

## FOCUSED REVIEW

# Genome downsizing after polyploidy: mechanisms, rates and selection pressures

Xiaotong Wang<sup>1,2</sup> , Joseph A. Morton<sup>1,2</sup> , Jaime Pellicer<sup>1,3</sup> , Ilia J. Leitch<sup>1,\*</sup>  and Andrew R. Leitch<sup>2,\*</sup> <sup>1</sup>Royal Botanic Gardens, Kew, Surrey TW9 3AB, UK,<sup>2</sup>Queen Mary University of London, Mile End Road, London E1 4NS, UK, and<sup>3</sup>Institut Botànic de Barcelona (IBB, CSIC-Ajuntament de Barcelona), Passeig del Migdia sn, Barcelona 08038, Spain

Received 11 March 2021; revised 7 May 2021; accepted 13 May 2021; published online 2 June 2021.

\*For correspondence (e-mail a.r.leitch@qmul.ac.uk; i.leitch@kew.org).

## SUMMARY

An analysis of over 10 000 plant genome sizes (GSs) indicates that most species have smaller genomes than expected given the incidence of polyploidy in their ancestries, suggesting selection for genome downsizing. However, comparing ancestral GS with the incidence of ancestral polyploidy suggests that the rate of DNA loss following polyploidy is likely to have been very low (4–70 Mb/million years, 4–482 bp/generation). This poses a problem. How might such small DNA losses be visible to selection, overcome the power of genetic drift and drive genome downsizing? Here we explore that problem, focussing on the role that double-strand break (DSB) repair pathways (non-homologous end joining and homologous recombination) may have played. We also explore two hypotheses that could explain how selection might favour genome downsizing following polyploidy: to reduce (i) nitrogen (N) and phosphate (P) costs associated with nucleic acid synthesis in the nucleus and the transcriptome and (ii) the impact of scaling effects of GS on cell size, which influences CO<sub>2</sub> uptake and water loss. We explore the hypothesis that losses of DNA must be fastest in early polyploid generations. Alternatively, if DNA loss is a more continuous process over evolutionary time, then we propose it is a byproduct of selection elsewhere, such as limiting the damaging activity of repetitive DNA. If so, then the impact of GS on photosynthesis, water use efficiency and/or nutrient costs at the nucleus level may be emergent properties, which have advantages, but not ones that could have been selected for over generational timescales.

**Keywords:** genome downsizing, whole-genome duplications, DNA repair, DNA loss rate, selection pressures.

## INTRODUCTION

Polyploidy or whole-genome duplication (WGD) is prevalent amongst many vascular plant lineages and is a major driver of evolutionary novelty and speciation (Escudero and Wendel, 2020; Fox *et al.*, 2020). It has been suggested that all angiosperms (flowering plants) have experienced one or more episodes of polyploidy in their ancestry (Bowers *et al.*, 2003; Jaillon *et al.*, 2007; Jiao *et al.*, 2011; Landis *et al.*, 2018; Vision *et al.*, 2000; Wendel, 2015) and it is commonly encountered amongst individuals of extant species (Kolář *et al.*, 2017). Indeed, angiosperms have the highest incidence of ancestral WGDs among all land plant groups (Van de Peer *et al.*, 2017). Post-polyploidy genome divergence is typically associated with various ‘diploidisation’ processes, including loss of gene duplicates and descending

dysploidy, which returns the chromosome number back to a diploid-like form (Dodsworth *et al.*, 2015a; Mandáková and Lysak, 2018; Wendel, 2015). In addition, since the genome sizes (GSs) of polyploids are typically smaller than expected given the incidence of WGDs in their ancestry, diploidisation is also considered to be accompanied by extensive loss of DNA or ‘genome downsizing’ (Leitch and Bennett, 2004; Pellicer *et al.*, 2018; Zenil-Ferguson *et al.*, 2016).

An analysis of the distribution of GSs across angiosperms reveals that in contrast to gymnosperms (the other extant group of seed plants), most angiosperms have small GSs, with the modal size being only 0.58 Gb/1C (Dodsworth *et al.*, 2015b; Pellicer *et al.*, 2018), despite a huge 2400-fold range in GS (Figure 1a). Whilst the GS of the ancestral seed plant is unclear largely because of the

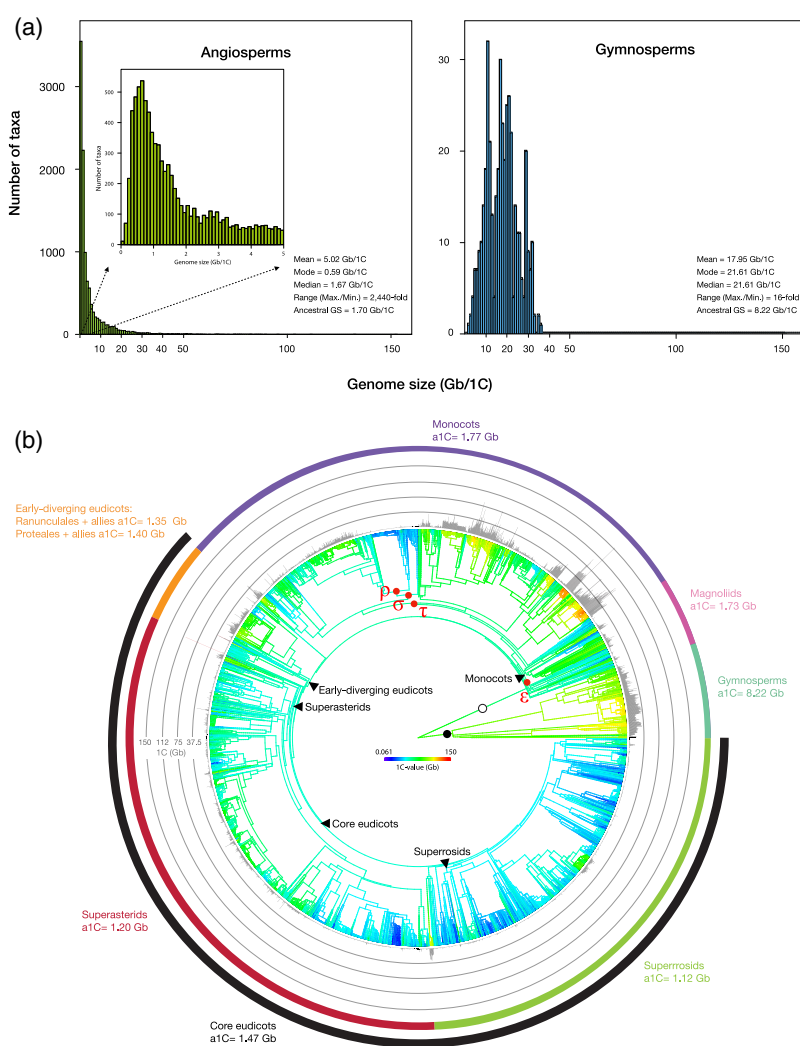
extinction of many key seed plant lineages (Hilton and Bateman, 2006), in contrast to gymnosperms, it is likely that genome downsizing has accompanied the divergence of angiosperms, given the frequency of polyploidy across many lineages (Van de Peer *et al.*, 2017). However, the mechanisms, rates and selection pressures driving this genome downsizing are not well understood. Here we explore each of these, in turn, to provide new insights into the impact of polyploidy on the immense diversity of GSs encountered in angiosperms and why most have small genomes.

### IMPACT OF WHOLE-GENOME DUPLICATIONS ON GENOME SIZE DIVERSITY IN ANGIOSPERMS

A large number of WGD events in the ancestry of angiosperm lineages is inferred from studies of mutation patterns in gene duplicates (Ks analysis), studies of divergence of paralogues (MAPS analysis) and synteny analyses (Jiao *et al.*, 2011; Li *et al.*, 2015; Van de Peer *et al.*, 2017; Landis

*et al.*, 2018). To look for signatures of the numerous WGD events experienced across angiosperms in the distribution of GSs amongst extant species, we reconstructed ancestral GSs across all angiosperms. Using the same approach as in Carta *et al.* (2020), but with an expanded dataset of 5866 species (Figure 1b), an ancestral GS of 1.70 Gb/1C at the base of all angiosperms was recovered, which is close to previous inferences based on smaller datasets (i.e. 1.42 Gb/1C by Puttick *et al.* (2015) and 1.69 Gb/1C by Carta *et al.* (2020)).

Figure 1b also shows that all major lineages (e.g. monocots, superasterids, superrosids) and most branches were similarly reconstructed as having small GSs, typically below 2 Gb/1C (see blue and green coloured branches), as reported previously using ancestral GS reconstructions (Puttick *et al.*, 2015; Soltis *et al.*, 2009; Simonin and Roddy, 2018; Carta *et al.*, 2020). We know, however, that the divergence of many of these species has been accompanied by numerous WGD events. For example, all Poaceae species



**Figure 1.** (a) The contrasting distribution of genome sizes (GSs) and mean, modal and median 1C-values and ancestral GS in angiosperms and gymnosperms. In angiosperms, despite GSs ranging 2400-fold, most species have small genomes (mean GS = 5.02 Gb/1C), with an ancestral GS of 1.70 Gb/1C. The distribution of GS between 0 and 5 Gb/1C is shown in detail. In gymnosperms, the GSs range only 16-fold but have a mean value five times greater than that of angiosperms (mean GS = 17.95 Gb/1C). Data for 10 770 angiosperms and 420 gymnosperms were taken from the Plant DNA C-values database (release 7.1) (Pellicer and Leitch, 2020). (b) Ancestral reconstruction of GS across the major lineages of seed plants. Reconstruction of ancestral GS (a1C) in Gb, following the approach used in Carta *et al.* (2020), but with an updated dataset of 5866 species extracted from the Plant DNA C-values database (Pellicer and Leitch, 2020) combined with the GBOTB large tree (GenBank taxa with a backbone provided by Open Tree of Life v 9.1) available from Smith and Brown (2018). After  $\log_{10}$  transforming 1C-values, a1C reconstruction was carried out under maximum likelihood and plotted using Phytools functions *fastAnc* and *contMap* (Revell, 2012). The rho ( $\rho$ ), sigma ( $\sigma$ ) and tau ( $\tau$ ) whole-genome duplication (WGD) events within monocots and the WGD at the base of all angiosperms (Jiao *et al.*, 2011) are shown. The outer ring shows the distribution of GS diversity across the tree with the value for each species shown as a bar (○ = ancestral branch leading to angiosperms, a1C = 1.70 Gb; ● = ancestral branch leading to gymnosperms, a1C = 8.22 Gb).

have undergone three WGD events ( $\sigma$ ,  $\rho$ ,  $\tau$ ), subsequent to a putative WGD event ( $\epsilon$ ) at the base of all angiosperms (Figure 1b, Van de Peer *et al.*, 2017; Clark and Donoghue, 2018). Indeed, the ancestral GS of the grass family Poaceae is 1.53 Gb/1C, which is smaller than that at the base of all monocots (i.e. 1.77 Gb/1C, Figure 1b). This extended analysis confirms that, despite the numerous ancient WGD events, overall genome downsizing subsequent to polyploidy must have been common and widespread during the evolution of angiosperms, or at least there is no signature of these polyploidy events in GS reconstructions using data from extant species.

