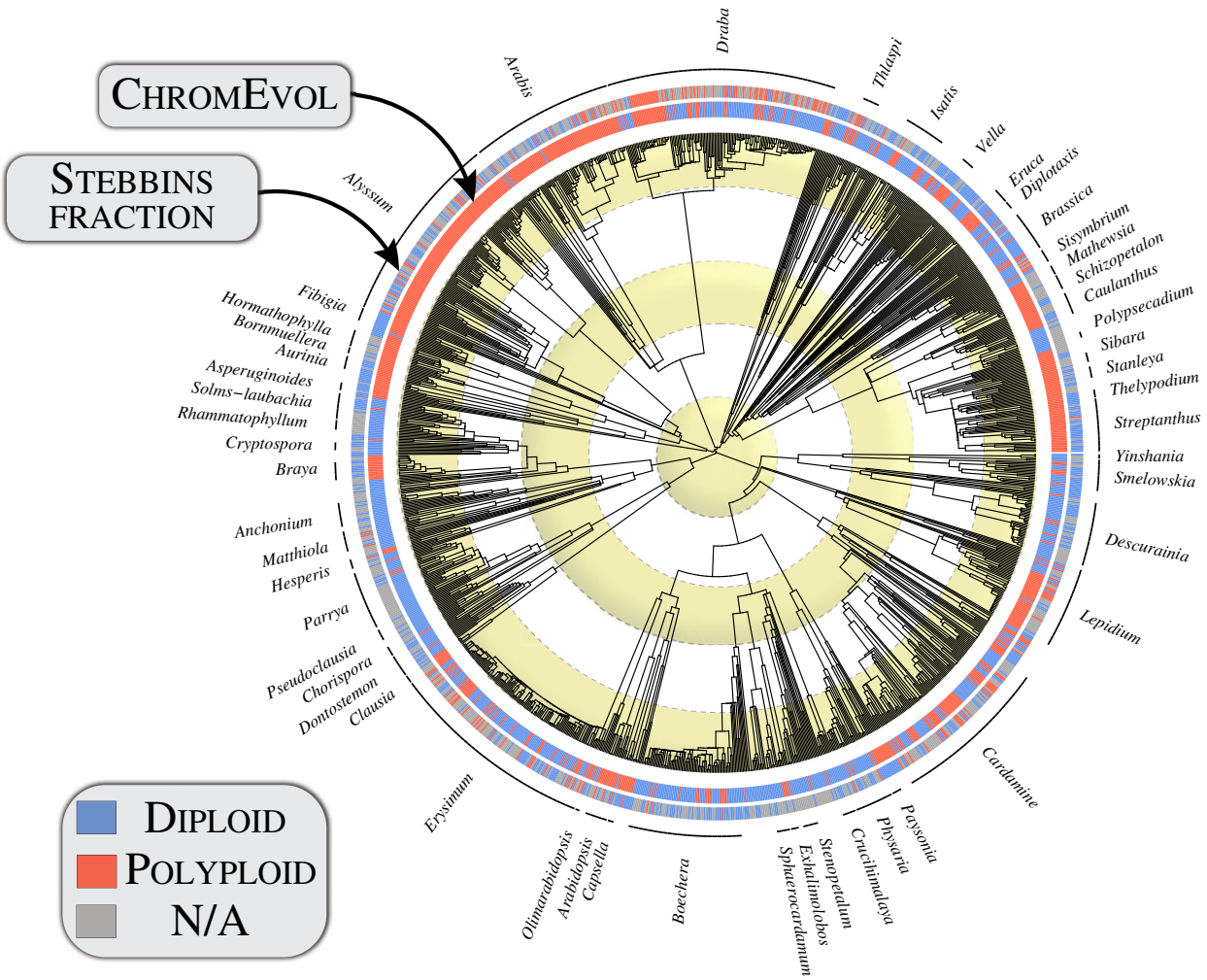

Figure 1. Phylogeny depicting the temporal evolution of ploidy across 1333 Brassicaceae species representing nearly 40% of the total family richness. Species are here coded as diploid or polyploid based on the evolution of haploid chromosome numbers across the tree. Ploidy was inferred using the ploidy inference pipeline (PIP) implemented in ChromEvol. We also show the ancestral state reconstruction of ploidy based on the best-fitting HiSSE model. We also depict variation in net diversification rates across branches of the phylogeny—fast-evolving branches are marginally highlighted in black. Ancestral state reconstructions shown here are for visualization of diversification rates and not used to ploidy changes across the tree. Instead, the ancestral state reconstruction for chromosomal numbers across the Brassicaceae is based on ChromEvol. Rings in the background are placed each 5 million years.

with state-level corrections based on the same dataset. Full results are shown in electronic supplementary material, tables S3–S5.

