CellPress
REVIEWS

Review

Whole-Genome Duplication and Plant Macroevolution

James W. Clark[1] and Philip C.J. Donoghue[1,*]

Whole-genome duplication (WGD) is characteristic of almost all fundamental lineages of land plants. Unfortunately, the timings of WGD events are loosely constrained and hypotheses of evolutionary consequence are poorly formulated, making them difficult to test. Using examples from across the plant kingdom, we show that estimates of timing can be improved through the application of molecular clock methodology to multigene datasets. Further, we show that phenotypic change can be quantified in morphospaces and that relative phenotypic disparity can be compared in the light of WGD. Together, these approaches facilitate tests of hypotheses on the role of WGD in plant evolution, underscoring the potential of plants as a model system for investigating the role WGD in macroevolution.

Highlights

WGD events continue to be discovered throughout plant systematic diversity as a consequence of genome sequencing programs.

The absolute timing of WGD events remains poorly constrained and poorly understood, but many hypotheses regarding the role of WGD in plant evolution depend on precise estimates.

The role of WGD in facilitating diversification has a strong theoretical basis but remains to be rigorously tested, and WGD in species-poor lineages cannot be ignored.

WGD as a driver of plant morphological diversity is an appealing hypothesis, but requires a framework which can quantify morphological variation between lineages and through time.

Whole-Genome Duplication

WGD encompasses multiple processes that lead to the formation of a polyploid organism with three or more sets of the base chromosome number. It has been invoked as a cause of macroevolutionary change [1], explaining everything from extinction resistance to fundamental evolutionary innovation. WGD has been proposed as a driver of diversity [2,3], herbivore interactions [4], geographic expansions [5], climatic niche shifts [6], and of facilitating lineage longevity [7]. Clustering of WGD events around the Cretaceous–Paleogene (K–Pg) interval has led to the hypothesis that genome duplication may have facilitated evolutionary success in the wake of the mass extinction event at the end of the Cretaceous [8,9] (Box 1). Equally, however, it is possible that the extensive history of WGD in plant evolution is incidental or inconsequential, and there are examples, such as mosses and horsetails [7,10], where a macroevolutionary-scale phenotypic impact is not evident. Ancient WGD events (paleopolyploidy) first appeared to be rare [11], but newly sequenced genomes have revealed duplication in an increasing diversity of plant lineages [12,13]. However, with few exceptions, it appears that most of the hypothesized macroevolutionary outcomes have neither been tested nor formulated as hypotheses that are readily testable, despite the diversity of comparative methods that are available for facilitating such tests. There are multiple emerging models explaining how complexity and novelty may arise through genome duplication [14], although fundamental questions remain as to why the outcomes of WGD are so disparate among lineages and whether the nature of the ploidy event influences the outcome (Box 2). Tests are necessary to quantify the macroevolutionary change in the wake of WGD, or else WGD risks becoming a phenomenon that explains everything and, therefore, nothing.

WGD has occurred across the breadth of eukaryote phylogeny [15–18], but the majority of WGD events have occurred within land plants (Embryophyta) (Figure 1). As such, plants provide very many natural experiments from which it may be possible to develop a general theory on the role of WGD in macroevolution. Patterns of diversification among extant taxa have pointed towards a scenario of rarely successful polyploids [19,20]. However, all members of the most diverse lineage

[1]School of Earth Sciences, University of Bristol, Life Sciences Building, Bristol BS8 1TH, UK

*Correspondence:
james.clark@bristol.ac.uk (J.W. Clark) and phil.donoghue@bristol.ac.uk (P.C.J. Donoghue).

Box 1. WGD and the K–Pg Boundary

The distribution of WGD events both across plant phylogeny and through time has revealed in multiple independent lineages that WGD events appear to cluster around the K–Pg boundary (Figure 1). This has led to two related hypotheses: that genome duplication may have conferred an 'extinction resistance' to particular lineages of plants, and that polyploid genomes may have allowed surviving lineages to rise to dominance in the wake of this mass extinction episode.

Polyploid plants are sometimes found towards the edge of species ranges, and polyploid genomes facilitate rapid radiations and invasiveness. Polyploid genomes also possess a 'mutational robustness' relative to diploids, and this may provide short-term advantages which could have allowed them to survive and then thrive. An alternative hypothesis suggests that it is not WGD itself that facilitated extinction resistance, but the coincidence that many newly formed polyploids rely on selfing for reproduction. Selfing is also associated with the extremes of novel habitats, but in the long term is seen as an evolutionary dead end. A return to outbreeding could allow the continued success of these lineages and may also explain the apparent lag between WGD and diversification.

These hypotheses are entirely dependent on the precise timing of each duplication event. As shown in Figure 2, current estimates for the timing of WGD are likely to change given a careful appraisal of the fossil record. As such, until the timing of each WGD event that is considered to lie close to the boundary is re-evaluated, this correlation should be treated with caution.

of land plants, the seed plants (Spermatophyta), are descended from an ancestor that underwent at least one round of WGD [21,22]. Furthermore, within Spermatophyta, another WGD is shared by all flowering plants (angiosperms) [21], and a further WGD is shared in turn by several major clades of flowering plants including the monocots [23], eudicots [24,25], Asteraceae [5,26], Brassicales [27], legumes [28], and the most economically important plants, the grasses [29,30] (Figure 2). The paucity of ancient WGD events that was perceived early in the history of genome sequencing increasingly appears to be an oversight, with denser sampling revealing multiple WGD events during the evolution of taxonomically large and small lineages [6].

Double Dates – The Absolute Timing of WGD

Hypotheses on the role of WGD in plant macroevolution are contingent on the phylogenetic (relative) and geological (absolute) timing of each event. Methods to identify WGD events are

Box 2. The Origins of WGD

Traditionally, polyploids are recognized as originating from a single parent species (autopolyploidy, xx to xxxx) or from two hybridizing species (allopolyploidy, xx + yy to xxyy). Current views maintain that these two outcomes exist along a spectrum, with segmental allopolyploids containing paralogs that display varying levels of synteny [77]. A segmental allopolyploid may form via hybridization between two closely related species, or through the process of homoeologous compensation [77]. Despite potential differences in outcome, both are likely to have had significant effects throughout plant evolution (both processes and their potential evolutionary outcomes have recently been reviewed [97–99]). Based on observations from neopolyploids, there is reason to believe that their outcomes may differ, and it is therefore a priority to establish whether ancient events were a consequence of autopolyploidy or allopolyploidy. Methods to differentiate between the two processes are under development, and in some instances ancient events have been successfully characterized. Genome dominance is a phenomenon observed in allopolyploids, where one subgenome shows lower expression and retention than the other (biased fractionation). Signal of a bias in gene retention between subgenomes could provide evidence for allo- rather than autopolyploidy [100]. Gene-tree methods are also capable of resolving allopolyploid WGDs by considering reticulate patterns of gene-tree evolution [17,101], and in some instances they have been able to identify the most likely parental lineages involved in the hybridization event [102].

The nature of the WGD affects the approach required for dating because auto- and allopolyploidy present different issues. The two subgenomes of an allopolyploid would have diverged at the point of speciation between the two parent lineages, rather than at the hybridization event itself [50,103]. Successful and viable hybrids are more likely to arise between closely related species, giving rise to 'segmental allopolyploids'. However, there are examples of hybridization between distantly related lineages of plants [104], which could lead to a significant overestimation of the age of the WGD. Similarly, as outlined previously, autopolyploidy can lead to a prolonged period of tetrasomic inheritance between ohnologs [59]. In this case there is potential to underestimate the age of the WGD because the ohnologs will only start to diverge once disomic inheritance has occurred, and we date the point at which they diverge rather than the date of duplication.