Given this analysis, it is not surprising that the average GS of angiosperm species is not correlated with ploidy level predicted from chromosome counts (Leitch and Bennett, 2004; Zenil-Ferguson *et al.*, 2016), and most angiosperm species have smaller GSs than might be expected given the incidence of polyploidy in their ancestries. For example, the lineage leading to *Arabidopsis thaliana* is thought to have undergone multiple WGD events (Van de Peer *et al.*, 2017), such that it is predicted to be 48-ploid (Wendel, 2015). Given an ancestral GS for angiosperms of 1.70 Gb/1C (Figure 1b), this would give a GS of approximately 82 Gb/1C, which is approximately 512 times larger than its actual GS of just 0.16 Gb/1C (Bennett *et al.*, 2003). In contrast, available data in ferns (Clark *et al.*, 2016) and gymnosperms (Farhat *et al.*, 2019; Ickert-Bond *et al.*, 2020; Wu *et al.*, 2020) show that GS does reflect ploidy level estimated from chromosome counts, and for gymnosperms, outside of Gnetales, there have been no widely accepted ancestral WGD events since the common ancestor of all seed plants (Van de Peer *et al.*, 2017; Zwaenepoel and Van de Peer, 2019).

In the forthcoming sections, we explore possible mechanisms that may be responsible for genome downsizing, provide the first estimates of the rates of genome downsizing and suggest possible sources of selection that might have played a relevant role in driving genome downsizing.

## MECHANISMS UNDERPINNING GENOME SIZE DIVERGENCE

The remarkable variation in GS across extant seed plants arose through differences in the activities of processes that increase and decrease GS. Tandem duplications, unequal recombination, insertion/deletion of (retro)transposable elements, chromosome non-disjunction and polyploidy, coupled with genetic drift and/or selection, all contribute to GS change (Bennetzen *et al.*, 2005; Panchy *et al.*, 2016). Of these, polyploidy is a driver of step changes in GS, acting to generate multiple copies of the genome, both genic and non-genic.

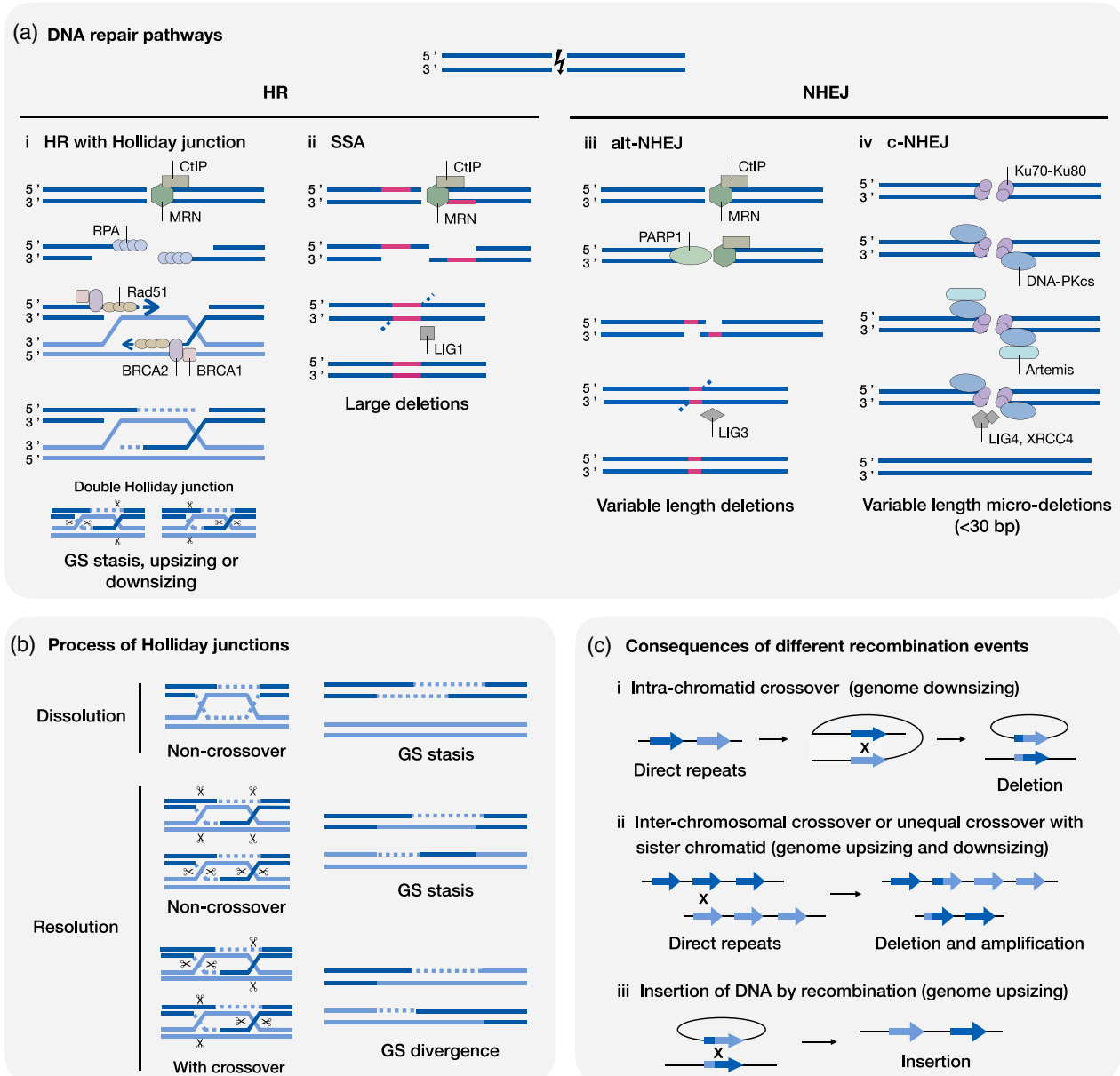
It is notable that many of the mechanisms that lead to GS change are associated with DNA double-strand break (DSB) repair involving DNA recombination (Schubert and Vu, 2016). These include (i) the deletion and insertion of

transposable elements, particularly, retroelements in plants (e.g. Cossu *et al.*, 2017; Hawkins *et al.*, 2009; Senerchia *et al.*, 2014), (ii) the amplification and deletion of tandem repeats (Viales *et al.*, 2019) and (iii) structural rearrangements of chromosomes which can move sequences into new chromatin domains with different local dynamics in recombination frequencies, leading to altered patterns of DNA loss and gains (Ren *et al.*, 2018). In addition, at meiosis, aberrant control of recombination can lead to the formation of univalents and multivalents, resulting in non-disjunction and polyploidy, all impacting GS. Thus, insights into GS diversity require a recognition of the activities of DSB repair and DNA recombination pathways.

There are numerous ways in which DSBs can be repaired (Figure 2a), each of which can lead to different outcomes in terms of DNA losses and gains, and if they remain stable over multiple generations it will drive GS divergence (Schubert and Vu, 2016). For example, homologous recombination (HR) involving Holliday junctions has the potential to lead to GS stasis, upscaling or downsizing, depending on the sequences involved and the way that junctions are resolved (Figure 2b). In contrast, single-strand annealing (SSA) and non-homologous end joining (NHEJ) will likely lead to varying degrees of genome downsizing (Figure 2a). In tobacco (*Nicotiana tabacum*), it is estimated that most DNA repair occurs via SSA, with about one-third being HR (Orel & Puchta, 2003), but such ratios could vary, depending on cell type, cell cycle and developmental stage. Variation can also arise in the extent of resectioning (Figure 2a) that occurs at a DSB, which, in turn, can influence the particular DNA repair pathway adopted and hence the amount of DNA lost during DSB repair (Clouaire and Legube, 2019; Shibata *et al.*, 2011). Mechanisms can also be intrinsically biased in their action; for example, in *Drosophila melanogaster* (fruit flies), non-allelic gene conversion (Figure 2b) can lead to approximately 3.5 times more deletions than insertions (Assis and Kondrashov, 2012).

There can also be differences in the types of repair pathways adopted dependent on the genomic location of the sequences or the proximity of homologous sequences. For example, in transgenic *A. thaliana*, deletion of DNA occurred in about one-third of repair events involving HR, if templating was possible from adjacent sequences, as in tandemly arranged genes (Orel *et al.*, 2003).