Overall, we found that changes in ploidal level were important drivers of diversification rates across the Brassicaceae phylogeny (figure 1; a full HiSSE model fitted better our dataset, next model ΔAICc = 9.786; electronic supplementary material, table S3). Based on the best-fitting model, we found similar rates of diversification (figure 2; mean diversification rate polyploids = 0.199 events Myr^{-1}, diploids = 0.217 events Myr^{-1}; electronic supplementary material, table S4) and turnover for polyploids and diploids (mean turnover rate polyploids = 1.927 events Myr^{-1}, diploids = 1.214 events Myr^{-1}). Nevertheless, we note that hidden state B significantly influences differences in diversification rates between diploids and polyploids by doubling diversification in diploids (figure 2; electronic supplementary material, table S5). Overall, in addition to recovering the importance of ploidy as an important driver of diversification rate differences across the Brassicaceae phylogeny, our analyses at the species level also highlight the complexity of the long-term evolutionary consequences of polyploidy in the evolution of the family.

Although these analyses provide insight into the different evolutionary dynamics of polyploid and diploid species, it is still unclear whether diversity generated via polyploidy significantly contributed to present-day differences in species richness across clades in the Brassicaceae. We used clade-level estimates of diversification rates to explore this question. These estimates of diversity were based on the MS estimator (electronic supplementary material, table S6) and the DR statistic (electronic supplementary material, table S7). We note, however, that DR estimates have been shown to actually be closer to speciation and not diversification rates [49]. Furthermore, the crown-based MS results excluded nearly 50% of genera in the phylogeny (see Methods). The main results we show for phylogenetic regressions and path analyses are therefore shown for MS stem-based rates of diversification that include 97% of genera in the phylogeny and 65% of the total generic richness within the family (electronic supplementary material, table S6).

Next, we used phylogenetic generalized least-squares models to test whether polyploid richness influences species richness through diversification rates. Strikingly, we found that 16–30% of the variance in net diversification rates was explained by polyploid richness alone (log-transformed

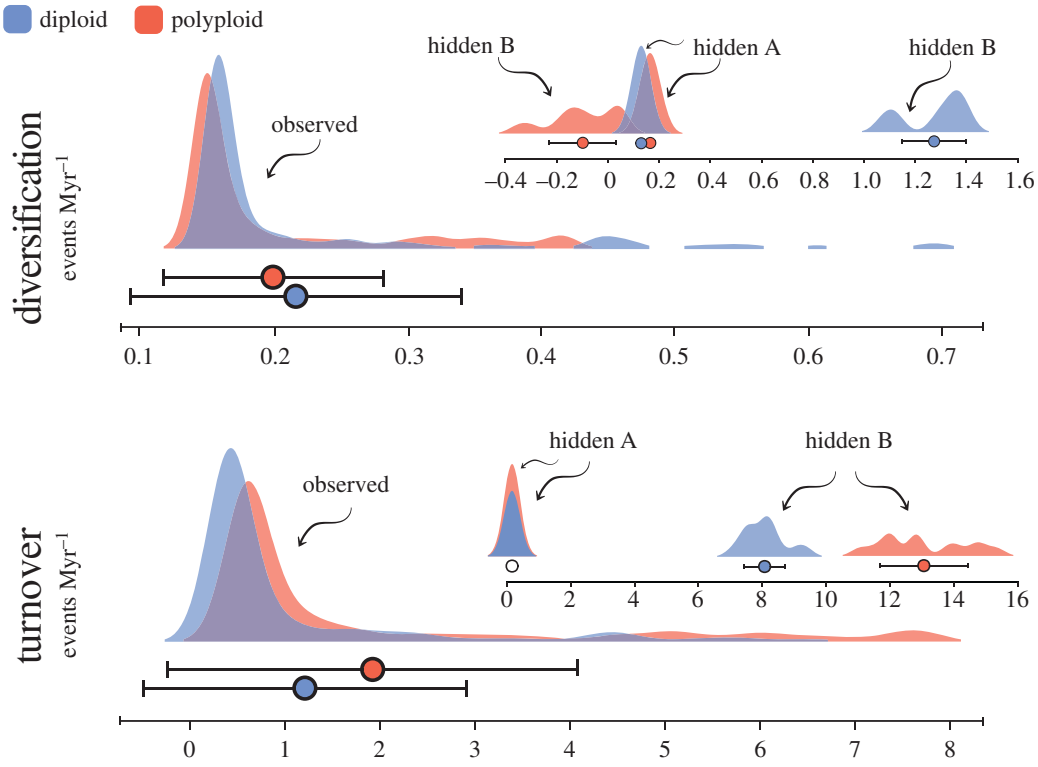

Figure 2. Macroevolutionary rates associated with polyploids and diploids. We show diversification (top) and turnover rates (bottom) for each diploids (blue) and polyploids (red) based on the best-fitting HiSSE model. We also present diversification and turnover rates for the hidden states (A and B) within each of the observed states (diploids and polyploids), as insets of the corresponding figure panels. Ploidy is based on the ChromEvol dataset with incomplete sampling corrections estimated from the same dataset. Rates for observed states were estimated using the utilhisse R package. We show mean (circles), minimum and maximum values for the rates within each state. Full results are presented in electronic supplementary material, table S4. Units for both diversification and turnover rates in events Myr^{-1}.