Glossary

Diploidization: sometimes termed fractionation, this is the period following WGD whereby through rearrangement, silencing, and loss of DNA the genome returns to a diploid expression pattern.
Morphospace: an *n*-dimensional multivariate space describing phenotypes, where points represent taxonomic units and the distances between them their (dis)similarities.
Neofunctionalization: following gene duplication, one copy of the gene takes on a novel function while the other copy continues to perform the previous function.
Paralogs, ohnologs, and homologs: two genes related by descent, typically with similar sequences, are homologs. If they share a 1:1 relationship between species, they are orthologs. If they deviate from this 1:1 relationship as a result of a duplication event, they become paralogs. Paralogs that have derived specifically from a WGD event are termed ohnologs, after Susumu Ohno.
Subfunctionalization: following gene duplication, each duplicate performs part of the original function, and in combination both maintain the original function of the gene.

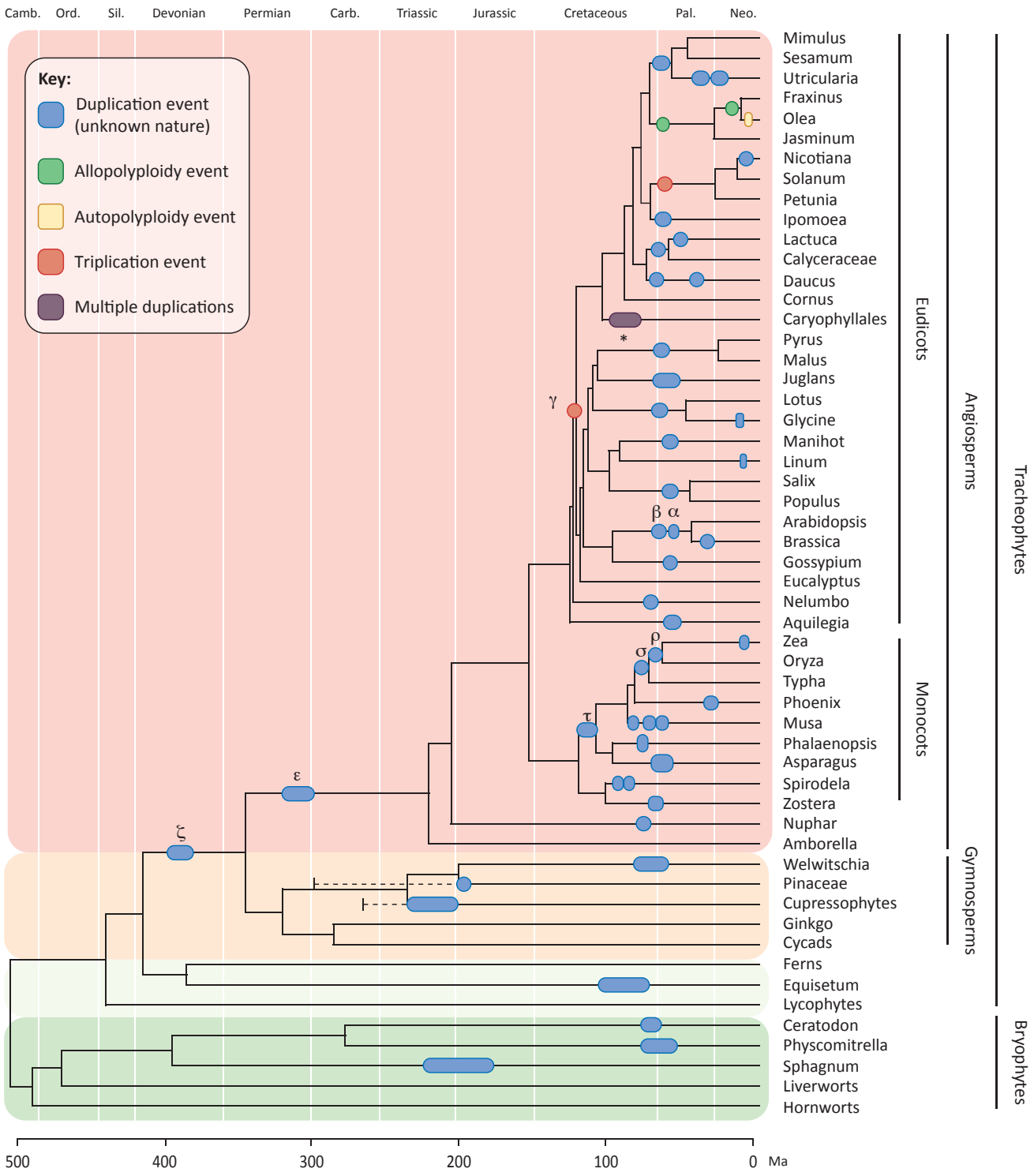

Figure 1. The Distribution of Known Whole-Genome Duplication (WGD) Events within the Plant Kingdom. Most events are shown from Van de Peer *et al.* [91] but have been updated. The length of each bar along the branch indicates the current estimate for its age. Duplication events of unknown origin are shown in navy

(Figure legend continued on the bottom of the next page.)

many and varied: these include **paralog** (see Glossary) substitution distributions (plots of the synonymous substitution rate, Ks) [31,32], phylogenomics [21], genome size, karyotype, gene copy-number analyses [33,34], and synteny [23,35,36]. Greater sampling of diversity helps resolve the phylogenetic (relative) timing of each WGD, but to refine these hypotheses it is important that their absolute ages are known with accuracy and precision. Absolute ages can be constrained by the age of bracketing speciation events since WGD must have occurred after the divergence of species that have not undergone WGD, and before those living species within the same lineage that have (Figure 2). When taxonomic sampling is dense and the WGD occurred on a short branch (such as with more recent events) this can yield relatively precise age estimates [37]. However, with increasing uncertainty in species divergence time estimates, longer branches, monotypic lineages, or less-dense sampling, it becomes more challenging to directly estimate the timing of a WGD.

As well as being a means to identify and relatively date WGD events, both Ks analyses and phylogenomic methods can be used to directly infer the age of WGD events [32,38–40]. Ks plots represent distributions of rates of synonymous substitutions between paralogs. A peak in the distribution is interpreted as a WGD event, and distributions compared between species can reveal shared duplication events. An external calibration can convert Ks rates into geological time, although this is often done by comparing the position of the peak in Ks rates to ages inferred from phylogenomic dating, for example a Ks value between 0.6 and 1.1 synonymous substitutions per site corresponds to an age of 50–70 Myr. These methods assume a strict rate of molecular evolution, and diff erent rates produce highly variable age estimates. The signature of increasingly ancient WGD events is eroded by sequence saturation, and therefore the dating of more ancient events is prone to error [32]. For example, a WGD event predicted in the early-diverging gymnosperm *Ginkgo biloba* was estimated at between 500 and 700 Ma – pre-dating most estimates for the origin of land plants [41–43].

Phylogenomic approaches exploit the signal of paralogy present in the history of gene families to directly estimate the age of the WGD event [21]. This requires the reconstruction of gene families across multiple species (also termed a phylome [44]) and subjecting them to molecular clock analysis. Molecular clock methodology has typically been applied to dating species divergences but can also be used to date both speciation and duplication events within gene trees. Typically, molecular clock analyses have investigated each gene family in isolation, producing both a topological and temporal estimate of WGD events. Molecular clock approaches to dating WGD have either been flawed by the underlying algorithm [45] or, when more powerful Bayesian uncorrelated methods have been used, by the limited sampling of taxa and a paucity of appropriate fossil calibrations [40]. Furthermore, dating individual gene families does not make best use of information available because individual gene families have low statistical power, yielding imprecise, if not inaccurate, estimates of gene and (by inference) genome duplication dates.