Furthermore, studies have shown that the repair of DSBs can differ between species, leading to different outcomes that have the potential to impact GS. For example, distinct differences in the DSB repair pathways were observed between *A. thaliana* and barley (*Hordeum vulgare*), which differ 35-fold in GS (i.e. 157 and 5500 Mb/1C, respectively) (Vu *et al.*, 2017). Whereas the DSB repairs in *Arabidopsis* were typically skewed towards more frequent and larger deletions, barley was characterised by repairs that led to



**Figure 2.** (a) DNA repair pathways. The repair of double-strand breaks (DSBs) in the DNA (solid blue lines) is achieved via either homologous recombination (HR, i and ii) or non-homologous end joining (NHEJ, iii and iv) pathways. (i) In HR involving Holliday junctions, resectioning (i.e. the removal of DNA) generates two 3' single-stranded DNA (ssDNA) overhangs which provide a platform for recruiting the HR repair-related proteins (e.g. Rad51, BRAC1 and BRAC2). The 3' single-strand overhangs then invade homologous DNA sequences, to form a structure known as the D-loop. A double Holliday junction arises after DNA synthesis (indicated by dashed blue lines extending from the DNA strands) replaces those DNA sequences lost by resectioning, and strand ends are ligated. (ii) In contrast, the repair of a DSB via the single-strand annealing (SSA) pathway of HR is associated with DNA loss. This is caused by resectioning, which can be extensive as it will continue until regions of sequence homology (as indicated by red segments in the DNA strands) are encountered between the two strands. These homologous regions will then form a duplex which is ligated, while any 3' overhangs generated that lack sequence homology are deleted. (iii) Alternative NHEJ (alt-NHEJ) requires only microhomologies (2–25 bp) generated by resectioning to enable strand annealing and end joining, accompanied by variable lengths of deletions and minor insertions at the break sites ( $\leq 3$  bp). (iv) In contrast, classical or canonical NHEJ (c-NHEJ) does not require resectioning but is usually associated with short microdeletions (<30 bp) and minor insertions at the DSB ( $\leq 3$  bp). See Table 1, which lists the functions and abbreviations for the proteins shown.

(b) Processing of Holliday junctions. How the Holliday junctions in (a) are processed will influence the genetic outcome and impact GS evolution. The Holliday junctions can undergo dissolution without cleavage of the DNA, resulting in no GS change. Alternatively, depending on which two of the four DNA strands are cleaved (as shown by the alternative positions of the scissors), the outcome will either be a non-crossover or a crossover event, the latter potentially leading to GS divergence.

(c) Consequences of different recombination events. (i) If recombination occurs between two homologous sequences (e.g. repeats) on the same chromosome, intervening DNA can be lost; (ii) if recombination occurs between homologous sequences on sister chromatids or homologues which are not in precise register, then the resulting chromatids can gain or lose DNA; (iii) and if recombination occurs between chromosomal and extrachromosomal DNA (e.g. retroelements or extrachromosomal DNA), then DNA can integrate at regions of homology.

more and larger insertions of DNA. Such differences were considered to arise in part through the more extensive resectioning at DSBs observed in *Arabidopsis* compared with barley, perhaps due to less efficient protection of free DNA ends at the breakpoints. Similarly, in comparisons of multiple species of *Drosophila* with *Laupala* (crickets), whose genomes are 11-fold larger, the studied non-long terminal repeat (non-LTR) transposable elements in *Laupala* had more insertions and lower rates of DNA deletion caused by fewer large (>15 bp) indels (Petrov *et al.*, 2000). In the tunicate *Oikopleura dioica*, genes essential for canonical NHEJ are missing and DSBs are instead repaired using alternative NHEJ (alt-NHEJ). The authors suggest that the mutations by alt-NHEJ ultimately led to the organism's small compact genomes (Deng *et al.*, 2018). In all such examples, differences in DSB repair over multiple generations can be expected to lead to GS divergence.

Overall, the particular repair pathways used (Figure 2a) can vary at the molecular level (e.g. histone phosphorylation levels, expression levels of repair protein genes), at the cellular level (e.g. cell type, cell cycle), between developmental stages and with species and environment (e.g. under stress) (Boyko *et al.*, 2006a,b; Clouaire and Legube, 2019; Harris *et al.*, 2018; McVey and Lee, 2008; Shrivastav *et al.*, 2008; Vu *et al.*, 2014). Any bias in the germline lineage over generations in the relative frequencies of these DSB repair pathways or the activity of the different proteins involved (Table 1) may result in GS divergence. This can potentially arise from bias in (i) the relative activities of the DNA repair mechanisms (Figure 2a), (ii) the tendency for Holliday junctions to resolve with a crossover or not (Figure 2b), (iii) the frequency of intra-chromatid crossovers (Figure 2c), (iv) the frequency of misalignment between homologous chromosomes or sister chromatids leading to inter-chromosomal or unequal crossovers, respectively (Figure 2c) and/or (v) the tendency of chromosomes to recombine with extrachromosomal DNA (e.g. during retroelement insertion). If selection is favouring any such bias such that DNA is lost, it will drive genome downsizing.

### PREDICTING RATES OF GENOME SIZE CHANGE FOLLOWING WHOLE-GENOME DUPLICATIONS

To gain insights into the drivers of genome downsizing, it is necessary to determine the rate of DNA loss following polyploidy, to see if DNA losses are sufficiently large to be visible to selection. Given that such data have not previously been estimated to the best of our knowledge, the following approach was taken to provide insights.

We used available information on (i) the phylogenetic relationships between species, (ii) GS data and (iii) the estimated dates of WGD events identified in Van de Peer *et al.* (2017) to predict the rate of GS change following the most recent WGD event (Figure 3). This was done by

reconstructing the ancestral GSs before WGDs and comparing the multiple of these values (to account for the WGDs) with the actual GSs. This required reconstructing the GSs of individual lineages leading to a WGD event, requiring data from 'diploid' lineages that diverged prior to that WGD event, a situation that occurs seven times on the phylogeny in Figure 3. The results reveal that the rate of DNA loss per generation or per million years is low (Figure 3). For example, in the lineage leading to the walnut (*Juglans regia*), which is predicted to have experienced a WGD approximately 50 million years ago, a predicted ancestral GS of 712 Mb before the WGD event means an expected GS of  $2 \times 712 \text{ Mb/1C} = 1412 \text{ Mb/1C}$  in the extant *J. regia*. However, its actual GS is 620 Mb/1C, indicating that over half the DNA has been lost since the WGD. Assuming a generation time of 30 years, this then equates to only 482 bp/generation. Over seven species, representing both eudicots and monocots, the GS losses per generation are similarly low (4–482 bp/1C per generation, depending on species). It is, therefore, difficult to explain what may be driving genome downsizing following WGD given such small changes in GS are taking place per generation (i.e. in the  $10^1$ – $10^2$  bp/1C range compared with the GSs of most plant species (in the  $10^8$ – $10^{11}$  bp/1C range), which are  $10^6$ – $10^{10}$  orders of magnitude bigger). Such small losses in DNA are unlikely to be visible to selection, so, assuming a constant rate of DNA loss per generation, we would expect GS to drift, leading to either genome upsizing or downsizing subsequent to polyploidy. Therefore, the following question arises 'why is there a general trend towards genome downsizing?'. We address this question in the following sections.

### POTENTIAL SOURCES OF SELECTION ACTING ON GENOME SIZE FOLLOWING WHOLE-GENOME DUPLICATIONS

There are at least two possible sources of selection acting to enhance plant fitness following a WGD event and leading to genome downsizing. In species with large GS these are: (i) the increased nutrient costs of nucleic acids in species and (ii) the impact of increased cell size detrimentally influencing gas exchange parameters and hence photosynthesis efficiency and water use efficiency. Each of these is considered below.

#### Selection for small genomes generated by the nutrient costs of nucleic acids

Nucleic acids are comprised of 15% nitrogen (N) and 9% phosphorus (P) and are, therefore, amongst the most N- and P-demanding molecules of the cell (Sterner and Elser, 2002). However, when nutrients are limited, as is typical of many natural systems, the allocation of N and P to nucleic acid synthesis, maintenance, repair and transcription must be traded off with other key processes, including

**Table 1** Key proteins involved in the repair of double-strand breaks (DSB) in DNA. The table lists some of the main proteins involved in the four DSB repair pathways shown in Box 2, whose activities, if biased over multiple generations, are predicted to lead to directional genome size change during species divergence

Pathway	Key proteins	Molecular function	References
c-NHEJ	Ku70- Ku80	DSB recognition and DSB-specific end-binding protein	(Chiruvella <i>et al.</i> , 2013)
	DNA- PKcs	Catalytic subunit of DNA-PK	(Chiruvella <i>et al.</i> , 2013)
	Artemis	End trimming	(Brandsma and Gent, 2012)
	XRCC4 LIG4	Recruitment of ligase and interaction with Lig4 End ligation	(Mari <i>et al.</i> , 2006) (Ranjha <i>et al.</i> , 2018)
alt- NHEJ	MRN - CtIP	The MRN complex (Mre11–Rad50–Nbs1) co-acts with CtIP and starts resectioning at breakpoints to generate ssDNA	(Brandsma and Gent, 2012; Uziel <i>et al.</i> , 2003; Zamariola <i>et al.</i> , 2014)
	PARP1	Synapsis activity	(Ottaviani <i>et al.</i> , 2014)
	LIG3	End ligation	(Brandsma and Gent, 2012)
HR	MRN- CtIP	The MRN complex (Mre11–Rad50–Nbs1) co-acts with CtIP and starts resectioning at breakpoints to generate ssDNA	(Brandsma and Gent, 2012; Uziel <i>et al.</i> , 2003; Zamariola <i>et al.</i> , 2014)
	RPA	Attaches to ssDNA and removes secondary structure in the DNA	(Sugiyama <i>et al.</i> , 1997)
	BRCA1	Promotes the assembly of other HR proteins	(Lafarge and Montane, 2003; Prakash <i>et al.</i> , 2015)
	BRCA2	Mediates the recruitment of Rad51	(Holloman, 2011; Siaud <i>et al.</i> , 2004; Walsh, 2015)
SSA	Rad51	Promotes single-strand end invasions of homologous intact duplexes	(Gao <i>et al.</i> , 2014)
	MRN- CtIP	The MRN complex (Mre11–Rad50–Nbs1) co-acts with CtIP and starts resectioning at breakpoints to generate ssDNA	(Brandsma and Gent, 2012; Uziel <i>et al.</i> , 2003; Zamariola <i>et al.</i> , 2014)
	LIG1	End ligation	(Ceccaldi <i>et al.</i> , 2016)

c-NHEJ, canonical non-homologous end joining; alt-NHEJ, alternative non-homologous end joining; HR, homologous recombination; SSA, single-strand annealing.