polyploid richness versus net diversification rates; PGLS slope = 0.085–0.444, $r^2 = 0.166$–0.304, $p < 0.001$; electronic supplementary material, table S7). We also analysed PGLS regressions that were fitted using alternative descriptors of polyploid richness within clades that were either unrelated (log-transformed proportion of polyploid richness in described richness versus net diversification rates, PGLS $r^2 = 0.001$–0.003, $p > 0.4$; electronic supplementary material, table S8) or weakly associated with net diversification rates (log-transformed proportion of polyploid richness in tree versus net diversification rates, $r^2 = 0.016$–0.017, all $p = 0.04$). Second, we used 14 phylogenetic path regression models (the additional models for the Stebbins fraction dataset, under crown-based MS and DR, are shown in electronic supplementary material, table S9) to test whether polyploid richness has indirectly shaped present-day patterns of species richness among clades through diversification rates (figure 3; electronic supplementary material, table S9). Based on the best-fitting model—a model accounting for the indirect effects of diploid and polyploid richness and the direct effect of diversification and clade age on richness (next model ΔCICc = 21.021; electronic supplementary material, table S9)—we found that polyploidy is an indirect driver of differences in species richness among clades across the Brassicaceae (figure 3). The positive association between polyploidy and diversification was consistently recovered using OLS models indicating the absence of the Simpson's paradox in our dataset (electronic supplementary material, table S10). In sum, despite diploids contributing four times more than polyploids to net diversification rates, polyploidy-generated diversity has significantly influenced differences in species richness among clades through positive net diversification rates.

4. Discussion

Our analyses revealed the complex interactions of polyploidy and diversification in flowering plants. First, leveraging one of the largest species-level phylogenies for the Brassicaceae with ploidal level inferences for nearly 40% of all extant species from 65% of genera in the family, we found that polyploidy is not rare in the family (26–49% of species are polyploids, range based on Stebbins fraction and ChromEvol). These values are consistent with other estimates of polyploid incidence for angiosperms [4,10,54]. Second, the results of our phylogenetic analyses of polyploidy and diversification using either estimates were consistent with each other. We found that polyploid species in the Brassicaceae experience slightly higher turnover rates than diploids [4,5]. However, our taxonomically wider analyses of the differences in species richness among clades found that polyploidy significantly influenced present-day differences in diversity in the Brassicaceae. In particular, the best-fitting model in a phylogenetic path analysis indicated that polyploid richness had a positive indirect effect on overall net diversification rates. Nevertheless, polyploids contributed less to richness differences among clades than diploids. Overall, while most previous research on this issue has focused only on the macroevolutionary differences between diploid and polyploid species [4,10,12,55–57], our analyses quantified the overall contributions of these differences to reveal that

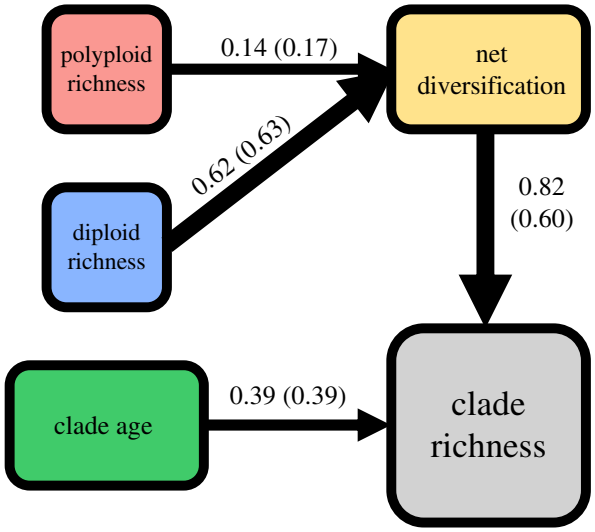

Figure 3. Phylogenetic path analysis showing the main drivers of differences in species richness among clades in the Brassicaceae. Results are based on 65% of the total generic richness within the family. We show results for the best-fitting phylogenetic path model. This model included the indirect effects of polyploid and diploid richness on species richness. We also note that the selected model assumed clade richness to be directly influenced by clade age and net rates of diversification. Here, polyploid richness is simply the number of polyploid species within each clade (based on ChromEvol analyses). We note that, contrary to the frequency of polyploidy, polyploid raw richness was a significant predictor of diversification rates explaining 16–30% of the variance. We present results for both stem-based MS and DR rates. We acknowledge that DR rates are usually referred to as related to speciation, but results are congruent to those based on MS. Path coefficient for the MS estimator is indicated below each arrow outside of parentheses. DR-related coefficients are indicated next to MS-related coefficients in parentheses. Finally, the clade age is always based on stem. (Online version in colour.)

polyploidy has a direct, positive impact on overall net diversification rates and an indirect effect on differences in clade richness across the Brassicaceae.