Paralog sets derived from a WGD share the same age and can be combined in a concatenated alignment that is capable of producing far more precise results than any single gene family [46,47]. Precision is not the sole concern, and improved accuracy is achieved by using conservative paleontological constraints on speciation events [48] alongside clock methods that can model both the uncertainty in the fossil evidence and the variation in rates of evolution between genes [46,49]. Box 3 shows a schematic analysis of the WGD present in the ancestor

blue, triplications in red, known autopolyploidy events in yellow, and allopolyploidy events in green. The white bar associated with Caryophyllales represents 26 independent WGD events, some of which are autopolyploidy and some allopolyploidy. Named duplication events are shown alongside their Greek letter. Abbreviations: Camb., Cambrian; Carb., Carboniferous; Ord., Ordovician; Neo., Neogene; Pal., Paleozoic; Sil., Silurian.

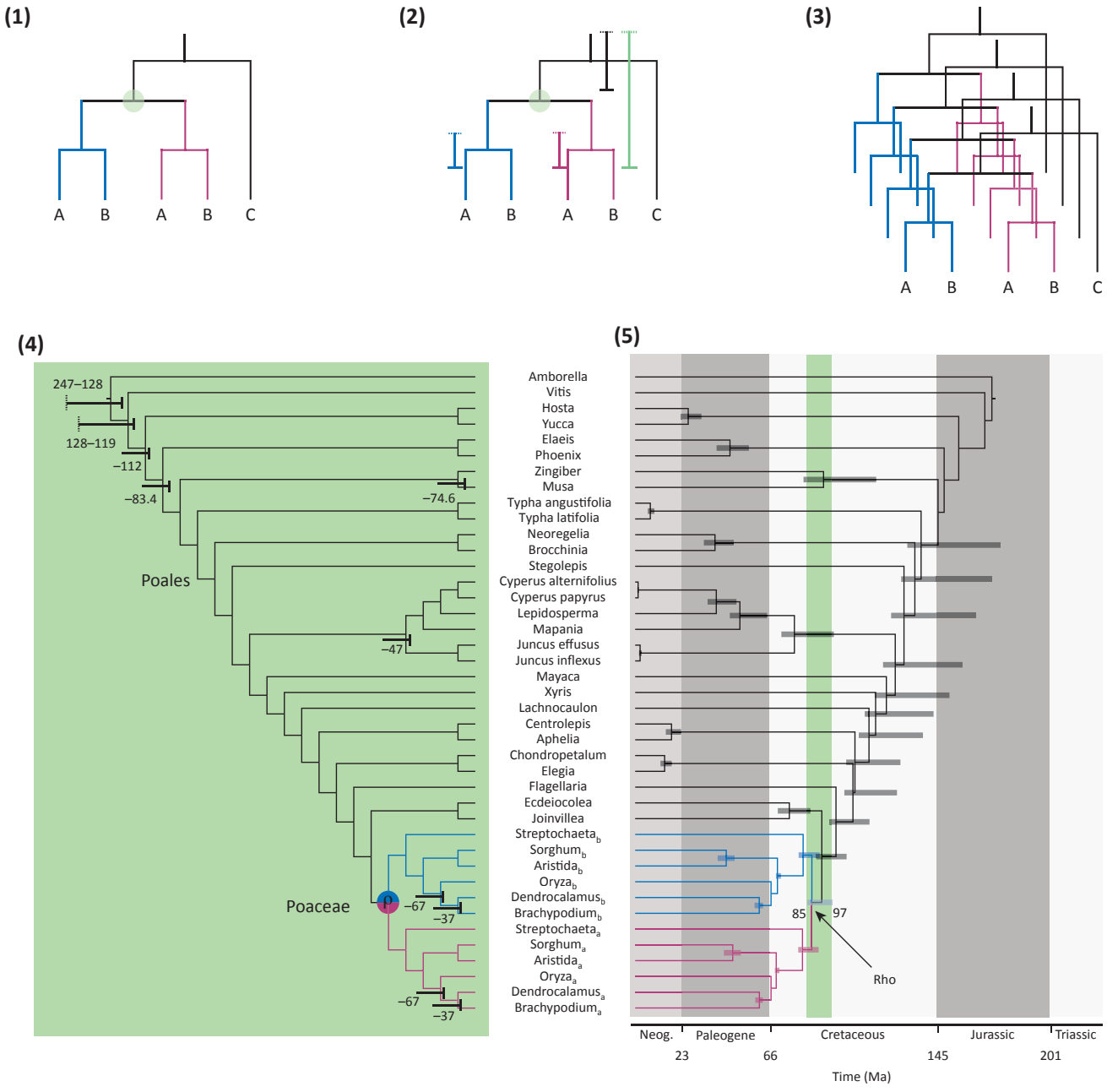

Figure 2. Dating Whole-Genome Duplication (WGD) by Combining Genomics and the Fossil Record. (1) the history of WGD is present in individual gene families. Taxa A and B have undergone a shared duplication event, which taxon C has not. (2) The timing of the duplication is bracketed by the timing of the divergence of A and B and the divergence of A + B and C. These divergence times can be calibrated using distributions between minimum and soft maximum ages. (3) Multiple gene families with a shared WGD signal can be concatenated to maximize the precision of the analysis. (4) Accuracy is achieved through careful appraisal of the fossil record and by modeling uncertainty through soft maximum ages [46,94]. (5) A Bayesian molecular clock analysis reveals that the grass duplication (Rho) occurred at 85–97 Ma (95% highest posterior density, HPD).

Box 3. Dating WGD in Grasses

Syntenic and phylogenomic evidence points towards a WGD event in the ancestor of all extant grasses (Poaceae) [23,30]. The Rho event has previously been dated through phylogenetic bracketing to ~70 Ma [90] and is one of the numerous plant WGDs hypothesized to approximate the K–Pg boundary [91]. We sampled the gene families that were previously shown to retain the signal of the Rho duplication (Figure 2.1) and concatenated them into an alignment (Figure 2.3; see File S2 in the supplemental information online). Fossil evidence constrains the minimum age on speciation nodes (see File S1 in the supplemental information online), and in some cases can be used to apply 'soft' maxima [92] (Figure 2.2). The Late Cretaceous fossil phytolith taxon *Changii indicum* is assigned to the crown group (i.e., the living clade) of the Oryzeae tribe, and provides a minimum age of 66 Ma based on radiometric dating [93–95]. This fossil placement of this fossil is contentious and can be instead used to calibrate the BOP (Bambusoideae, Oryzoideae, Pooideae) + PACMAD (Panicoideae, Arundinoideae, Chloridoideae, Micrairoideae, Aristidoideae, Danthonioideae) clade of grasses [95]. We applied further fossil constraints and, combined with the concatenated alignment, these calibrations inform a Bayesian molecular clock analysis performed on the fixed topology of McKain *et al*. in the phylogenetic program MCMCTREE [96]. The results predict that the WGD took place in the 97–85 Ma period, and in this case is not compatible with the hypothesis that this event coincides with the K–Pg boundary (Figure 2.5). Abbreviation: ANA grade.

of all grasses (Rho). This event is evident in the genomes and phylomes of multiple extant grass species which, owing to their economic value as food crops, have been well-sampled by sequencing projects [30].

As well as being able to inform on the coincidence of WGD with geological or biogeographic events, these approaches coestimate the timing of duplication alongside the timing of speciation. This allows us to see how early or late WGD occurred relative to the crown (extant) clade and to directly estimate lag between the WGD event and any hypothesized macroevolutionary consequences [46].