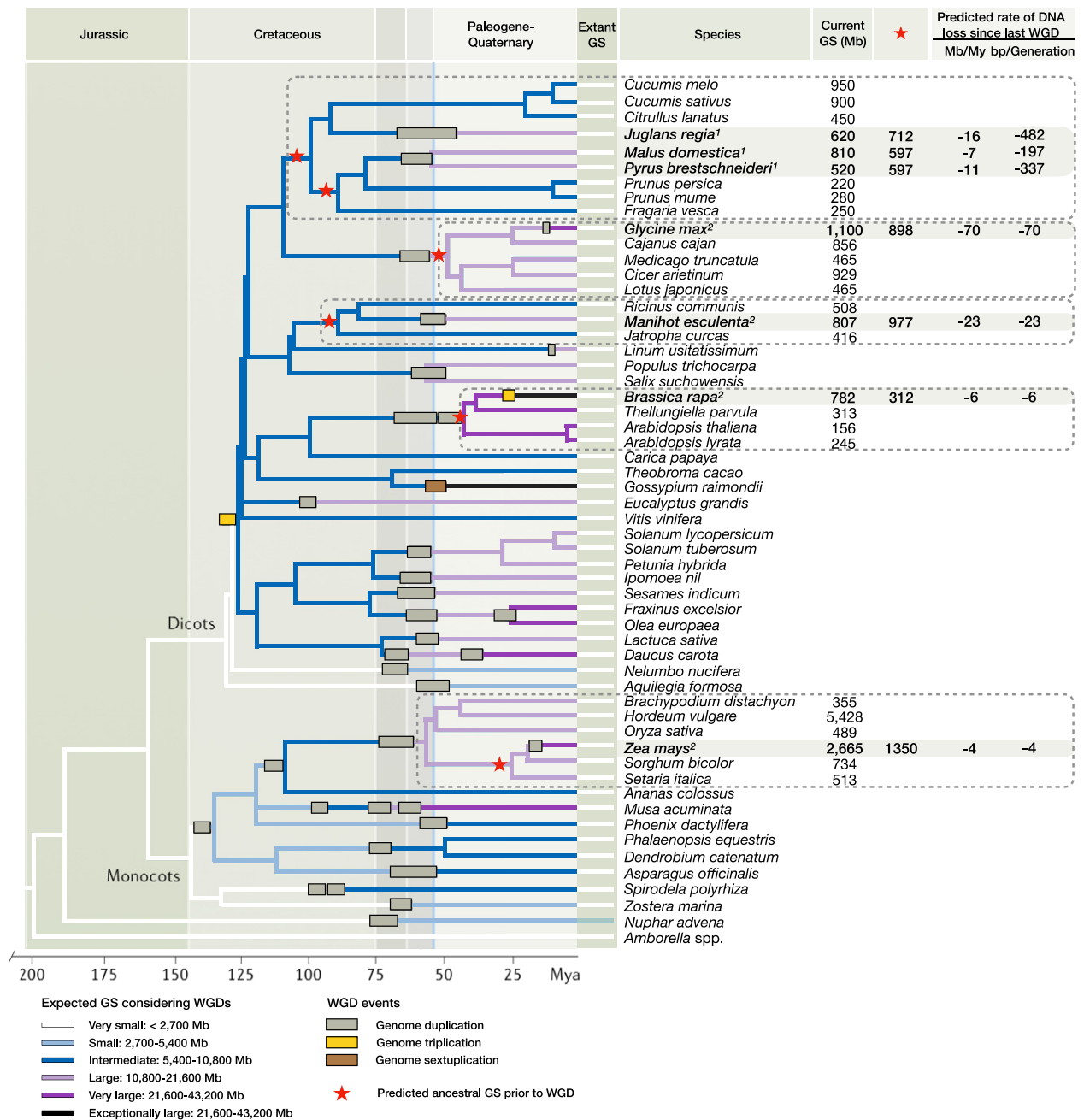
photosynthesis and protein synthesis, and hence impacts growth and fitness (Hessen *et al.*, 2010).

Evidence of selection for N-efficient nucleotide and amino acid usage has been observed in bacteria (Shenhav and Zeevi, 2020) and plants (Acquisti *et al.*, 2009a,b; Kelly, 2018). For example, Kelly (2018) revealed the strongest selection for codons with efficient N usage in the genomes of plants needing more N for photosynthesis (i.e. C<sub>3</sub> versus C<sub>4</sub> plants). These examples suggest there is selection for N and P use efficiency at the genomic level and hence support a model in which species with larger genomes may be more compromised when competing with smaller-genome plants in nutrient-poor environments. This prediction is supported by studies showing that GS influences the ecological selection of species under nutrient limitation. GS analysis of species growing in long-term fertilised grassland in the United Kingdom and Germany showed that polyploid species with large GSs competed most successfully only in the presence of high levels of N and P, while species with smaller GSs were more successful when one or both nutrients were limited (Guignard *et al.*, 2016; Šmarda *et al.*, 2013).

These data support the hypothesis that GS represents a significant cost to a plant, influencing fitness. In species of the plant genus *Primulina* that are specialists to limestone

karst, leaf N levels are positively correlated with GS, indicative of a nutrient cost of the genome (Kang *et al.*, 2015). Furthermore, lab-based experiments using both *Chamaerion angustifolium* and *A. thaliana* found that synthetic neotetraploids (with double the GS of diploids) only had a competitive advantage over their diploid progenitors when growing in both N- and P-enriched conditions, where they generated more biomass and had increased reproductive fitness (Anneberg and Segraves, 2020; Bales and Hersch-Green, 2019; Walczyk and Hersch-Green, 2019). Of note, the parasitic species mistletoe (*Viscum album*) has a particularly large genome (approximately 85 Gb/1C), perhaps because its parasitic lifestyle means there is no N and P limitation and the genome has been free to drift to large GS. In contrast, some of the smallest genomes are found in insectivorous plants (e.g. *Genlisea* and *Utricularia*), which typically have a streamlined genome, perhaps due to evolutionary or ecological selection on GS in low-nutrient environments (Veleba *et al.*, 2020).

Selection for increased N and P use efficiency under nutrient limitation may, therefore, result in selection pressures to eliminate DNA, especially 'junk' DNA such as certain repetitive DNA sequences. The selective advantage of this DNA elimination may not only lie in lower N and P costs of the repetitive DNA itself (e.g. arising from proteins



**Figure 3.** The predicted impact of major whole-genome duplication (WGD) events on genome size (GS) for 56 angiosperms and the predicted rate of GS change following the most recent WGD for five different angiosperm lineages on a dated phylogenetic tree. Each rectangle on a branch indicates a WGD event (grey, yellow or brown). The colour of the branches shows the predicted GS considering the impact of the WGD. The colour of branches in the 'Extant GS' panel indicates the GS range in which the GS of the species falls. The polyploid species with predicted rates of GS change following the most recent WGD event are given in red for five angiosperm lineages (dotted grey boxes). All extant GSs are given as 1C-values. Ancestral GS reconstruction of the sister lineages and outgroups to the polyploid (excluding the polyploid itself) provides an estimate of the ancestral GS before the WGD event (red star). For each polyploid species analysed, a comparison is made between the multiple of this ancestral GS (to account for the WGD, i.e. double the value for a genome duplication, or triple for a genome triplication) and the actual GS of the polyploid species to calculate the rate of DNA loss since the WGD event. The image with polyploidy events is modified from Van de Peer *et al.* (2017), while GS data are taken from the Plant DNA C-values database (Pellicer and Leitch, 2020). Mya, million years ago; My, million years; <sup>1</sup>assumed generation time of 30 years; <sup>2</sup>assumed generation time of 1 year.

needed for packaging DNA into chromatin [e.g. histones], DNA replication and maintenance and DNA repair), but also in the RNA costs associated with repeats. The latter

includes transcriptional costs associated with the selfish (retro)transposition of repeats in their amplification as well as costs associated with their silencing, via canonical and

non-canonical RNA-directed DNA methylation (RdDM) pathways (Kenchanmane Raju *et al.*, 2019). Although the magnitude of these costs and the extent to which cell nutrient allocation trade-offs are influenced by GS remain to be determined, it appears from the available experimental data noted above that these costs may be visible to selection in angiosperms.

It is also possible that selection for species with small genomes and increased nutrient use efficiency is an emergent property of genome downsizing processes that are not themselves under such selection (e.g. some DNA recombination and repair pathways). Over many generations, the cumulative effects of many small DNA losses could nevertheless lead to sufficient nutrient use efficiency gains that individuals with substantially smaller GSs have ecological advantages.

### **Selection for small genomes to maximise photosynthesis and/or water use efficiency**

Genome size is positively correlated with minimum cell size (Beaulieu *et al.*, 2008; Jovtchev *et al.*, 2006), as the quantity of nuclear DNA plays a role in setting the lower limit in the volume of the nucleus and consequently a lower limit in the size of the cell that houses and supports it (Cavalier-Smith, 2005; Doyle and Coate, 2019; Greilhuber and Leitch, 2013; Šimová and Herben, 2012). These patterns are observed in animals and plants. In animal cells, such as erythrocytes, which carry oxygen to tissues, the scaling relationship between GS and cell size impacts the surface area-to-volume ratios of cells and hence the efficiency of oxygen delivery (i.e. species with bigger GS and hence larger cells have a lower surface area-to-volume ratio, which reduces the speed of oxygen delivery). Selection for small GS in animals with high rates of metabolism (e.g. bats and hummingbirds) may be driven by the need to maximise rapid delivery of oxygen. In contrast, selection on GS may be relaxed in animals with slow rates of metabolism and hence lower oxygen demands (e.g. lungfish), enabling GS and cell size to drift upwards (Gregory, 2001). In plants, such scaling relationships can impact gas exchange in the leaf by determining the minimum stomatal pore size, which, in turn, is inversely correlated with stomatal density (Beaulieu *et al.*, 2008). Indeed, ploidal levels and/or GS impact the density of cells in leaves, the maximum rate of carboxylation in photosynthesis and water use efficiency (Roddy *et al.*, 2020; Wilson *et al.*, 2021), which may also explain why GS is also correlated with seed number and seed mass (Herben *et al.*, 2012). [Correction added on 31 July 2021, after first online publication: The text in the first line of this paragraph has been corrected in this version.]