If polyploidy has a positive effect on the diversity of the Brassicaceae, should it really be considered an evolutionary 'dead end'? The traditional perspective that polyploid plant species are evolutionary 'dead ends' dates back to Stebbins [7,8] and Wagner [9]. These authors suggested that polyploidy in plants, although common, probably does not contribute significantly to the long-term evolution of lineages. However, recent studies [4,5] have used this phrase to describe the macroevolutionary differences in extinction rates among polyploids and diploids. Although extinction rates are ultimately related to the long-term significance of polyploidy, the contribution of polyploid species to present-day diversity patterns should also be measured in terms of net diversification, turnover rates and their actual effects on driving richness patterns across clades. Our finding that polyploidy positively influenced diversification rates and species richness within the Brassicaceae indicates that polyploidy is not an evolutionary 'dead end' in the classical sense. Additionally, our analyses suggested that non-assessed traits have played a significant role in driving differences in diversification rates between and within diploids and polyploids (see electronic supplementary material, table S5). Specifically, differences in diversification rates are also affected by a second character (hidden state B) that either increases diversification rates in diploids or slows down the same rates in polyploids. This second character may explain why some but not all polyploids influence diversity patterns within the Brassicaceae. Even in a family like the Brassicaceae where polyploid species have lower net diversification rates, polyploidy has an overall positive effect on diversification and species richness.

Despite our finding that the contribution of polyploid species to clade richness was significant, diploids contributed more than polyploids to variation in clade richness within the Brassicaceae. This likely reflects the complexity of evolution and persistence at higher ploidal levels. Previous studies have discussed the evolutionary trade-offs associated with polyploidy [58–62]. Specifically, polyploids tend to speciate faster than diploids but also to experience faster rates of extinction (see electronic supplementary material, table S4). On the one hand, faster speciation rates are potentially related to the rate of the successful establishment of polyploids to novel habitats [63–72], or even to their ability to mask deleterious mutations [58–62,73]. On the other hand, polyploidy may trigger extinction rates by decreasing effective population sizes [74,75], increasing rates of segregation errors during meiosis [76] or even by increasing the number of deleterious mutations that may contribute to extinction during diploidization [5,75]. In fact, despite widespread evidence of diploidization being weak at shorter time scales (<50 Myr [57]), we show that diploidization is not effectively zero in the Brassicaceae and that this process has significantly influenced diversification rates within the family (see electronic supplementary material, table S3).

The scale of our analyses may also play a role in the measured differences of net diversification rates between diploid and polyploid species. Specifically, both the positive correlation between polyploidy and diversification at deep time scales [6] (but see recent methodological caveats in [77]) in addition to the observation that ancient polyploidy is correlated with mass extinction events and major geological transitions [76,78] suggest that polyploidy may facilitate lineage survival during periods of rapid ecological change. Our analyses within the Brassicaceae are not deep enough in time to capture that scale of ecological change. Further, the methods do not yet exist to formally estimate rates of polyploidy and diversification from inferences based on chromosome counts at the tips of a phylogeny with inferences of ancient whole-genome duplications from genomic data deeper in a phylogeny. Progress on developing these types of integrated analytical frameworks and denser data sampling will permit further testing of the hypotheses on the role of whole-genome duplications in the evolution of plant diversity.

5. Conclusion

Our results recovering polyploidy as a major driver of diversification and species richness differences among clades shows that polyploids are not evolutionary 'dead ends' in the Brassicaceae. Our findings are potentially extended to other branches of the plant Tree of Life that exhibit similar levels of polyploidy (e.g. Asteraceae, Poaceae, Solanaceae, Fabaceae, Cleomaceae [10,13,79–82]) and ratios of polyploid and diploid net diversification rates [4]. The influence of ploidy in the long-term evolution is potentially not only restricted to the Brassicaceae but likely to drive diversity in other lineages of green plants [6]. Polyploidy has many impacts on the population genetics and ecology of

species [5,36,37,83] that produce a complex association between whole-genome duplications and macroevolutionary changes in diversity. Further quantitative studies are required to examine the specific relationship between diversification rates, clade richness and ploidy across all green plants.

Data accessibility. All data and code are available in the electronic supplementary materials. Electronic supplementary material, data S1–S5 are available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.31zcrjdhw [84].

Author contributions. C.R.P., Y.F.M.-H. and M.S.B. designed research; C.R.P. and Y.F.M.-H. performed research and analysed data, and C.R.P., Y.F.M.-H. and M.S.B. wrote the paper.

Competing interests. We declare we have no competing interests.