Whole Genomes and Diversification

Diversification is one of the most widely proposed consequences of WGD in plants. This relationship has been explored at multiple levels across angiosperms, but support for a correlation remains equivocal [2,29,50,51]. There is little evidence supporting a direct shift in diversification immediately following WGD. Instead, there is some support for the proposed 'WGD lag-time' model, wherein diversification follows WGD – but only after a protracted period of geological time [2]. The lag period has been measured either as a period of absolute time or as an arbitrary measure of time such as the number of nodes separating a WGD event and a subsequent shift in the rate of diversification. When the age of the duplication event and the subsequent speciation events are coestimated, the absolute age and duration of the lag can be estimated directly [46]. Estimates for the timing of the angiosperm-specific genome duplication event imply that it occurred 65–35 Myr before the divergence of crown angiosperms (the living clade of flowering plants), closer to 70 Myr before the radiation of the Mesangiospermae and over 100 Myr before a detectable angiosperm radiation in the fossil record [46,52]. Such an extensive lag raises two questions: first, is it plausible to associate two events that are separated by such a long interval of time? Second, why did the early diverging lineages of angiosperms (the ANA grade: Amborellales, Nymphaeales, and Austrobaileyales) not undergo a similar radiation?

Schranz *et al*. [53] proposed a model in which WGD provides latent evolvability that may be later triggered by a shift in environment and promote diversification. This has been further refined, and several new models have emerged to explain the lag phase, some of which are readily testable. Among these is the suggestion that it is not WGD, but the ensuing process of genome fractionation (or **diploidization**), that may be responsible for diversification. During this process the organism undergoes large-scale genome rearrangements and redundant gene copies are silenced and excised, leading to potentially novel patterns of expression [54]. Most angiosperm lineages have undergone multiple rounds of WGD and exhibit the fastest rate of genome size

evolution among land plants [55], and it has been proposed that their ability to rapidly downsize their genome in the wake of WGD has led to their global dominance [56]. Ferns show a higher rate of genome duplication than angiosperms but appear not to undergo such extensive genome downsizing and are considerably less diverse than angiosperms [33,57]. The observed lag between WGD and diversification in angiosperms may be explained by the period of genome fractionation, although the long-term rate of fractionation is uncertain. It seems appropriate to ask whether the extent or rate of genome reorganization post-WGD correlates with observed shifts in the rate of diversification. The WGD event associated with one of the most dramatic shifts in diversification, the gamma event at the base of the eudicots, involved extensive genome reorganization [25,58]. Speciation post-WGD would lead to fractionation occurring independently in separate lineages, which could explain the differences between lineages that emerge from WGD [54].

In the specific case of autopolyploidy (duplication involving a single parental lineage) the newly duplicated paralogs can pair randomly at meiosis. This pattern of tetrasomic inheritance facilitates ongoing exchange between paralogous chromosomes and may prevent them from diverging until a state of disomic inheritance is restored [59,60]. The period required to attain a state of disomic inheritance could also explain the macroevolutionary lag between WGD and phenotypic evolution. As with the duplication/fractionation model, speciation occurring before the restoration of disomic inheritance will result in independent diploidization of lineages. Robertson *et al*. [59] demonstrated this 'lineage-specific **ohnolog** resolution' (LORe) model in the descendants of the salmonid fish-specific WGD event, and showed that independent diploidization was present in 27% of salmonid paralogs. Although untested in plants, its predictions of a long lag period and disparate evolutionary trajectories suggest that it may also fit the patterns observed after the angiosperm-specific WGD.

The case for a general theory of WGD as an intrinsic driver of diversification is undermined by the multiple cases where WGD does not accompany any shift in diversity. Non-seed-plant lineages, such as paleopolyploid mosses and horsetails, remain species-poor despite repeated duplications [7,10]. This can be partly reconciled by the differing rates of genome downsizing and rearrangement exhibited by these clades relative to angiosperms. However, further research on the mechanisms for rapidly altering genome structure is required. Beyond plants and, in particular, among teleost fish, the paleontological record shows no evidence in support of a role for WGD in directly promoting diversification [61]. There is some evidence supporting a direct role for WGD in promoting diversity in yeasts where reciprocal gene loss can lead to reproductive isolation [62], although on a macroevolutionary scale this effect is small [63].

WGD and Morphological Innovation

The link between WGD and morphological evolution in plants has remained both pervasive and speculative [1,64]. Some have proposed that polyploids may survive and evolve in extreme or marginal habitats, allowing them a competitive advantage over their parent species at range margins [65]. However, the range of many extant polyploids does not exceed that of their parents [66], while genes related to stress tolerance appear to have evolved via tandem duplication rather than by WGD [67,68]. The evolution of morphological diversity, like species diversity, may also require a lag phase. For selection to act on innovation, developmental robustness is required [69], and hence it is possible that morphological diversification may occur only after a period of developmental lability. At the genetic level, WGD may free a lineage from the constraints of purifying selection and allow genes to take on new functions [1], such as through **neofunctionalization** and **subfunctionalization**. At the phenotypic level this may allow the evolution of novel forms and body plans. Indeed, formative innovations within the plant kingdom have been