Plant species with larger genomes typically have larger, sparser stomata, limiting stomatal conductance of CO<sub>2</sub> and water (Franks and Farquhar, 2001; Simonin and Roddy,

2018) and slowing the speed at which stomata respond to daily changes in environmental conditions (Lawson and Vialet-Chabrand, 2019; Roddy *et al.*, 2019). Given these relationships, selection acting upon stomatal size to optimise trade-offs between photosynthesis and water use efficiency may select for genome downsizing, particularly under water stress and during periods of low atmospheric CO<sub>2</sub> concentrations, as exist today (Franks and Beerling, 2009). Under such conditions, small stomata at a high density better supply CO<sub>2</sub> to photosynthetic tissues, while also increasing water use efficiency as small stomata close more rapidly than larger ones (Lawson and Vialet-Chabrand, 2019; Roddy *et al.*, 2019). This is supported by a study of vernal geophytes, which showed that species with adaptations to maintain humid conditions and higher CO<sub>2</sub> concentrations around the stomata had larger stomata and GSs than those without (Veselý *et al.*, 2020). Nevertheless, existing evidence relies mostly on correlations, and hence more research is urgently required to demonstrate the causal effects of GS change on photosynthetic rates.

However, as with nutrient use efficiency potentially being an emergent property of other processes acting over multiple generations (see above), so too could any beneficial effects of smaller GS on cell size, and its consequences for photosynthesis and water use efficiency be a by-product of those processes. If so, species with smaller genomes may have advantages, but not ones that would have been under selection over generational timescales.

### **HOW MIGHT GENOME DOWNSIZING ARISE?**

Lynch and Conery (2003) suggested that population sizes of complex eukaryotes may be too small for selection to act on GS. Furthermore, Lynch and Marinov (2015) proposed that the energetic costs of nucleic acids (DNA replication, transcription and translation costs of associated histones) are unlikely to be sufficient to offer a selective pressure for genome downsizing. Instead, they argued that GS has drifted to be large in the most complex multicellular organisms with relatively small effective population sizes. However, these analyses are based on theoretical predictions using data from prokaryotes and eukaryotes with small GSs (i.e. up to 157 Mb/1C). It is, therefore, possible that such dynamics may not be applicable across the full 2400-fold range of GSs found in plants, with new dynamics becoming apparent as the GS increases. Indeed, the energetic cost related to GS may be more visible to selection than might be expected from these analyses, and other costs (e.g. N and P) have not been explored so thoroughly. Certainly, the evolution of GS does not appear to be dominated by genetic drift in angiosperms, as this would give rise to a normal distribution of GS centred around large genomes (due to the frequency of ancestral WGDs in angiosperms), rather than the strongly skewed distribution dominated by small genomes (Figure 1a). By



contrast, gymnosperms, which have a much lower frequency of polyploidy compared with angiosperms, have a near-normal distribution of GS (Figure 1a). Thus, the overall loss of DNA following polyploidy in angiosperms points towards a role for selection to reduce GS.

In order to explain the observation that most angiosperms have small GSs despite (i) the high frequency of polyploidy, (ii) the small amounts of DNA predicted to be lost per generation following WGD (Figure 3) and (iii) the power of drift in driving GS divergence (Lynch and Conery, 2003), we need to establish where a strong selection pressure might act on GS or where processes might be triggered that indirectly lead to DNA loss.

### Genome downsizing arises through selection against genome upsizing

It is possible that genome downsizing arises because processes that result in increases in GS (e.g. [retro]transposition) may be more deleterious than those causing downsizing (e.g. deletions) and be more strongly selected against. This might then lead to the rates of DNA loss exceeding the rates of gain, which would inevitably lead to GS decreasing over time. This might continue until such times that DNA losses themselves (e.g. through loss of gene function) become more deleterious than gains, when the balance of selection pressures may reverse. In the context of a young polyploid such a scenario may, indeed, occur, since gene redundancy might provide minimal selection against gene loss in early generations, yet amplification of elements, e.g. via (retro)transposons, may have costs associated with mobility and insertion, including RdDM costs and those associated with disrupted gene networks (e.g. through altered promoter and regulator activity) (Lisch, 2013). In contrast, in ancient polyploids, duplicate gene losses may have resulted in little or no gene redundancy and so selection against processes that cause the deletion of DNA may be stronger than those associated with GS increases.

### Genome downsizing is selected for in early-generation polyploids

The only way to reconcile a role for photosynthesis, water use efficiency and/or N and P costs in acting as selection pressures driving genome downsizing is to assume that DNA loss is extensive and rapid in the early polyploid generations. For example, this has been reported in the first two generations of colchicine-doubled *Phlox drummondii* (25% decrease in GS over two generations, Raina *et al.*, 1994). Such large reductions in GS might then be visible to selection. However, direct evidence for rapid reductions in GS following natural polyploid formation is currently lacking, although we do know from studies on retention patterns of gene duplicates in yeast lineages that gene loss is at first random, impacting orthologs and paralogs equally, but

later, amongst the more slowly diverging genes, the losses are no longer random and instead are frequently targeted at specific copies (Scannell *et al.*, 2007). Such data suggest that genome dynamics may change with the age of the polyploid, and potentially such changes could also impact rates of DNA loss. Certainly, in early-generation polyploids there can be a much higher frequency of recombination between homologs and homeologs leading to multivalent formation during meiosis compared with older, more established polyploids (Lloyd and Bomblies, 2016; Yant *et al.*, 2013). This will not only generate considerable chromosomal diversity in young polyploids, but also impact the segregation of diversity, including indels. For example, some young polyploid species show much karyotype diversity between individuals, as in allotetraploids of *Tragopogon* (approximately 40 generations and 80 years old) (Chester *et al.*, 2012). Indeed, it has been hypothesised that a positive feedback loop exists, whereby homeologous recombination in young allopolyploids causes depletion in DNA mismatch repair proteins, which enhances aberrant recombination and DNA loss, leading to even more homeologous recombination in future generations (Comai, 2000). Potentially from such diversity, selection could favour variants with smaller GS. It would be particularly relevant to know if the recombination frequency and/or fitness of polyploid individuals with smaller GS increase when growing under nutrient-limited or water stress conditions, but such data are lacking.

### Genome downsizing is a byproduct of high recombination frequencies in polyploids

Genome downsizing may indirectly arise as a consequence of changes in recombination frequencies triggered by WGD. Initially, WGD is expected to increase the number of recombination events per generation assuming regular bivalent formation, because of the increased chromosome number in the polyploid nucleus and the typical occurrence of one or two chiasma per bivalent (Stapley *et al.*, 2017). If such increased meiotic recombination is associated with an increased frequency of recombination-based DNA excision (e.g. frequency of intra-chromatid recombination), this could lead to greater DNA losses in early polyploids, which would decline as polyploid genomes diploidise. Nevertheless, in young polyploids, selection is predicted to favour reduced chiasma frequency to facilitate bivalent formation (Le Comber *et al.*, 2010; Yant *et al.*, 2013). Over longer timeframes, reductions in chromosome number as part of the diploidisation process following WGD (Dodsworth *et al.*, 2015a; Wendel, 2015) may reduce chiasma frequency. Chromosome number reduction may itself be driven by selection on chromosome fusions that bring adaptive genes into linkage groups (Hoffmann and Rieseberg, 2008; Ritz *et al.*, 2017). Whatever the cause, the return to a diploid-like chromosome number with polyploid age might act to reduce the rate of DNA loss over time.

### Genome downsizing is a byproduct of selection for a nutrient-efficient transcriptome

Although data are limited, studies have shown that the size of the transcriptome in a cell can change with development (Sulpice *et al.*, 2014), in association with polyploidy (Doyle and Coate, 2019) and growth rate (Reef *et al.*, 2010). Nevertheless, available estimates of transcriptome size reveal that RNA:DNA ratios typically range from about 1:1 to 20:1 in plants (e.g. *Avicennia marina* [0.51 Gbp/1C]; Reef *et al.*, 2010). We are unaware of evidence that suggests that the RNA:DNA ratio scales with GS. Assuming so, species with larger genomes tend to have bigger transcriptomes. Much of the increase in transcriptome size may arise through the transcription of repeats, which themselves scale in number with GS (Novák *et al.*, 2020) and which are likely to induce a rising nutrient cost with GS upon which selection may act.

Although there are rather few studies, it is estimated that as much as 50–80% of the total cellular P is contained in nucleic acids (Hessen *et al.*, 2010), probably mostly ribosomal RNA, but the value of this component can also be regulated, as can the relative size of mRNA and other RNA fractions (Doyle and Coate, 2019), which include transcripts from repeats. Such differences arise through variation in (i) rates of transcription, (ii) the amount of RNA stored (especially pre-ribosomes in the nucleolus) and/or (iii) rates of RNA cycling (i.e. transcription and breakdown, Hessen *et al.*, 2010).

As previously noted, the nutrient costs of the transcriptome can influence the selection of N-efficient nucleotides in plants with less N-efficient photosynthesis (Kelly, 2018). Perhaps these costs also influence GS indirectly through selection towards lower N and P costs from the transcriptome by reducing the transcription of repetitive DNA. Wasteful transcription results in reductions in available N and P for photosynthesis and biomass production, a cost that may select for processes that eliminate repeats. Deletion of repeats could potentially remove a substantial nutrient burden, far beyond that of the genome itself, since the transcription of repeats is ongoing in every cell, whereas the cost of the DNA in the genome may be less demanding once it has replicated (but still requiring DNA maintenance and repair). Even repeats present in low copy numbers may have a substantial cost if they are transcriptionally active, as seen in some transposing repeats (Dias *et al.*, 2015). In addition to transcriptome costs, even the small-scale elimination of an actively transcribed and/or transposing repeat could have a selective advantage if it disrupted gene regulation and function (Lisch, 2013).