Funding. Y.F.M.-H. was supported by a Graduate Fellowship from the Department of Organismic and Evolutionary Biology at Harvard University; the Colombian Administrative Department of Science, Technology, and Innovation (COLCIENCIAS) 538-2012; and the Fulbright Exchange Visitor Program G-1-12218. M.S.B. was supported by National Science Foundation (NSF) IOS-1339156 and NSF EF-1550838.

Acknowledgements. We thank Jeremy M. Beaulieu for help with HiSSE, Zheng Li and Elizabeth C. Miller for discussions. Elena M. Kramer provided comments on an early version of the manuscript. This material is based upon High-Performance Computing (HPC) resources supported by the University of Arizona TRIF, UITS, and RDI and maintained by the UA Research Technologies department. Additional hosting infrastructure and services were provided by the FAS Research Computing Facility at Harvard University.

References

1. Leebens-Mack JH *et al.* 2019 One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **574**, 679–685. (doi:10.1038/s41586-019-1693-2)
2. Soltis DE *et al.* 2009 Polyploidy and angiosperm diversification. *Am. J. Bot.* **96**, 336–348. (doi:10.3732/ajb.0800079)
3. Van de Peer Y, Maere S, Meyer A. 2009 The evolutionary significance of ancient genome duplications. *Nat. Rev. Genet.* **10**, 725–732. (doi:10.1038/nrg2600)
4. Mayrose I, Zhan SH, Rothfels CJ, Magnuson-Ford K, Barker MS, Rieseberg LH, Otto SP. 2011 Recently formed polyploid plants diversify at lower rates. *Science* **333**, 1257–1257. (doi:10.1126/science.1207205)
5. Arrigo N, Barker MS. 2012 Rarely successful polyploids and their legacy in plant genomes. *Curr. Opin Plant Biol.* **15**, 140–146. (doi:10.1016/j.pbi.2012.03.010)
6. Landis JB, Soltis DS, Li Z, Marx HE, Barker MS, Tank DC, Soltis PS. 2018 Impact of whole-genome duplication events on diversification rates in angiosperms. *Am. J. Bot.* **105**, 348–363. (doi:10.1002/ajb2.1060)
7. Stebbins GL. 1950 *Variation and evolution in plants*. New York, NY: Columbia University Press.
8. Stebbins GL. 1970 Variation and evolution in plants: Progress during the past twenty years. In *Essays in evolution and genetics in honor of Theodosius Dobzhansky: a supplement to evolutionary biology* (eds MK Hecht, WC Steere), pp. 173–208. Berlin, Germany: Springer.
9. Wagner W. 1970 Biosystematics and evolutionary noise. *Taxon* **19**, 146–151. (doi:10.2307/1217945)
10. Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH. 2009 The frequency of polyploid speciation in vascular plants. *Proc. Natl Acad. Sci. USA* **106**, 13 875–13 879. (doi:10.1073/pnas.0811575106)
11. Mayrose I, Zhan SH, Rothfels CJ, Arrigo N, Barker MS., Rieseberg LH, Otto SP. 2015 Methods for studying polyploid diversification and the dead end hypothesis: a reply to Soltis *et al.* (2014). *New Phytol.* **206**, 27–35. (doi:10.1111/nph.13192)
12. Estep MC *et al.* 2014 Allopolyploidy, diversification, and the Miocene grassland expansion. *Proc. Natl Acad. Sci. USA* **111**, 15 149–15 154. (doi:10.1073/pnas.1404177111)
13. Paterson AH, Bowers JE, Chapman BA. 2004 Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proc. Natl Acad. Sci. USA* **101**, 9903–9908. (doi:10.1073/pnas.0307901101)
14. De Bodt S, Maere S, Van de Peer Y. 2005 Genome duplication and the origin of angiosperms. *Trends Ecol. Evol.* **20**, 591–597. (doi:10.1016/j.tree.2005.07.008)
15. Jaillon O *et al.* 2007 The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* **449**, 463–467. (doi:10.1038/nature06148)
16. Fawcett JA, Maere S, Van de Peer Y. 2009 Plants with double genomes might have had a better chance to survive the Cretaceous–Tertiary extinction event. *Proc. Natl Acad. Sci. USA* **106**, 5737–5742. (doi:10.1073/pnas.0900906106)
17. Jiao Y *et al.* 2011 Ancestral polyploidy in seed plants and angiosperms. *Nature* **473**, 97–100. (doi:10.1038/nature09916)
18. Al-Shehbaz IA, Beilstein MA, Kellogg EA. 2006 Systematics and phylogeny of the Brassicaceae (Cruciferae): an overview. *Plant Syst. Evol.* **259**, 89–120. (doi:10.1007/s00606-006-0415-z)
19. Beilstein MA, Al-Shehbaz IA, Kellogg EA. 2006 Brassicaceae phylogeny and trichome evolution. *Am. J. Bot.* **93**, 607–619. (doi:10.3732/ajb.93.4.607)
20. Hohmann N, Wolf EM, Lysak MA, Koch MA. 2015 A time-calibrated road map of Brassicaceae species radiation and evolutionary history. *Plant Cell* **27**, 2770–2784.
21. The Plant List. 2013 A working list for all plant species. See http://www.theplantlist.org/ (accessed 10 January 2018).
22. Smith SA, Brown JW. 2018 Constructing a broadly inclusive seed plant phylogeny. *Am. J. Bot.* **105**, 302–314. (doi:10.1002/ajb2.1019)
23. Rice A, Glick L, Abadi S, Einhorn M, Kopelman NM, Salman-Minkov A, Mayzel J, Chay O, Mayrose I. 2015 The Chromosome Counts Database (CCDB): a community resource of plant chromosome numbers. *New Phytol.* **206**, 19–26. (doi:10.1111/nph.13191)
24. Lysak MA. 2018 Brassicales: an update on chromosomal evolution and ancient polyploidy. *Plant Syst. Evol.* **304**, 757–762. (doi:10.1007/s00606-018-1507-2)
25. Mayrose I, Barker MS, Otto SP. 2010 Probabilistic models of chromosome number evolution and the inference of polyploidy. *Syst. Biol.* **59**, 132–144. (doi:10.1093/sysbio/syp083)
26. Glick L, Mayrose I. 2014 ChromEvol: assessing the pattern of chromosome number evolution and the inference of polyploidy along a phylogeny. *Mol. Biol. Evol.* **31**, 1914–1922. (doi:10.1093/molbev/msu122)
27. Beaulieu JM, O'Meara BC. 2016 Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. *Syst. Biol.* **65**, 583–601. (doi:10.1093/sysbio/syw022)
28. Maddison WP, Midford PE, Otto SP. 2007 Estimating a binary character's effect on speciation and extinction. *Syst. Biol.* **56**, 701–710. (doi:10.1080/10635150701607033)
29. FitzJohn RG, Maddison WP, Otto SP. 2009 Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. *Syst. Biol.* **58**, 595–611. (doi:10.1093/sysbio/syp067)
30. Gonzalez-Voyer A, Von Hardenberg A. 2014 An introduction to phylogenetic path analysis. In *Modern phylogenetic comparative methods and their application in evolutionary biology* (ed LZ Garamszegi), pp. 201–229. Berlin, Germany: Springer.
31. Paradis E, Claude J, Strimmer K. 2004 APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290. (doi:10.1093/bioinformatics/btg412)
32. Pennell M. 2016 chromer: Interface to chromosome counts database API. R package version 0.1.2.9000. See https://github.com/ropensci/chromer.
33. Stebbins GL. 1938 Cytological characteristics associated with the different growth habits in the dicotyledons. *Am. J. Bot.* **25**, 189–198. (doi:10.1002/j.1537-2197.1938.tb09203.x)
34. Salman-Minkov A, Sabath N, Mayrose I. 2016 Whole-genome duplication as a key factor in crop