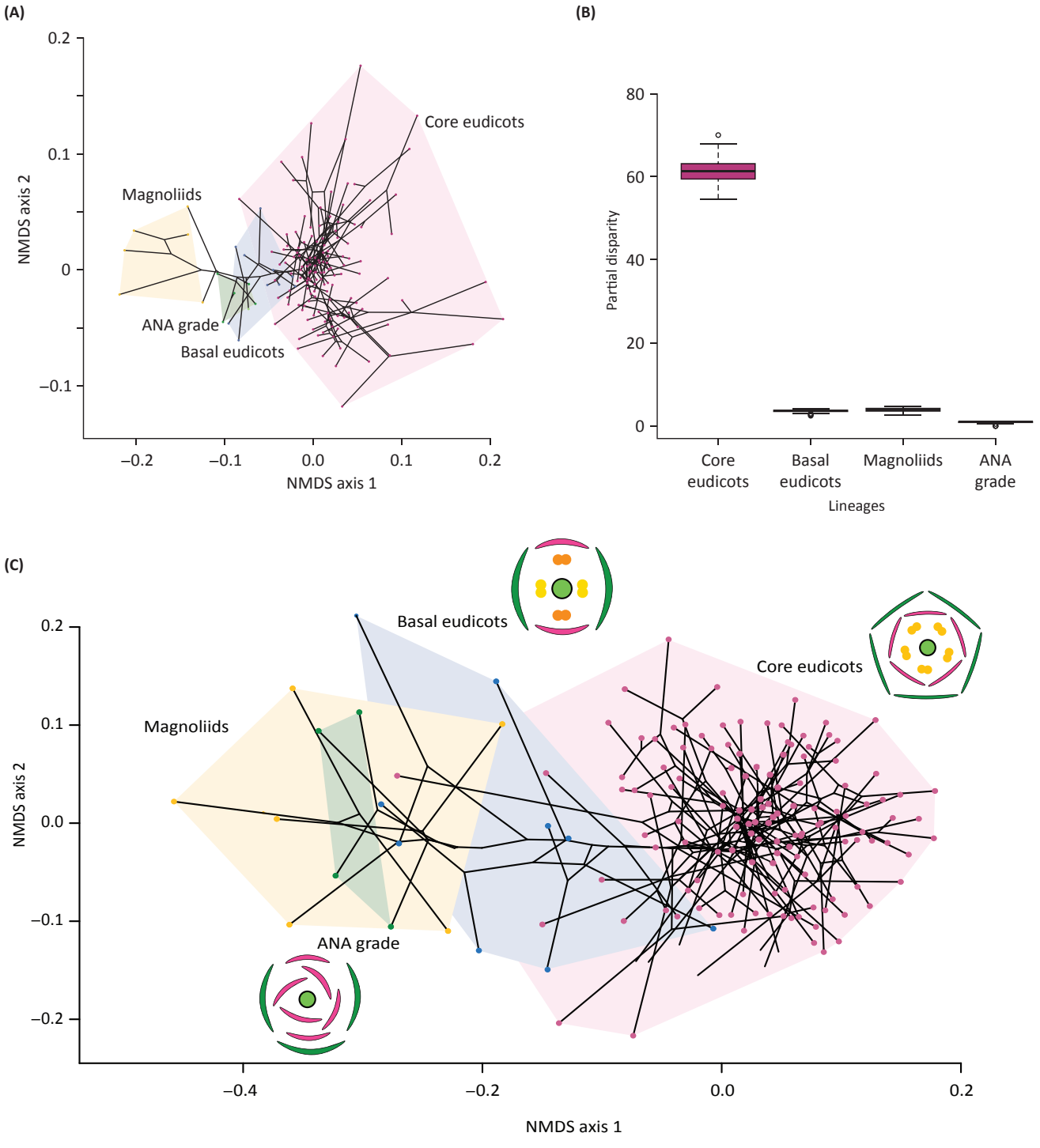

Figure 3. Phenotypic Evolution in the Wake of the Gamma Triplication Which Occurred before the Evolution of the Core Eudicots. (A) An empirical morphospace based on a morphological matrix [85]. Morphological characters form the basis of a distance metric (Gower's index) which was subjected to non-metric multidimensional scaling (NMDS) to display variation in two axes. A consensus phylogeny is mapped onto the morphospace (see File S4 in the supplementary

(Figure legend continued on the bottom of the next page.)

associated with the expansion of families of regulatory genes [70,71]. Patterns of gene retention post-WGD are not random, and in repeated cases genes encoding proteins that function as part of networks and signaling cascades are retained preferentially [72–74]. This has been explained in terms of dosage balance and the need to maintain stoichiometric ratios of proteins within the cell [75,76]. The dosage-balance hypothesis is exemplified during the diploidization process in allopolyploids, where exchanges can occur between homoeologous chromosomes of subgenomes [77]. These exchanges can result in novel gene expression and gene copy number [78], but can also result in the deleterious loss of chromosome regions or entire chromosomes. Homoeologous compensation has been proposed as a mechanism to prevent dosage imbalances, and has been demonstrated to lead to increased phenotypic variation in newly synthesized allopolyploids [77]. The dosage-balance hypothesis does not predict the evolution of morphological diversity until such constraints are relaxed and retained paralogs are selected to evolve new functions [14,79]. These constraints may relax under different selection pressures, although a quantitative model of compensatory drift has also been proposed [80]. Compensatory drift is the process whereby paralogs are initially retained due to dosage sensitivity, but over time the expression levels of the individual genes drift until one paralog is free of the dosage-dependent constraint [80]. This model not only provides a mechanism for neofunctionalization to arise from a state of dosage balance but also a potential explanation for the emergence of evolutionary novelty after prolonged periods of evolutionary time.

It is difficult to ascribe adaptive evolution to WGD, especially with ancient events. The link between WGD and novelty has been elegantly shown in the glucosinolate pathway in Brassicales [4]. This gene family has expanded over several rounds of WGD and is involved in plant–herbivore interactions. It has also been proposed that gene families underpinning floral patterning expanded during the angiosperm-specific WGD [71]. These genes are implicated in the origin and diversification of the flower, a structure that has shaped recent plant and animal evolution [81]. The evolution of pentamerous flowers in the core eudicots also coincides with a genome triplication (gamma, Figure 1) [25,82]. The coincidence of the gamma event with this major synapomorphy, a large increase in the rate of diversification, and extensive genome reorganization [58] makes it a tantalizing system in which to investigate the link between WGD and morphological evolution.

Regulatory gene retention and large shifts in their transcription patterns suggest a role for WGD in the evolution of eudicot floral diversity [82]. To make such a hypothesis testable, the increase in phenotypic complexity must be quantified for comparative analysis [83]. To achieve this we can borrow from paleontology, which has a strong tradition in comparative analysis of phenotype through multivariate statistics – manifest as **morphospace** analyses. The hypothesis that WGD drives innovation would predict that events coincide either with movement to a new 'island' within phenotype 'design space' or with continued expansion of an existing island. These predictions can be tested explicitly with datasets that use discrete phenotypic characters to describe the traits that unify and distinguish taxa [84]. For example, one can characterize the disparity of extant angiosperms to test the hypothesis that the gamma triplication event coincides with an increase in phenotypic diversity. To do this we used a morphological dataset that captures the disparity of early angiosperms, basal eudicots, and core eudicots [85] (see File S3 in the supplemental information online). We used these data to calculate the dissimilarity between each taxa, as measured using Gower's dissimilarity metric [86]. To visualize this dissimilarity we performed non-

information online). (B) The contribution to total disparity (partial disparity) of each clade calculated from distance matrix (1000 bootstrap replicates) [105]. (C) A morphospace constructed from floral characters. Major trends in floral evolution are displayed next to the lineages; spiral phyllotaxis is present in early angiosperms, dimerous flowers are common among basal eudicots, and the pentamerous flower is associated with the core eudicots. Abbreviation: ANA grade, from Amborellales, Nymphaeales, and Austrobaileyales.

Box 4. Duplication and Disparity in the Conifers

Some explosive WGD events, such as that associated with the core eudicots, coincide with rapid diversification and an increase in phenotypic variation. However, many WGD events in species-poor lineages are not closely associated with macroevolutionary phenomena. Most conifers are thought to have undergone at least two rounds of WGD during their evolution, one shared with seed plants, and then two lineage-specific events on the branches leading to Pinaceae and Cupressophytes [22]. Preliminary analyses of diversity and disparity in the pines indicate a rapid increase in phenotypic variance during the Late Jurassic and Early Cretaceous [83], and the Pinaceae occupy a highly distinct area of morphospace (Figure I). This provides some corroborative support for the hypothesis that WGD resulted in phenotypic variation among conifers during their early evolution. However, the age of the pine WGD is currently estimated at between 342 and 200 Ma [22] (Figure I); given such uncertainty it is not presently possible to link WGD to the shift in phenotypic disparity. This example highlights the need to employ methods that can accurately and precisely estimate the timing of WGD events because a temporal framework is essential for testing macroevolutionary hypotheses [46].

