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.

**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

---

**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.

---

1School 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).
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 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.

**Box 2. The Origins of WGD**

Traditionally, polyploids are recognized as originating from a single parent species (autopolyploidy, xx to xxy) or from two hybridizing species (allopolyploidy, xx + yy to xxy). 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.
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 different 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.
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).
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 autopolyplody (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 material).
associated with the expansion of families of regulatory genes. 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. This has been explained in terms of dosage balance and the need to maintain stoichiometric ratios of proteins within the cell. The dosage-balance hypothesis is exemplified during the diploidization process in allopolyploids, where exchanges can occur between homoeologous chromosomes of subgenomes. These exchanges can result in novel gene expression and gene copy number, 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.

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. These constraints may relax under different selection pressures, although a quantitative model of compensatory drift has also been proposed. 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. 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 Brassicaceae. 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 pattern expanded during the angiosperm-specific WGD. These genes are implicated in the origin and diversification of the flower, a structure that has shaped recent plant and animal evolution. The evolution of pentamerous flowers in the core eudicots also coincides with a genome triplication (gamma) in the core eudicots. The coincidence of the gamma event with this major synapomorphy, a large increase in the rate of diversification, and extensive genome reorganization 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. To make such a hypothesis testable, the increase in phenotypic complexity must be quantified for comparative analysis. 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. 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 (see File S3 in the supplemental information online). We used these data to calculate the dissimilarity between each taxon, as measured using Gower’s dissimilarity metric. To visualize this dissimilarity we performed non-
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 estimate 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).

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

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

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.