We know that polyploidy can stimulate much epigenetic reprogramming (Nieto Feliner *et al.*, 2020), which can lead to increased transcriptional and transpositional activity of repeats (Petit *et al.*, 2010) that are not silenced by the epigenetic machinery. In early polyploid lineages, this may

#### Box 1. Summary of the review.

- Following polyploidy or WGDs, the genomes of angiosperms are typically downsized, pointing to the action of selection on GS.
- The differential activities of HR and NHEJ DNA repair pathways might be responsible for genome downsizing and the divergence of GS following polyploidy.
- Ancestral GS reconstructions suggest that rates of genome downsizing (base pairs per generation) following polyploidy are likely to be very low, raising the conundrum as to how selection might act.
- Possible sources of selection driving genome downsizing include reduced nutrient costs of the genome and transcriptome and/or reductions in cell size, facilitating efficient gas exchange for photosynthesis and enhanced water use efficiency.
- Genome downsizing may be a byproduct of processes that lead to smaller genomes, and derived smaller might offer a selective advantage as an emergent property.

#### Box 2. Open questions.

- Are genomes of different angiosperm lineages downsized at different rates following polyploidy?
- How do rates of GS loss after polyploidy correlate with the activities of different mechanisms of DNA recombination and DNA repair?
- To what extent does the transcription of repeats provide a selection pressure for genome downsizing?
- To what extent does GS itself directly contribute to selection for genome downsizing after polyploidy?

result in the upregulation of NHEJ and/or unequal HR in DNA repair, leading to elevated frequencies of DNA elimination (Figure 2). Such upregulation may have a selective advantage if the eliminated DNA 'kills' actively transcribed and transposing repeats. Potentially, gene redundancy in young polyploids means that there is little selective disadvantage to any random loss of genes this process might also cause. The fixation of such a process may set in train processes that ultimately lead to substantial and significant genome downsizing over time. This would continue until gene redundancy is exhausted and there is once again selection for more faithful HR DNA repair, without

substantial, or any, DNA loss. A prediction, therefore, is that the altered regulation of, and selection for, different recombination pathways will vary with polyploid age.

### CONCLUDING REMARKS AND FUTURE PERSPECTIVES

Clearly, there is still much to learn about the mechanisms, rates and selection pressures to distinguish between these different hypotheses underpinning genome downsizing following polyploidy. Nevertheless, addressing these uncertainties is now within reach given the growing availability of high-quality chromosome-level genome assemblies, large transcriptomic and methylome datasets, resolved phylogenetic relationships between species and advances in machine learning approaches. These data will provide information on genome structure and organisation, the recombinational landscape and measures of selection acting on nucleotides, codons (in the genome and transcriptome) and amino acids. Together they will enable us to derive detailed insights into how and why genomes are downsized by searching for signals of selection across genes, non-coding sequences and chromosomal regions, as well as signals of DNA deletions and insertions in comparative analyses of polyploids of different ages (recent to ancient). In addition, there is a need to combine these genomic studies with robust ecological experiments to track the fitness (e.g. photosynthesis and water use efficiency) of polyploids with different GSs and from different phylogenetic lineages growing under different environmental conditions. Together, these data will enable us to distinguish between species-specific and general responses to selection pressures, and hence generate a holistic understanding of the evolutionary and ecological forces responsible for genome downsizing.

### ACKNOWLEDGEMENTS

XW is funded by the China Scholarship Council (CSC, Grant 201908370151) from the Ministry of Education of the P.R. China. JP is supported by a Ramón y Cajal Fellowship (2017-RYC-2274) funded by the Ministry of Science and Technology (Government of Spain).

### CONFLICT OF INTEREST

The authors declare no conflict of interest.

### REFERENCES

Acquisti, C., Elser, J.J. & Kumar, S. (2009a) Ecological nitrogen limitation shapes the DNA composition of plant genomes. *Molecular Biology and Evolution*, **26**, 953–956.

Acquisti, C., Kumar, S. & Elser, J.J. (2009b) Signatures of nitrogen limitation in the elemental composition of the proteins involved in the metabolic apparatus. *Proceedings of the Royal Society B: Biological Sciences*, **276**, 2605–2610.

Anneberg, T.J. & Segraves, K.A. (2020) Nutrient enrichment and neopolyploidy interact to increase lifetime fitness of *Arabidopsis thaliana*. *Plant and Soil*, 1–15.

Assis, R. & Kondrashov, A.S. (2012) A strong deletion bias in nonallelic gene conversion. *PLoS Genetics*, **8**, e1002508.

Bales, A.L. & Hersch-Green, E.I. (2019) Effects of soil nitrogen on diploid advantage in fireweed, *Chamerion angustifolium* (Onagraceae). *Ecology and Evolution*, **9**, 1095–1109.

Beaulieu, J.M., Leitch, I.J., Patel, S., Pendharker, A. & Knight, C.A. (2008) Genome size is a strong predictor of cell size and stomatal density in angiosperms. *New Phytologist*, **179**, 975–986.

Bennett, M.D., Leitch, I.J., Price, H.J. & Johnston, J.S. (2003) Comparisons with *Caenorhabditis* (approximately 100 Mb) and *Drosophila* (approximately 175 Mb) using flow cytometry show genome size in *Arabidopsis* to be approximately 157 Mb and thus approximately 25% larger than the *Arabidopsis* genome initiative estimate of approximately 125 Mb. *Annals of Botany*, **91**, 547–557.

Bennetzen, J.L., Ma, J. & Devos, K.M. (2005) Mechanisms of recent genome size variation in flowering plants. *Annals of Botany*, **95**, 127–132.

Bowers, J.E., Chapman, B.A., Rong, J. & Paterson, A.H. (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature*, **422**, 433–438.

Boyko, A., Hudson, D., Bhomkar, P., Kathiria, P. & Kovalchuk, I. (2006a) Increase of homologous recombination frequency in vascular tissue of *Arabidopsis* plants exposed to salt stress. *Plant and Cell Physiology*, **47**, 736–742.

Boyko, A., Zemp, F., Filkowski, J. & Kovalchuk, I. (2006b) Double-strand break repair in plants is developmentally regulated. *Plant Physiology*, **141**, 488–497.

Brandtsma, I. & Gent, D.C. (2012) Pathway choice in DNA double strand break repair: observations of a balancing act. *Genome Integrity*, **3**, 9.

Carta, A., Bedini, G. & Peruzzi, L. (2020) A deep dive into the ancestral chromosome number and genome size of flowering plants. *New Phytologist*, **228**(3), 1097–1106.

Cavalier-Smith, T. (2005) Economy, speed and size matter: evolutionary forces driving nuclear genome miniaturization and expansion. *Annals of Botany*, **95**, 147–175.

Ceccaldi, R., Rondinelli, B. & D'Andrea, A.D. (2016) Repair pathway choices and consequences at the double-strand break. *Trends in Cell Biology*, **26**, 52–64.

Chester, M., Gallagher, J.P., Symonds, V.V., Cruz da Silva, A.V., Mavrodiev, E.V., Leitch, A.R. *et al.* (2012) Extensive chromosomal variation in a recently formed natural allopolyploid species, *Tragopogon miscellus* (Asteraceae). *Proceedings of the National Academy of Sciences of United States of America*, **109**, 1176–1181.

Chiruvella, K.K., Liang, Z. & Wilson, T.E. (2013) Repair of double-strand breaks by end joining. *Cold Spring Harbor Perspectives in Biology*, **5**, a012757.

Clark, J.W. & Donoghue, P.C.J. (2018) Whole-genome duplication and plant macroevolution. *Trends in Plant Science*, **23**, 933–945.

Clark, J., Hidalgo, O., Pellicer, J., Liu, H., Marquardt, J., Robert, Y. *et al.* (2016) Genome evolution of ferns: evidence for relative stasis of genome size across the fern phylogeny. *New Phytologist*, **210**, 1072–1082.

Clouaire, T. & Legube, G. (2019) A snapshot on the cis chromatin response to DNA double-strand breaks. *Trends in Genetics*, **35**, 330–345.

Comai, L. (2000) Genetic and epigenetic interactions in allopolyploid plants. *Plant Molecular Biology*, **43**, 387–399.

Cossu, R.M., Casola, C., Giacomello, S., Vidalis, A., Scofield, D.G. & Zuccolo, A. (2017) LTR retrotransposons show low levels of unequal recombination and high rates of intraelement gene conversion in large plant genomes. *Genome Biology and Evolution*, **9**, 3449–3462.

Deng, W., Henriet, S. & Chourrout, D. (2018) Prevalence of mutation-prone microhomology-mediated end joining in a chordate lacking cNHEJ DNA repair. *Current Biology*, **28**, 3337–3341.

Dias, E.S., Hatt, C., Hamon, S., Hamon, P., Rigoreau, M., Crouzillat, D. *et al.* (2015) Large distribution and high sequence identity of a Copia-type retrotransposon in angiosperm families. *Plant Molecular Biology*, **89**, 83–97.

Dodsworth, S., Chase, M.W. & Leitch, A.R. (2015a) Is post-polyploidization diploidization the key to the evolutionary success of angiosperms? *Botanical Journal of the Linnean Society*, **180**, 1–5.

Dodsworth, S., Leitch, A.R. & Leitch, I.J. (2015b) Genome size diversity in angiosperms and its influence on gene space. *Current Opinion in Genetics Development*, **35**, 73–78.

Doyle, J.J. & Coate, J.E. (2019) Polyploidy, the nucleotype, and novelty: the impact of genome doubling on the biology of the cell. *International Journal of Plant Sciences*, **180**, 1–52.