domestication. *Nat. Plants* **2**, 16115. (doi:10.1038/nplants.2016.115)

35. Zenil-Ferguson R, Burleigh JG, Ponciano JM. 2018 chromploid: an R package for chromosome number evolution across the plant tree of life. *Appl. Plant Sci.* **6**, e1037. (doi:10.1002/aps3.1037)
36. Klekowski Jr EJ. 1973 Sexual and subsexual systems in homosporous pteridophytes: a new hypothesis. *Am. J. Bot.* **60**, 535–544. (doi:10.1002/j.1537-2197.1973.tb05955.x)
37. Goldblatt P. 1980 Polyploidy in angiosperms: monocotyledons. In *Polyploidy* (ed. E Lewis), pp. 219–239. Berlin, Germany: Springer.
38. Pagel M. 1999 Inferring the historical patterns of biological evolution. *Nature* **401**, 877–884. (doi:10.1038/44766)
39. Fritz SA, Purvis A. 2010 Selectivity in mammalian extinction risk and threat types: a new measure of phylogenetic signal strength in binary traits. *Conserv. Biol.* **24**, 1042–1051. (doi:10.1111/j.1523-1739.2010.01455.x)
40. Revell LJ. 2012 phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223. (doi:10.1111/j.2041-210X.2011.00169.x)
41. Orme D. 2013 The caper package: comparative analysis of phylogenetics and evolution in R. *R Package Version* **5**, 1–36.
42. FitzJohn RG. 2012 *Diversitree*: comparative phylogenetic analyses of diversification in R. *Methods Ecol. Evol.* **3**, 1084–1092. (doi:10.1111/j.2041-210X.2012.00234.x)
43. Beaulieu JM, O'Meara B. 2019 HiSSE: hidden state speciation and extinction. R package version 1.8. see https://cran.r-project.org/web/packages/hisse/index.html.
44. Akaike H. 1974 A new look at the statistical model identification. *IEEE Trans. Autom. Control* **19**, 716–723. (doi:10.1109/TAC.1974.1100705)
45. Nakov T, Beaulieu JM, Alverson AJ. 2019 Diatoms diversify and turn over faster in freshwater than marine environments. *Evolution* **73**, 2497–2511. (doi:10.1111/evo.13832)
46. van der Bijl W. 2018 *phylopath*: easy phylogenetic path analysis in R. *PeerJ* **6**, e4718. (doi:10.7717/peerj.4718)
47. Magallón S, Sanderson MJ. 2001 Absolute diversification rates in angiosperm clades. *Evolution* **55**, 1762–1780. (doi:10.1111/j.0014-3820.2001.tb00826.x)
48. Pennell M, Eastman JM, Slater GJ, Brown JW, Uyeda JC, FitzJohn RG, Alfaro ME, Harmon LJ. 2014 geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. *Bioinformatics* **30**, 2216–2218. (doi:10.1093/bioinformatics/btu181)
49. Title PO, Rabosky DL. 2019 Tip rates, phylogenies and diversification: what are we estimating, and how good are the estimates? *Methods Ecol. Evol.* **10**, 821–834. (doi:10.1111/2041-210X.13153)
50. Jetz W, Thomas GH, Joy JB, Hartmann K, Mooers AO. 2012 The global diversity of birds in space and time. *Nature* **491**, 444–448. (doi:10.1038/nature11631)
51. Nee S, Read AF, Greenwood JJ, Harvey PH. 1991 The relationship between abundance and body size in British birds. *Nature* **351**, 312–313. (doi:10.1038/351312a0)
52. Uyeda JC, Zenil-Ferguson R, Pennell MW. 2018 Rethinking phylogenetic comparative methods. *Syst. Biol.* **67**, 1091–1109. (doi:10.1093/sysbio/syy031)
53. R Core Team. 2013 *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
54. Barker MS, Arrigo N, Baniaga AE, Li Z, Levin DA. 2016 On the relative abundance of autopolyploids and allopolyploids. *New Phytol.* **210**, 391–398. (doi:10.1111/nph.13698)
55. Scarpino SV, Levin DA, Meyers LA. 2014 Polyploid formation shapes flowering plant diversity. *Am. Nat.* **184**, 456–465. (doi:10.1086/677752)
56. Huang XC, German DA, Koch MA. 2020 Temporal patterns of diversification in Brassicaceae demonstrate decoupling of rate shifts and mesopolyploidization events. *Ann. Bot.* **125**, 29–47. (doi:10.1093/aob/mcz123).
57. Zenil-Ferguson R, Gordon Burleigh J, Freyman WA, Igić B, Mayrose I, Goldberg EE. 2019 Interaction between ploidy, breeding system, and lineage diversification. *New Phytol.* **224**, 1252–1265. (doi:10.1111/nph.16184)
58. Comai L. 2005 The advantages and disadvantages of being polyploid. *Nat. Rev. Genet.* **6**, 836–846. (doi:10.1038/nrg1711)
59. Otto SP. 2007 The evolutionary consequences of polyploidy. *Cell* **131**, 452–462. (doi:10.1016/j.cell.2007.10.022)