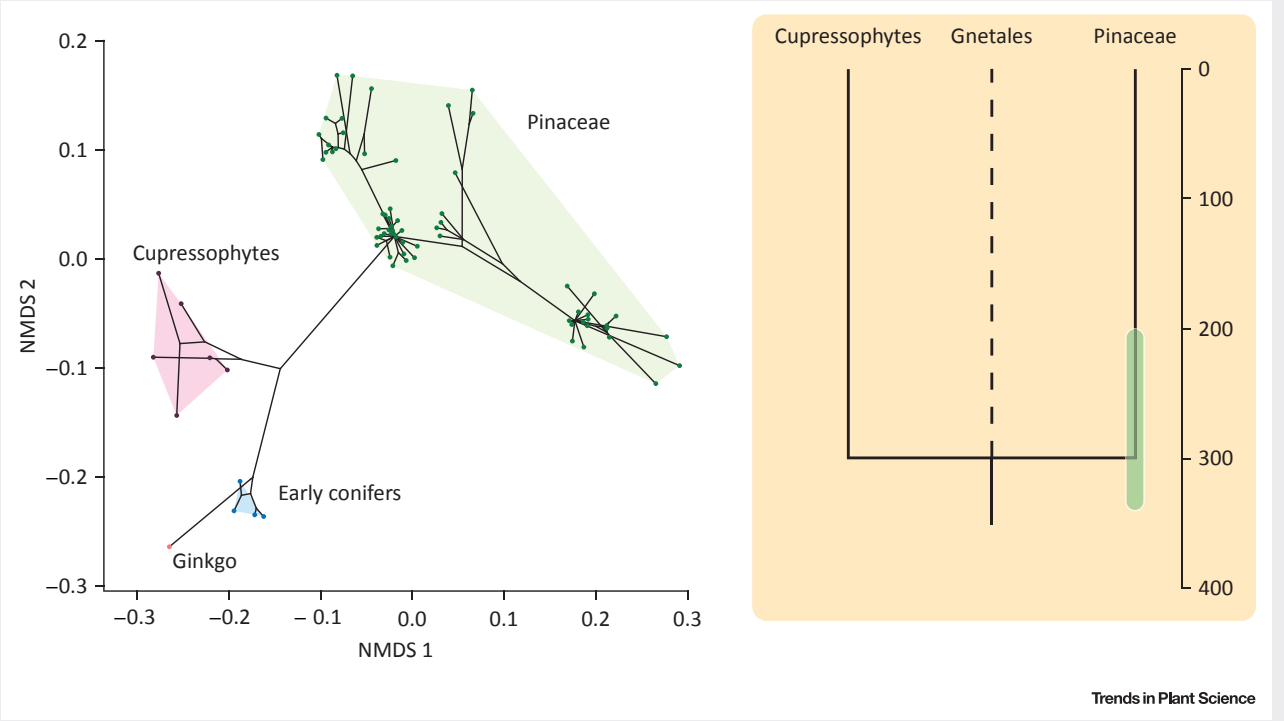

Figure I. Evolution in Pinaceae. An empirical morphospace of Pinaceae and relatives built from phenotypic characters [106] (see File S5 in the supplemental information online) which formed the basis of a distance matrix (Gower's index) that was subjected to non-metric multidimensional scaling (NMDS). A consensus phylogeny is mapped onto the morphospace (see File S6 in the supplemental information online). The uncertainty of both the relative (phylogenetic) and absolute timing of the event limits our understanding of the consequences because the position of the Gnetales remains contentious and the current estimate for the age of the WGD spans over 100 Myr.

metric multidimensional scaling, a non-metric ordination method that summarizes variation over a specified number of axes – in this instance, two. The result is presented in Figure 3, which shows that the core eudicots occupy a far greater area of morphospace than the basal eudicots. Furthermore, relative to other early diverging lineages of angiosperms, they occupy the largest proportion of morphospace (partial disparity, Figure 3B). In addition, we subsampled the character matrix for floral characters only, relating specifically to the gamma-derived hypothesis (Figure 3C). The resulting morphospace shows less separation between the lineages, but core eudicots still occupy the largest area and, therefore, exhibit the greatest variation. The construction of a morphospace can be subjective in that it is dependent on the choice of taxa and characters – but there is strong evidence to suggest that the gamma triplication coincides with the rapid evolution of morphological disparity among eudicots. A comparable analysis of the impact of WGD in pines finds support for increased variance in morphospace occupation, but gross uncertainty in the estimate of the timing of WGD relative to the age of the disparate clade undermines the hypothesis of a causal link (Box 4).

Quantifying phenotypic evolution across multiple lineages will be instrumental in understanding the role of WGD in the evolution of morphologic complexity. The inclusion of fossil taxa and recent methods used to estimate disparity through time may allow us to measure the tempo of phenotypic evolution post-WGD. The impact of key innovations that are attributed to WGD can be tested by considering their impact on the shape of a morphospace or whether the innovation has resulted in diversification. A further question arises as to what degree WGD is essential for the appearance of major innovations. For example, the origin of seed and flowering plants coincides with a WGD event, but, arguably, a greater number of characters unite the vascular plants whose origin was independent of any known WGD event [87]. While it is plausible that saltational evolution has been caused by WGD in the plant kingdom [88], phenotypic complexity may also arise through the evolution more nuanced *trans*- and *cis*-acting regulation [89].

Concluding Remarks

WGD is associated with a macroevolutionary outcome in some but not all lineages, and it remains unclear how and why this is the case. As the number of identified WGD events in plant evolutionary history increases, there is an ever-growing need for a general theory on the role of WGD in macroevolution. However, to establish whether WGD is a class of event with characteristic and predictable outcomes, further work will be necessary to place, both relatively and absolutely, each event in time. There are many outstanding questions to be answered, but a precise temporal framework forms the basis for tests that can quantify any macroevolutionary consequences and inform and refine hypotheses about the relationship between WGD, diversification, and morphological evolution. Plants are the best system in which to elucidate the effects of WGD because of the prevalence of these genomic events in plant phylogeny. This will be crucial as we seek to explain the consequences beyond any single event and, given the role that genome duplication has had in the evolution of many crop species, being able to make general predictions about the outcome of WGD is of crucial interest (see Outstanding Questions).

Outstanding Questions

Questions remain about the absolute timing of many of the identified WGD events among plants – of particular interest in the clustering of events around the K–Pg boundary.

The origin of duplication events is important – it has implications for both the timing and evolutionary consequences.

Is morphological evolution accelerated in the wake of WGD, and what impact has WGD had on the plant morphospace?

Disparate outcomes between lineages, in terms of morphology and diversity, still require investigation.

Acknowledgments

We thank Simon Hiscock, Kyle Martin, Mark Puttick, Harald Schneider, and Ziheng Yang for discussion. J.E.C. is funded by a PhD studentship from the Biotechnology and Biological Sciences Research Council (BBSRC; swDTP); P.C.J.D. is funded by grants from the Natural Environment Research Council (NERC; NE/N002067/1) and the BBSRC (BB/N000919/1).

Appendix A Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.tplants.2018.07.006.

References

1. Ohno, S. (1970) *Evolution by Gene Duplication*, Springer Science Business Media
2. Tank, D.C. *et al.* (2015) Nested radiations and the pulse of angiosperm diversification: increased diversification rates often follow whole genome duplications. *New Phytol.* 207, 454–467
3. Ren, R. *et al.* (2018) Widespread whole genome duplications contribute to genome complexity and species diversity in angiosperms. *Mol. Plant* 11, 414–428
4. Edger, P.P. *et al.* (2015) The butterfly plant arms-race escalated by gene and genome duplications. *Proc. Natl. Acad. Sci. U.S.A.* 112, 8362–8366
5. Barker, M.S. *et al.* (2016) Most Compositae (Asteraceae) are descendants of a paleohexaploid and all share a paleotetraploid ancestor with the Calyceraceae. *Am. J. Bot.* 103, 1203–1211
6. Smith, S.A. *et al.* (2018) Disparity, diversity, and duplications in the Caryophyllales. *New Phytol.* 217, 836–854
7. Vanneste, K. *et al.* (2015) Horsetails are ancient polyploids: evidence from *Equisetum giganteum*. *Plant Cell* 27, 1567–1578
8. Lohaus, R. and Van de Peer, Y. (2016) Of dups and dinos: evolution at the K/Pg boundary. *Curr. Opin. Plant Biol.* 30, 62–69
9. Fawcett, J.A. *et al.* (2009) Plants with double genomes might have had a better chance to survive the Cretaceous–Tertiary extinction event. *Proc. Natl. Acad. Sci.* 106, 5737–5742
10. Devos, N. *et al.* (2016) Analyses of transcriptome sequences reveal multiple ancient large-scale duplication events in the ancestor of Sphagnopsida (Bryophyta). *New Phytol.* 211, 300–318
11. Cui, L. *et al.* (2006) Widespread genome duplications throughout the history of flowering plants. *Genome Res.* 16, 738–749
12. Walker, J.F. *et al.* (2017) Widespread paleopolyploidy, gene tree conflict, and recalcitrant relationships among the carnivorous Caryophyllales. *Am. J. Bot.* 104, 858–867