- Escudero, M. & Wendel, J.F. (2020) The grand sweep of chromosomal evolution in angiosperms. *New Phytologist*, **228**(3), 805–808.
- Farhat, P., Hidalgo, O., Robert, T., Siljak-Yakovlev, S., Leitch, I.J., Adams, R.P. et al. (2019) Polyploidy in the conifer genus *Juniperus*: an unexpectedly high rate. *Frontiers in Plant Science*, **10**, 676.
- Fox, D.T., Soltis, D.E., Soltis, P.S., Ashman, T.L. & Van de Peer, Y. (2020) Polyploidy: a biological force from cells to ecosystems. *Trends in Cell Biology*, **30**, 688–694.
- Franks, P.J. & Beerling, D.J. (2009) Maximum leaf conductance driven by CO<sub>2</sub> effects on stomatal size and density over geologic time. *Proceedings of the National Academy of Sciences of United States of America*, **106**, 10343–10347.
- Franks, P.J. & Farquhar, G.D. (2001) The effect of exogenous abscisic acid on stomatal development, stomatal mechanics, and leaf gas exchange in *Tradescantia virginiana*. *Plant Physiology*, **125**, 935–942.
- Gao, M., Wei, W., Li, M.M., Wu, Y.S., Ba, Z., Jin, K.X. et al. (2014) Ago2 facilitates Rad51 recruitment and DNA double-strand break repair by homologous recombination. *Cell Research*, **24**, 532–541.
- Gregory, T.R. (2001) Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. *Biological Reviews*, **76**, 65–101.
- Greilhuber, J. & Leitch, I.J. (2013) Genome size and the phenotype. In: Greilhuber, J., Dolezel, J. & Wendel, J.F. (Eds.) *Plant genome diversity volume 2: physical structure, behaviour and evolution of plant genomes*. Vienna: Springer Vienna, pp. 323–344.
- Guignard, M.S., Nichols, R.A., Knell, R.J., Macdonald, A., Romila, C.A., Trimmer, M. et al. (2016) Genome size and ploidy influence angiosperm species' biomass under nitrogen and phosphorus limitation. *New Phytologist*, **210**, 1195–1206.
- Harris, J.L., Rabellino, A. & Khanna, K.K. (2018) RAD51 paralogs promote genomic integrity and chemoresistance in cancer by facilitating homologous recombination. *Annals of Translational Medicine*, **6**, S122.
- Hawkins, J.S., Proulx, S.R., Rapp, R.A. & Wendel, J.F. (2009) Rapid DNA loss as a counterbalance to genome expansion through retrotransposon proliferation in plants. *Proceedings of the National Academy of Sciences of United States of America*, **106**, 17811–17816.
- Herben, T., Suda, J., Klimešová, J., Mihulka, S., Říha, P. & Šimová, I. (2012) Ecological effects of cell-level processes: genome size, functional traits and regional abundance of herbaceous plant species. *Annals of Botany*, **110**, 1257–1267.
- Hessen, D.O., Jayasingh, P.D., Neiman, M. & Weider, L.J. (2010) Genome streamlining and the elemental costs of growth. *Trends in Ecology & Evolution*, **25**, 75–80.
- Hilton, J. & Bateman, R.M. (2006) Pteridosperms are the backbone of seed-plant phylogeny. *Journal of the Torrey Botanical Society*, **133**, 119–168.
- Hoffmann, A.A. & Rieseberg, L.H. (2008) Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation? *Annual Review of Ecology Evolution and Systematics*, **39**, 21–42.
- Holloman, W.K. (2011) Unraveling the mechanism of BRCA2 in homologous recombination. *Nature Structural & Molecular Biology*, **18**, 748–754.
- Ickert-Bond, S.M., Sousa, A., Min, Y., Loera, I., Metzgar, J., Pellicer, J. et al. (2020) Polyploidy in gymnosperms – Insights into the genomic and evolutionary consequences of polyploidy in *Ephedra*. *Molecular Phylogenetics and Evolution*, **147**, 106786.
- Jailon, O., Aury, J.M., Noel, B., Policriti, A., Clepet, C., Casagrande, A. et al. (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, **449**, 463–467.
- Jiao, Y., Wickert, N.J., Ayyampalayam, S., Chanderbali, A.S., Landherr, L., Ralph, P.E. et al. (2011) Ancestral polyploidy in seed plants and angiosperms. *Nature*, **473**, 97–100.
- Jovtchev, G., Schubert, V., Meister, A., Barow, M. & Schubert, I. (2006) Nuclear DNA content and nuclear and cell volume are positively correlated in angiosperms. *Cytogenetic and Genome Research*, **114**, 77–82.
- Kang, M., Wang, J. & Huang, H. (2015) Nitrogen limitation as a driver of genome size evolution in a group of karst plants. *Scientific Reports*, **5**, 11636.
- Kelly, S. (2018) The amount of nitrogen used for photosynthesis modulates molecular evolution in plants. *Molecular Biology and Evolution*, **35**, 1616–1625.
- Kenchanmane Raju, S.K., Ritter, E.J. & Niederhuth, C.E. (2019) Establishment, maintenance, and biological roles of non-CG methylation in plants. *Essays in Biochemistry*, **63**, 743–755.
- Kolář, F., Certner, M., Suda, J., Schonswetter, P. & Husband, B.C. (2017) Mixed-ploidy species: progress and opportunities in polyploid research. *Trends in Plant Science*, **22**, 1041–1055.
- Lafarge, S. & Montane, M.H. (2003) Characterization of *Arabidopsis thaliana* ortholog of the human breast cancer susceptibility gene 1: AtBRCA1, strongly induced by gamma rays. *Nucleic Acids Research*, **31**, 1148–1155.
- Landis, J.B., Soltis, D.E., Li, Z., Marx, H.E., Barker, M.S., Tank, D.C. et al. (2018) Impact of whole-genome duplication events on diversification rates in angiosperms. *American Journal of Botany*, **105**, 348–363.
- Lawson, T. & Violet-Chabrand, S. (2019) Speedy stomata, photosynthesis and plant water use efficiency. *New Phytologist*, **221**, 93–98.
- Le Comber, S.C., Ainouche, M.L., Kovarik, A. & Leitch, A.R. (2010) Making a functional diploid: from polysomic to disomic inheritance. *New Phytologist*, **186**, 113–122.
- Leitch, I.J. & Bennett, M.D. (2004) Genome downsizing in polyploid plants. *Biological Journal of the Linnean Society*, **82**, 651–663.
- Li, Z., Baniaga, A.E., Sessa, E.B., Scascitelli, M., Graham, S.W., Rieseberg, L.H. et al. (2015) Early genome duplications in conifers and other seed plants. *Science Advances*, **1**, e1501084.
- Lisch, D. (2013) How important are transposons for plant evolution? *Nature Reviews Genetics*, **14**, 49–61.
- Lloyd, A. & Bomblies, K. (2016) Meiosis in autopolyploid and allopolyploid *Arabidopsis*. *Current Opinion in Plant Biology*, **30**, 116–122.
- Lynch, M. & Conery, J.S. (2003) The origins of genome complexity. *Science*, **302**, 1401–1404.
- Lynch, M. & Marinov, G.K. (2015) The bioenergetic costs of a gene. *Proceedings of the National Academy of Sciences of United States of America*, **112**, 15690–15695.
- Mandáková, T. & Lysak, M.A. (2018) Post-polyploid diploidization and diversification through dysploid changes. *Current Opinion in Plant Biology*, **42**, 55–65.
- Mari, P.-O., Florea, B.I., Persengiev, S.P., Verkaik, N.S., Brüggewirth, H.T., Modesti, M. et al. (2006) Dynamic assembly of end-joining complexes requires interaction between Ku70/80 and XRCC4. *Proceedings of the National Academy of Sciences of United States of America*, **103**, 18597.
- McVey, M. & Lee, S.E. (2008) MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends in Genetics*, **24**, 529–538.
- Nieto Feliner, G., Casacuberta, J. & Wendel, J.F. (2020) Genomics of evolutionary novelty in hybrids and polyploids. *Frontiers in Genetics*, **11**, 792.
- Novák, P., Guignard, M.S., Neumann, P., Kelly, L.J., Mlinarec, J., Koblížková, A. et al. (2020) Repeat-sequence turnover shifts fundamentally in species with large genomes. *Nature Plants*, **6**, 1325–1329.
- Orel, N., Kyryk, A. & Puchta, H. (2003) Different pathways of homologous recombination are used for the repair of double-strand breaks within tandemly arranged sequences in the plant genome. *The Plant Journal*, **35**, 604–612.
- Orel, N. & Puchta, H. (2003) Differences in the processing of DNA ends in *Arabidopsis thaliana* and tobacco: possible implications for genome evolution. *Plant Molecular Biology*, **51**, 523–531. Available from: <https://link.springer.com/article/10.1023/A:1022342405661>
- Ottaviani, D., LeCain, M. & Sheer, D. (2014) The role of microhomology in genomic structural variation. *Trends in Genetics*, **30**, 85–94.
- Panchy, N., Lehti-Shiu, M. & Shiu, S.H. (2016) Evolution of gene duplication in plants. *Plant Physiology*, **171**, 2294–2316.
- Pellicer, J., Hidalgo, O., Dodsworth, S. & Leitch, I.J. (2018) Genome size diversity and its impact on the evolution of land plants. *Genes (Basel)*, **9**, 88.
- Pellicer, J. & Leitch, I.J. (2020) The Plant DNA C-values database (release 7.1): an updated online repository of plant genome size data for comparative studies. *New Phytologist*, **226**, 301–305.
- Petit, M., Guidat, C., Daniel, J., Denis, E., Montoriol, E., Bui, Q.T. et al. (2010) Mobilization of retrotransposons in synthetic allotetraploid tobacco. *New Phytologist*, **186**, 135–147.
- Petrov, D.A., Sangster, T.A., Johnston, J.S., Hartl, D.L. & Shaw, K.L. (2000) Evidence for DNA loss as a determinant of genome size. *Science*, **287**, 1060–1062.
- Prakash, R., Zhang, Y., Feng, W. & Jasin, M. (2015) Homologous recombination and human health: the roles of BRCA1, BRCA2, and associated proteins. *Cold Spring Harbor Perspectives in Biology*, **7**, a016600.
- Puttick, M.N., Clark, J. & Donoghue, P.C.J. (2015) Size is not everything: rates of genome size evolution, not C-value, correlate with speciation in