60. Sémon M, Wolfe KH. 2007 Consequences of genome duplication. *Curr. Opin. Genet. Dev.* **17**, 505–512. (doi:10.1016/j.gde.2007.09.007)
61. Madlung A. 2013 Polyploidy and its effect on evolutionary success: old questions revisited with new tools. *Heredity* **110**, 99–104. (doi:10.1038/hdy.2012.79)
62. Van De Peer Y, Mizrachi E, Marchal K. 2017 The evolutionary significance of polyploidy. *Nat. Rev. Genet.* **18**, 411–424. (doi:10.1038/nrg.2017.26)
63. Baniaga AE, Marx HE, Arrigo N, Barker MS. 2020 Polyploid plants have faster rates of multivariate niche evolution than their diploid relatives. *Ecol. Lett.* **23**, 68–78. (doi:10.1111/ele.13402)
64. Theodoridis S, Randin C, Broennimann O, Patsiou T, Conti E. 2013 Divergent and narrower climatic niches characterize polyploid species of European primroses in *Primula* sect. *Aleuritia*. *J. Biogeogr.* **40**, 1278–1289. (doi:10.1111/jbi.12085)
65. Harbert RS, Brown ADH, Doyle JJ. 2014 Climate niche modeling in the perennial *Glycine* (Leguminosae) allopolyploid complex. *Am. J. Bot.* **101**, 710–721. (doi:10.3732/ajb.1300417)
66. Thompson KA, Husband BC, Maherali H. 2014 Climatic niche differences between diploid and tetraploid cytotypes of *Chamerion angustifolium* (Onagraceae). *Am. J. Bot.* **101**, 1868–1875. (doi:10.3732/ajb.1400184)
67. Hulber K, Sonnleitner M, Suda J, Krejčíková J, Schönswetter P, Schneeweiss GM, Winkler M. 2015 Ecological differentiation, lack of hybrids involving diploids, and asymmetric gene flow between polyploids in narrow contact zones of *Senecio carniolicus* (syn. *Jacobaea carniolica*, Asteraceae). *Ecol. Evol.* **5**, 1224–1234. (doi:10.1002/ece3.1430)
68. Lopez-Alvarez *et al.* 2015 Environmental niche variation and evolutionary diversification of the *Brachypodium distachyon* grass complex species in their native circum-Mediterranean range. *Am. J. Bot.* **102**, 1073–1088. (doi:10.3732/ajb.1500128)
69. Zozomova-Lihova J, Malánová-Krásná I, Vít P, Urfus T, Senko D, Svitok M, Kempa M, Marhold K. 2015 Cytotype distribution patterns, ecological differentiation, and genetic structure in a diploid-tetraploid contact zone of *Cardamine amara*. *Am. J. Bot.* **102**, 1380–1395. (doi:10.3732/ajb.1500052)
70. Marchant DB, Soltis DE, Soltis PS. 2016 Patterns of abiotic niche shifts in allopolyploids relative to their progenitors. *New Phytol.* **212**, 708–718. (doi:10.1111/nph.14069)
71. Visger CJ, Germain-Aubrey CC, Patel M, Sessa EB, Soltis PS, Soltis DE. 2016 Niche divergence between diploid and autotetraploid *Tolmiea*. *Am. J. Bot.* **103**, 1396–1406. (doi:10.3732/ajb.1600130)
72. Molina-Henao YF, Hopkins R. 2019 Autopolyploid lineage shows climatic niche expansion but not divergence in *Arabidopsis arenosa*. *Am. J. Bot.* **106**, 61–70. (doi:10.1002/ajb2.1212)
73. Werth CR, Windham MD. 1991 A model for divergent, allopatric speciation of polyploid pteridophytes resulting from silencing of duplicate-gene expression. *Am. Nat.* **137**, 515–526. (doi:10.1086/285180)
74. Soltis DE, Buggs RJA, Doyle JJ, Soltis PS. 2010 What we still don't know about polyploidy. *Taxon* **59**, 1387–1403. (doi:10.1002/tax.595006)
75. Otto SP, Whitton J. 2000 Polyploid incidence and evolution. *Annu. Rev. Genet.* **34**, 401–437. (doi:10.1146/annurev.genet.34.1.401)
76. Vanneste K, Baele G, Maere S, Van de Peer Y. 2014 Analysis of 41 plant genomes supports a wave of successful genome duplications in association with the Cretaceous–Paleogene boundary. *Genome Res.* **24**, 1334–1347. (doi:10.1101/gr.168997.113)
77. Caetano DS, O'Meara BC, Beaulieu JM. 2018 Hidden state models improve state-dependent diversification approaches, including biogeographical models. *Evolution* **72**, 2308–2324. (doi:10.1111/evo.13602)
78. Cai L, Xi Z, Amorim AM, Sugumaran M, Rest JS, Liu L, Davis CC. 2019 Widespread ancient whole-genome duplications in Malpighiales coincide with Eocene global climatic upheaval. *New Phytol.* **221**, 565–576. (doi:10.1111/nph.15357)
79. Góralski G, Bal M, Gacek P, Orzechowski TM, Kosecka-Wierzejska A. 2014 Chromosome numbers