13. Xiang, Y. *et al.* (2017) Evolution of Rosaceae fruit types based on nuclear phylogeny in the context of geological times and genome duplication. *Mol. Biol. Evol.* 34, 262–281
14. Conant, G.C. *et al.* (2014) Dosage, duplication, and diploidization: clarifying the interplay of multiple models for duplicate gene evolution over time. *Curr. Opin. Plant Biol.* 19, 91–98
15. Donoghue, P.C. and Purnell, M.A. (2005) Genome duplication, extinction and vertebrate evolution. *Trends Ecol. Evol.* 20, 312–319
16. Schwager, E.E. *et al.* (2017) The house spider genome reveals an ancient whole-genome duplication during arachnid evolution. *BMC Biol.* 15, 62
17. Marcet-Houben, M. and Gabaldón, T. (2015) Beyond the whole-genome duplication: phylogenetic evidence for an ancient interspecies hybridization in the baker's yeast lineage. *PLoS Biol.* 13, e1002220
18. Li, Z. *et al.* (2018) Multiple large-scale gene and genome duplications during the evolution of hexapods. *Proc. Natl. Acad. Sci.* 115, 4713–7718
19. Arrigo, N. and Barker, M.S. (2012) Rarely successful polyploids and their legacy in plant genomes. *Curr. Opin. Plant Biol.* 15, 140–146
20. Soltis, D.E. *et al.* (2011) Are polyploids really evolutionary dead-ends (again)? A critical reappraisal of Mayrose *et al*. (2011). *New Phytol.* 202, 1105–1117
21. Jiao, Y. *et al.* (2011) Ancestral polyploidy in seed plants and angiosperms. *Nature* 473, 91–100
22. Li, Z. *et al.* (2015) Early genome duplications in conifers and other seed plants. *Sci. Adv.* 1, e1501084
23. Jiao, Y. *et al.* (2014) Integrated syntenic and phylogenomic analyses reveal an ancient genome duplication in monocots. *Plant Cell* 26, 2792–2802
24. Jaillon, O. *et al.* (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449, 463–467
25. Jiao, Y. *et al.* (2012) A genome triplication associated with early diversification of the core eudicots. *Genome Biol.* 13, R3
26. Huang, C.-H. *et al.* (2016) Multiple polyploidization events across Asteraceae with two nested events in the early history revealed by nuclear phylogenomics. *Mol. Biol. Evol.* 33, 2820–2835
27. Kagale, S. *et al.* (2014) Polyploid evolution of the Brassicaceae during the Cenozoic era. *Plant Cell* 26, 2777–2791
28. Cannon, S.B. *et al.* (2015) Multiple polyploidy events in the early radiation of nodulating and nonnodulating legumes. *Mol. Biol. Evol.* 32, 193–210
29. Estep, M.C. *et al.* (2014) Allopolyploidy, diversification, and the Miocene grassland expansion. *Proc. Natl. Acad. Sci.* 111, 15149–15154
30. McKain, M.R. *et al.* (2016) A phylogenomic assessment of ancient polyploidy and genome evolution across the poales. *Genome Biol. Evol.* 8, 1150–1164
31. Lynch, M. and Conery, J.S. (2000) The evolutionary fate and consequences of duplicate genes. *Science* 290, 1151–1155
32. Vanneste, K. *et al.* (2013) Inference of genome duplications from age distributions revisited. *Mol. Biol. Evol.* 30, 177–190
33. Clark, J. *et al.* (2016) Genome evolution of ferns: evidence for relative stasis of genome size across the fern phylogeny. *New Phytol.* 210, 1072–1082
34. Tiley, G.P. *et al.* (2016) Evaluating and characterizing ancient whole-genome duplications in plants with gene count data. *Genome Biol. Evol.* 8, 1023–1037
35. Tang, H. *et al.* (2008) Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome Res.* 18, 1944–1954
36. Lyons, E. *et al.* (2008) Finding and comparing syntenic regions among *Arabidopsis* and the outgroups papaya, poplar, and grape: CoGe with Rosids. *Plant Physiol.* 148, 1772–1781
37. Edger Patrick, P. *et al.* (2018) Brassicales phylogeny inferred from 72 plastid genes: a reanalysis of the phylogenetic localization of two paleopolyploid events and origin of novel chemical defenses. *Am. J. Bot.* 105, 463–469
38. Lynch, M. and Conery, J.S. (2000) The evolutionary fate and consequences of duplicate genes. *Science* 290, 1151–1155
39. Blanc, G. and Wolfe, K.H. (2004) Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell* 16, 1667–1678
40. Vanneste, K. *et al.* (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
41. Guan, R. *et al.* (2016) Draft genome of the living fossil *Ginkgo biloba*. *GigaScience* 5, 49
42. Roodt, D. *et al.* (2017) Evidence for an ancient whole genome duplication in the cycad lineage. *PLoS One* 12, e0184454
43. Morris, J.L. *et al.* (2018) The timescale of early land plant evolution. *Proc. Natl. Acad. Sci.* 115, E2274–E2283
44. Huerta-Cepas, J. *et al.* (2014) PhylomeDB v4: zooming into the plurality of evolutionary histories of a genome. *Nucleic Acids Res.* 42, D897–D902
45. Ruprecht, C. *et al.* (2018) Revisiting ancestral polyploidy in plants. *Sci. Adv.* 3, e1603195
46. Clark, J.W. and Donoghue, P.C.J. (2017) Constraining the timing of whole genome duplication in plant evolutionary history. *Proc. Biol. Sci.* 284, 20170912
47. Macqueen, D.J. and Johnston, I.A. (2014) A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. *Proc. Biol. Sci.* 281, 20132881
48. Parham, J.F. *et al.* (2012) Best practices for justifying fossil calibrations. *Syst. Biol.* 61, 346–359
49. Warnock, R.C. *et al.* (2015) Calibration uncertainty in molecular dating analyses: there is no substitute for the prior evaluation of time priors. *Proc. Biol. Sci.* 282, 20141013
50. Kellogg, E.A. (2016) Has the connection between polyploidy and diversification actually been tested? *Curr. Opin. Plant Biol.* 30, 25–32
51. Soltis, D.E. *et al.* (2009) Polyploidy and angiosperm diversification. *Am. J. Bot.* 96, 336–348
52. Silvestro, D. *et al.* (2015) Revisiting the origin and diversification of vascular plants through a comprehensive Bayesian analysis of the fossil record. *New Phytol.* 207, 425–436
53. Schranz, M.E. *et al.* (2012) Ancient whole genome duplications, novelty and diversification: the WGD radiation lag-time model. *Curr. Opin. Plant Biol.* 15, 147–153
54. Dodsworth, S. *et al.* (2016) Is post-polyploidization diploidization the key to the evolutionary success of angiosperms? *Bot. J. Linn. Soc.* 180, 1–5
55. Puttick, M.N. *et al.* (2015) Size is not everything: rates of genome size evolution, not C-value, correlate with speciation in angiosperms. *Proc. Biol. Sci.* 282, 20152289
56. Simonin, K.A. and Roddy, A.B. (2018) Genome downsizing, physiological novelty, and the global dominance of flowering plants. *PLoS Biol.* 16, e2003706
57. Wood, T.E. *et al.* (2009) The frequency of polyploid speciation in vascular plants. *Proc. Natl. Acad. Sci.* 106, 13875–13879
58. Wang, Y. *et al.* (2016) Large-scale gene relocations following an ancient genome triplication associated with the diversification of core eudicots. *PLoS One* 11, e0155637
59. Robertson, F.M. *et al.* (2017) Lineage-specific rediploidization is a mechanism to explain time-lags between genome duplication and evolutionary diversification. *Genome Biol.* 18, 111
60. Martin, K.J. and Holland, P.W.H. (2014) Enigmatic orthology relationships between Hox clusters of the African butterfly fish and other teleosts following ancient whole-genome duplication. *Mol. Biol. Evol.* 31, 2592–2611