- angiosperms. *Proceedings of the Royal Society B: Biological Sciences*, **282**(1820), 20152289.
- Raina, S.N., Parida, A., Koul, K.K., Salimath, S.S., Bisht, M.S., Raja, V. *et al.* (1994) Associated chromosomal DNA changes in polyploids. *Genome*, **37**, 560–564.
- Ranjha, L., Howard, S.M. & Cejka, P. (2018) Main steps in DNA double-strand break repair: an introduction to homologous recombination and related processes. *Chromosoma*, **127**, 187–214.
- Reef, R., Ball, M.C., Feller, I.C. & Lovelock, C.E. (2010) Relationships among RNA: DNA ratio, growth and elemental stoichiometry in mangrove trees. *Functional Ecology*, **24**, 1064–1072.
- Ren, L., Huang, W., Cannon, E.K.S., Bertoli, D.J. & Cannon, S.B. (2018) A mechanism for genome size reduction following genomic rearrangements. *Frontiers in Genetics*, **9**, 454.
- Revell, L.J. (2012) phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, **3**, 217–223.
- Ritz, K.R., Noor, M.A.F. & Singh, N.D. (2017) Variation in recombination rate: adaptive or not? *Trends in Genetics*, **33**, 364–374.
- Roddy, A.B., Thérroux-Rancourt, G., Abbo, T., Benedetti, J.W., Brodersen, C.R., Castro, M. *et al.* (2019) The scaling of genome size and cell size limits maximum rates of photosynthesis with implications for ecological strategies. *International Journal of Plant Sciences*, **181**, 75–87.
- Roddy, A.B., Thérroux-Rancourt, G., Abbo, T., Benedetti, J.W., Brodersen, C.R., Castro, M. *et al.* (2020) The scaling of genome size and cell size limits maximum rates of photosynthesis with implications for ecological strategies. *International Journal of Plant Sciences*, **181**, 75–87.
- Scannell, D.R., Frank, A.C., Conant, G.C., Byrne, K.P., Woolfit, M. & Wolfe, K.H. (2007) Independent sorting-out of thousands of duplicated gene pairs in two yeast species descended from a whole-genome duplication. *Proceedings of the National Academy of Sciences of United States of America*, **104**, 8397.
- Schubert, I. & Vu, G.T.H. (2016) Genome stability and evolution: attempting a holistic view. *Trends in Plant Science*, **21**, 749–757.
- Senerchia, N., Felber, F. & Parisod, C. (2014) Contrasting evolutionary trajectories of multiple retrotransposons following independent allopolyploidy in wild wheats. *New Phytologist*, **202**, 975–985.
- Shenhav, L. & Zeevi, D. (2020) Resource conservation manifests in the genetic code. *Science*, **370**, 683–687.
- Shibata, A., Conrad, S., Birraux, J., Geuting, V., Barton, O., Ismail, A. *et al.* (2011) Factors determining DNA double-strand break repair pathway choice in G2 phase. *The EMBO Journal*, **30**, 1079–1092.
- Shrivastav, M., De Haro, L.P. & Nickoloff, J.A. (2008) Regulation of DNA double-strand break repair pathway choice. *Cell Research*, **18**, 134–147.
- Siaud, N., Dray, E., Gy, I., Gérard, E., Takvorian, N. & Doutriaux, M.-P. (2004) Brca2 is involved in meiosis in *Arabidopsis thaliana* as suggested by its interaction with Dmc1. *The EMBO Journal*, **23**, 1392–1401.
- Simonin, K.A. & Roddy, A.B. (2018) Genome downsizing, physiological novelty, and the global dominance of flowering plants. *PLoS Biology*, **16**, e2003706.
- Šimová, I. & Herben, T. (2012) Geometrical constraints in the scaling relationships between genome size, cell size and cell cycle length in herbaceous plants. *Proceedings of the Royal Society B: Biological Sciences*, **279**, 867–875.
- Šmarda, P., Hejčman, M., Březinová, A., Horová, L., Steigerová, H., Zedek, F. *et al.* (2013) Effect of phosphorus availability on the selection of species with different ploidy levels and genome sizes in a long-term grassland fertilization experiment. *New Phytologist*, **200**, 911–921.
- Smith, S.A. & Brown, J.W. (2018) Constructing a broadly inclusive seed plant phylogeny. *American Journal of Botany*, **105**, 302–314.
- Soltis, D.E., Albert, V.A., Leebens-Mack, J., Bell, C.D., Paterson, A.H., Zheng, C. *et al.* (2009) Polyploidy and angiosperm diversification. *American Journal of Botany*, **96**, 336–348.
- Stapley, J., Feulner, P.G.D., Johnston, S.E., Santure, A.W. & Smadja, C.M. (2017) Variation in recombination frequency and distribution across eukaryotes: patterns and processes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **372**(1736), 20160455.
- Sterner, R.W. & Elser, J.J. (2002) *Ecological stoichiometry: the biology of elements from molecules to the biosphere*. Princeton, NJ: Princeton University Press.
- Sugiyama, T., Zaitseva, E.M. & Kowalczykowski, S.C. (1997) A single-stranded DNA-binding protein is needed for efficient presynaptic complex formation by the *Saccharomyces cerevisiae* Rad51 protein. *Journal of Biological Chemistry*, **272**, 7940–7945.
- Sulpice, R., Ishihara, H., Schlereth, A., Cawthray, G.R., Encke, B., Giavalisco, P. *et al.* (2014) Low levels of ribosomal RNA partly account for the very high photosynthetic phosphorus-use efficiency of Proteaceae species. *Plant, Cell & Environment*, **37**, 1276–1298.
- Uziel, T., Lerenthal, Y., Moyal, L., Andegeko, Y., Mittelman, L. & Shiloh, Y. (2003) Requirement of the MRN complex for ATM activation by DNA damage. *The EMBO Journal*, **22**, 5612–5621.
- Van de Peer, Y., Mizrahi, E. & Marchal, K. (2017) The evolutionary significance of polyploidy. *Nature Reviews Genetics*, **18**, 411–424.
- Veleba, A., Zedek, F., Horová, L., Veselý, P., Srba, M., Šmarda, P. *et al.* (2020) Is the evolution of carnivory connected with genome size reduction? *American Journal of Botany*, **107**, 1253–1259.
- Veselý, P., Šmarda, P., Bureš, P., Stirton, C., Muasya, A.M., Mucina, L. *et al.* (2020) Environmental pressures on stomatal size may drive plant genome size evolution: evidence from a natural experiment with Cape geophytes. *Annals of Botany*, **126**(2), 323–330.
- Vision, T.J., Brown, D.G. & Tanksley, S.D. (2000) The origins of genomic duplications in *Arabidopsis*. *Science*, **290**, 2114–2117.
- Vitales, D., Álvarez, I., García, S., Hidalgo, O., Nieto Feliner, G., Pellicer, J. *et al.* (2019) Genome size variation at constant chromosome number is not correlated with repetitive DNA dynamism in *Anacyclus* (Asteraceae). *Annals of Botany*, **125**, 611–623.
- Vu, G.T.H., Cao, H.X., Reiss, B. & Schubert, I. (2017) Deletion-bias in DNA double-strand break repair differentially contributes to plant genome shrinkage. *The New Phytologist*, **214**, 1712–1721.
- Vu, G.T., Cao, H.X., Watanabe, K., Hensel, G., Blattner, F.R., Kumlehn, J. *et al.* (2014) Repair of site-specific DNA double-strand breaks in barley occurs via diverse pathways primarily involving the sister chromatid. *The Plant Cell*, **26**, 2156–2167.
- Walczyk, A.M. & Hersch-Green, E.I. (2019) Impacts of soil nitrogen and phosphorus levels on cytotype performance of the circumboreal herb *Chamerion angustifolium*: implications for polyploid establishment. *American Journal of Botany*, **106**, 906–921.
- Walsh, C.S. (2015) Two decades beyond BRCA1/2: Homologous recombination, hereditary cancer risk and a target for ovarian cancer therapy. *Gynecologic Oncology*, **137**, 343–350.
- Wendel, J.F. (2015) The wondrous cycles of polyploidy in plants. *American Journal of Botany*, **102**, 1753–1756.
- Wilson, M.J., Fradera-Soler, M., Summers, R., Sturrock, C.J. & Fleming, A.J. (2021) Ploidy influences wheat mesophyll cell geometry, packing and leaf function. *Plant Direct*, **5**, e00314.
- Wu, H., Yu, Q., Ran, J.-H. & Wang, X.-Q. (2020) Unbiased subgenome evolution in allotetraploid species of *Ephedra* and its implications for the evolution of large genomes in gymnosperms. *Genome Biology and Evolution*, **13**, evaa236.
- Yant, L., Hollister, J.D., Wright, K.M., Arnold, B.J. & Higgins, J.D. (2013) Meiotic adaptation to genome duplication in *Arabidopsis arenosa*. *Current Biology*, **23**, 2151–2156.
- Zamariola, L., Tiang, C.L., De Storme, N., Pawlowski, W. & Geelen, D. (2014) Chromosome segregation in plant meiosis. *Frontiers in Plant Science*, **5**, 279.
- Zenil-Ferguson, R., Ponciano, J.M. & Burleigh, J.G. (2016) Evaluating the role of genome downsizing and size thresholds from genome size distributions in angiosperms. *American Journal of Botany*, **103**, 1175–1186.
- Zwaenepoel, A. & Van de Peer, Y. (2019) Inference of ancient whole-genome duplications and the evolution of gene duplication and loss rates. *Molecular Biology and Evolution*, **36**, 1384–1404.