and polyploidy in life forms of Asteraceae, Poaceae and Rosaceae in Polish flora. *Acta Biol. Cracov., Ser. Bot.* **56**, 7–15.

80. Blanc G, Wolfe KH. 2004 Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell* **16**, 1667–1678. (doi:10.1105/tpc.021345)
81. Leitch IJ, Hanson L, Lim KY, Kovarik A, Chase MW, Clarkson JJ, Leitch AR. 2008 The ups and downs of genome size evolution in polyploid species of *Nicotiana* (Solanaceae). *Ann. Bot.* **101**, 805–814. (doi:10.1093/aob/mcm326)
82. Morales M, Wulff AF, Fortunato RH, Poggio L. 2014 Chromosome studies in southern species of *Mimosa* (Fabaceae, Mimosoideae) and their taxonomic and evolutionary inferences. *Plant Syst. Evol.* **300**, 803–817. (doi:10.1007/s00606-013-0920-9)
83. Baduel P, Bray S, Vallejo-Marin M, Kolář F, Yant L. 2018 The 'polyploid hop': shifting challenges and opportunities over the evolutionary lifespan of genome duplications. *Front. Ecol. Evol.* **6**, 117. (doi:10.3389/fevo.2018.00117)
84. Román-Palacios C, Molina-Henao YF, Barker MS. 2020 Data from: Polyploids increase overall diversity despite higher turnover than diploids in the Brassicaceae. Dryad Digital Repository. (doi:10.5061/dryad.31zcrjdhw)