61. Laurent, S. *et al.* (2017) No evidence for the radiation time lag model after whole genome duplications in Teleostei. *PLoS One* 12, e0176384
62. Maclean, C.J. and Greig, D. (2011) Reciprocal gene loss following experimental whole-genome duplication causes reproductive isolation in yeast. *Evolution* 65, 932–945
63. Muir, C.D. and Hahn, M.W. (2015) The limited contribution of reciprocal gene loss to increased speciation rates following whole-genome duplication. *Am. Nat.* 185, 70–86
64. Crow, K.D. and Wagner, G.P. (2006) What is the role of genome duplication in the evolution of complexity and diversity? *Mol. Biol. Evol.* 23, 887–892
65. Stebbins, G.L. (1947) Types of polyploids: their classification and significance. *Adv. Genet.* 1, 403–429
66. Glennon, K. *et al.* (2014) Evidence for shared broad-scale climatic niches of diploid and polyploid plants. *Ecol. Lett.* 17, 574–582
67. Panchy, N. *et al.* (2016) Evolution of gene duplication in plants. *Plant Physiol.* 171, 2294–2316
68. Hanada, K. *et al.* (2008) Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli. *Plant Physiol.* 148, 993–1003
69. Melzer, R. and Theißen, G. (2016) The significance of developmental robustness for species diversity. *Ann. Bot.* 117, 725–732
70. Rensing, S.A. (2014) Gene duplication as a driver of plant morphogenetic evolution. *Curr. Opin. Plant Biol.* 17, 43–48
71. Chanderbali, A.S. *et al.* (2016) Evolving ideas on the origin and evolution of flowers: new perspectives in the genomic era. *Genetics* 202, 1255–1265
72. Seoighe, C. and Gehring, C. (2004) Genome duplication led to highly selective expansion of the *Arabidopsis thaliana* proteome. *Trends Genet.* 20, 461–464
73. Qiao, X. *et al.* (2018) Different modes of gene duplication show divergent evolutionary patterns and contribute differently to the expansion of gene families involved in important fruit traits in pear (*Pyrus bretschneideri*). *Front. Plant Sci.* 9, 161
74. Veron, A.S. *et al.* (2007) Evidence of interaction network evolution by whole-genome duplications: a case study in MADS-box proteins. *Mol. Biol. Evol.* 24, 670–678
75. Veitia, R.A. *et al.* (2008) Cellular reactions to gene dosage imbalance: genomic, transcriptomic and proteomic effects. *Trends Genet.* 24, 390–397
76. Birchler, J.A. and Veitia, R.A. (2012) Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. *Proc. Natl. Acad. Sci. U. S. A.* 109, 14746–14753
77. Xiong, Z. *et al.* (2011) Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid *Brassica napus*. *Proc. Natl. Acad. Sci. U. S. A.* 108, 7908–7913
78. Lloyd, A. *et al.* (2017) Homoeologous exchanges cause extensive dosage-dependent gene expression changes in an allopolyploid crop. *New Phytol.* 217, 367–377
79. Freeling, M. and Thomas, B.S. (2006) Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. *Genome Res.* 16, 805–814
80. Thompson, A. *et al.* (2016) Compensatory drift and the evolutionary dynamics of dosage-sensitive duplicate genes. *Genetics* 202, 765–774
81. Fernández-Mazuecos, M. and Glover, B.J. (2017) The evo-devo of plant speciation. *Nat. Ecol. Evol.* 1, 110
82. Chanderbali, A.S. *et al.* (2017) Evolution of floral diversity: genomics, genes and gamma. *Philos. Trans. R. Soc. B Biol. Sci.* 372, 20150509
83. Oyston, J.W. *et al.* (2016) Why should we investigate the morphological disparity of plant clades? *Ann. Bot.* 117, 859–879
84. Hetherington, A.J. *et al.* (2015) Do cladistic and morphometric data capture common patterns of morphological disparity? *Palaeontology* 58, 393–399
85. Nandi, O.I. *et al.* (1998) A combined cladistic analysis of angiosperms using rbcL and non-molecular data sets. *Ann. Mo. Bot. Gard.* 85, 137–212
86. Gower, J.C. (1971) A general coefficient of similarity and some of its properties. *Biometrics* 27, 857–871
87. Banks, J.A. *et al.* (2011) The *Selaginella* genome identifies genetic changes associated with the evolution of vascular plants. *Science* 332, 960–963
88. Minelli, A. (2018) *Plant Evolutionary Developmental Biology: The Evolvability of the Phenotype,* Cambridge University Press
89. Chen, C.-Y. *et al.* (2012) Lengthening of 3′UTR increases with morphological complexity in animal evolution. *Bioinformatics* 28, 3178–3181
90. Paterson, A. *et al.* (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proc. Natl. Acad. Sci. U. S. A.* 101, 9903–9908
91. Van de Peer, Y. *et al.* (2017) The evolutionary significance of polyploidy. *Nat. Rev. Genet.* 18, 411–424
92. Donoghue, P.C. and Benton, M.J. (2007) Rocks and clocks: calibrating the Tree of Life using fossils and molecules. *Trends Ecol. Evol.* 22, 424–431
93. Prasad, V. *et al.* (2011) Late cretaceous origin of the rice tribe provides evidence for early diversification in Poaceae. *Nat. Commun.* 2, 480
94. Iles, W.J.D. *et al.* (2015) Monocot fossils suitable for molecular dating analyses. *Bot. J. Linnean Soc.* 178, 346–374
95. Christin, P.-A. *et al.* (2014) Molecular dating, evolutionary rates, and the age of the grasses. *Syst. Biol.* 63, 153–165
96. Yang, Z. (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591
97. Steige, K.A. and Slotte, T. (2016) Genomic legacies of the progenitors and the evolutionary consequences of allopolyploidy. *Curr. Opin. Plant Biol.* 30, 88–93
98. Spoelhof, J.P. *et al.* (2017) Pure polyploidy: closing the gaps in autopolyploid research. *J. Syst. Evol* 55, 340–352
99. Bottani, S. *et al.* (2018) Gene expression dominance in allopolyploids: hypotheses and models. *Trends Plant Sci.* 23, 393–402
100. Garsmeur, O. *et al.* (2014) Two evolutionarily distinct classes of paleopolyploidy. *Mol. Biol. Evol.* 31, 448–454
101. Julca, I. *et al.* (2018) Phylogenomics of the olive tree (*Olea europaea*) disentangles ancient allo- and autopolyploidizations in Lamiales. *BMC Biol.* 16, 15
102. Thomas, G.W.C. *et al.* (2017) Gene-tree reconciliation with MUL-trees to resolve polyploidy events. *Syst. Biol.* 66, 1007–1018
103. Doyle, J.J. and Egan, A.N. (2010) Dating the origins of polyploidy events. *New Phytol.* 186, 73–85
104. Rothfels, C.J. *et al.* (2015) Natural hybridization between genera that diverged from each other approximately 60 million years ago. *Am. Nat.* 185, 433–442
105. Guillerme, T. (2015) dispRity: a modular R package for measuring disparity. *Methods Ecol. Evol.* 9, 1755–1763
106. Smith, S.Y. *et al.* (2017) A new species of *Pityostrobus* (Pinaceae) from the Cretaceous of California: moving towards understanding the Cretaceous radiation of Pinaceae. *J. Syst. Palaeontol.* 15, 69–81