RESEARCH ARTICLE

AMERICAN JOURNAL OF BOTANY

INVITED PAPER

For the Special Issue: The Evolutionary Importance of Polyploidy

Evaluating the role of genome downsizing and size thresholds from genome size distributions in angiosperms[1]

Rosana Zenil-Ferguson[2], José M. Ponciano, and J. Gordon Burleigh

PREMISE OF THE STUDY: Whole-genome duplications (WGDs) can rapidly increase genome size in angiosperms. Yet their mean genome size is not correlated with ploidy. We compared three hypotheses to explain the constancy of genome size means across ploidies. The genome downsizing hypothesis suggests that genome size will decrease by a given percentage after a WGD. The genome size threshold hypothesis assumes that taxa with large genomes or large monoploid numbers will fail to undergo or survive WGDs. Finally, the genome downsizing and threshold hypothesis suggests that both genome downsizing and thresholds affect the relationship between genome size means and ploidy.

METHODS: We performed nonparametric bootstrap simulations to compare observed angiosperm genome size means among species or genera against simulated genome sizes under the three different hypotheses. We evaluated the hypotheses using a decision theory approach and estimated the expected percentage of genome downsizing.

KEY RESULTS: The threshold hypothesis improves the approximations between mean genome size and simulated genome size. At the species level, the genome downsizing with thresholds hypothesis best explains the genome size means with a 15% genome downsizing percentage. In the genus level simulations, the monoploid number threshold hypothesis best explains the data.

CONCLUSIONS: Thresholds of genome size and monoploid number added to genome downsizing at species level simulations explain the observed means of angiosperm genome sizes, and monoploid number is important for determining the genome size mean at the genus level.

KEY WORDS genome downsizing; large genome thresholds; monoploid number; polyploidy; whole-genome duplications

Whole-genome duplications (WGDs) have occurred frequently throughout the evolution of angiosperms (e.g., Leitch and Bennett, 1997; Soltis and Soltis, 1999; Adams and Wendel, 2005; Jiao et al., 2011), and consequently, there has been much interest in the effects of WGD on genome size and composition in plants (Bennetzen and Kellogg, 1997; Bennetzen, 2002, Otto, 2007; Knight and Beaulieu, 2008; Mayrose et al., 2010; Soltis et al., 2012; Arrigo and Barker, 2012). Although WGDs initially increase both the genome size and number of chromosomes, changes in the genome size mean across angiosperms are not positively correlated with the ploidy level (Leitch and Bennett, 2004; Michael, 2014). In fact, the genome size means remain relatively constant at all ploidies (Leitch and Bennett, 2004). This constancy in the face of frequent WGDs is a central mystery in plant genome evolution.

Whole-genome duplications in plants are usually followed by rapid genome downsizing, or fractionation (Soltis and Soltis, 1999; Petrov et al., 2000; Bennetzen, 2002; De Smet et al., 2013), which could reduce the correlation between genome size mean and ploidy level (Leitch and Bennett, 2004). Furthermore, the lack of correlation could be due to the accumulation of transposable elements (Sanmiguel and Bennetzen, 1998; Hawkins et al., 2006) or the removal of large blocks of DNA via intra-strand homologous recombination and illegitimate recombination events (Vitte and Bennetzen, 2006; Hawkins et al., 2008). Genome size changes also have been associated with recombination rates (Ross-Ibarra, 2007; Tiley and Burleigh, 2015). Thus, genome size encompasses the net difference between increases and decreases of functional and nonfunctional DNA.

Biases in the species that undergo or survive WGDs could also affect the relationship between ploidy and mean genome size. For

[1] Manuscript received 17 September 2016; revision accepted 7 January 2016.
Department of Biology, University of Florida, P. O. Box 118525, Gainesville, Florida 32611-8525 USA
[2] Author of correspondence (e-mail: rzenil@ufl.edu)
doi:10.3732/ajb.1500408

example, WGD may be difficult or impossible in species with large genomes (Knight et al., 2005) if large genome sizes increase nutrient demands (Šmarda et al., 2013) or if large genomes are associated with cellular traits that cannot scale with the duplication (Knight and Beaulieu, 2008). Furthermore, for taxa with many chromosomes, WGDs could also be detrimental if the total number of chromosomes impedes homologous chromosome recognition during meiosis following a WGD (Zielinski and Scheid, 2012; Moore, 2013).

In this study, we used data for genome size and monoploid number (defined as chromosome number divided by ploidy) across angiosperms to evaluate the role of genome downsizing and thresholds on the genome size means at different ploidies. We performed simulations at species and genus levels to assess the ability of three hypotheses to explain observed constancy of genome size mean across ploidies (Leitch and Bennett, 2004). First, the genome downsizing (GD) hypothesis assumes that after a WGD, genome sizes decrease by a constant amount, explaining the distribution of mean genome sizes at different ploidy levels. Next, the threshold (T) hypothesis assumes that WGD is successful only for taxa with small genomes (GST) and/or small monoploid numbers (MNT), and this size limit determines the distribution and mean of genome sizes at each ploidy level. Finally, we examined the genome downsizing and threshold (GD+T) hypothesis, which assumes both genome downsizing and the presence of a maximum genome size and/or monoploid number threshold explain the genome size means across ploidies. We compared the results of our simulations to the observed mean of the distribution of genome sizes at each ploidy obtained from the Kew Royal Botanic Gardens Angiosperm C-value data set (Bennett and Leitch, 1995, 2012) to determine which scenario best explains the angiosperm genome size means across ploidy levels.

MATERIALS AND METHODS

Data set—We extracted data from 7187 angiosperm taxa from the Kew Royal Botanical Gardens Angiosperm DNA C-value database (http://data.kew.org/cvalues; Bennett and Leitch, 1995, 2012) in December 2013. The data set, which includes information from scientific publications from 1976 to 2012, contains taxon name, chromosome number (*2n*), ploidy level (*j*), and genome size (1C) data. The angiosperm data set contains 248 families and 1630 genera. Summary statistics for genome size at family and genus levels can be found in Appendix S1 (see supplemental data with online version of this article).

Polyploidy series data and genome downsizing percentage—Of the 7187 taxa in the Angiosperm C-value database, 304 taxa had data from multiple ploidy levels (Appendix S1). We refer to these as the polyploidy series taxa. We identified the maximum genome size $y^{j}_{\text{polyploid}}$ and maximum monoploid number $x^{j}_{\text{polyploid}}$ (defined as chromosome number divided by ploidy) at each ploidy $j = 2x, 3x, \ldots, 20x$ for all polyploidy series taxa (Fig. 1). To determine which ploidies are most frequently associated with each other, we assembled a ploidy matrix based on the number of subsets of two ploidies that appear in the same polyploidy series taxa. For example, *Artemisia dracunculus* has ploidy values 2*x*, 4*x*, 6*x*, and 10*x*; thus, there are six subsets of two ploidies that can potentially be associated in this species (i.e., 2*x* and 4*x*, 2*x* and 6*x*, 2*x* and 10*x*, 4*x* and 6*x*, 4*x* and 10*x*, and 6*x* and 10*x*). We estimated the relative frequencies of the associated ploidies by dividing each value of the ploidy series matrix by the sum of the total associations counted in their respective row (Table 1).

To investigate the possible genome downsizing percentage experienced by polyploidy series taxa, we first calculated expected

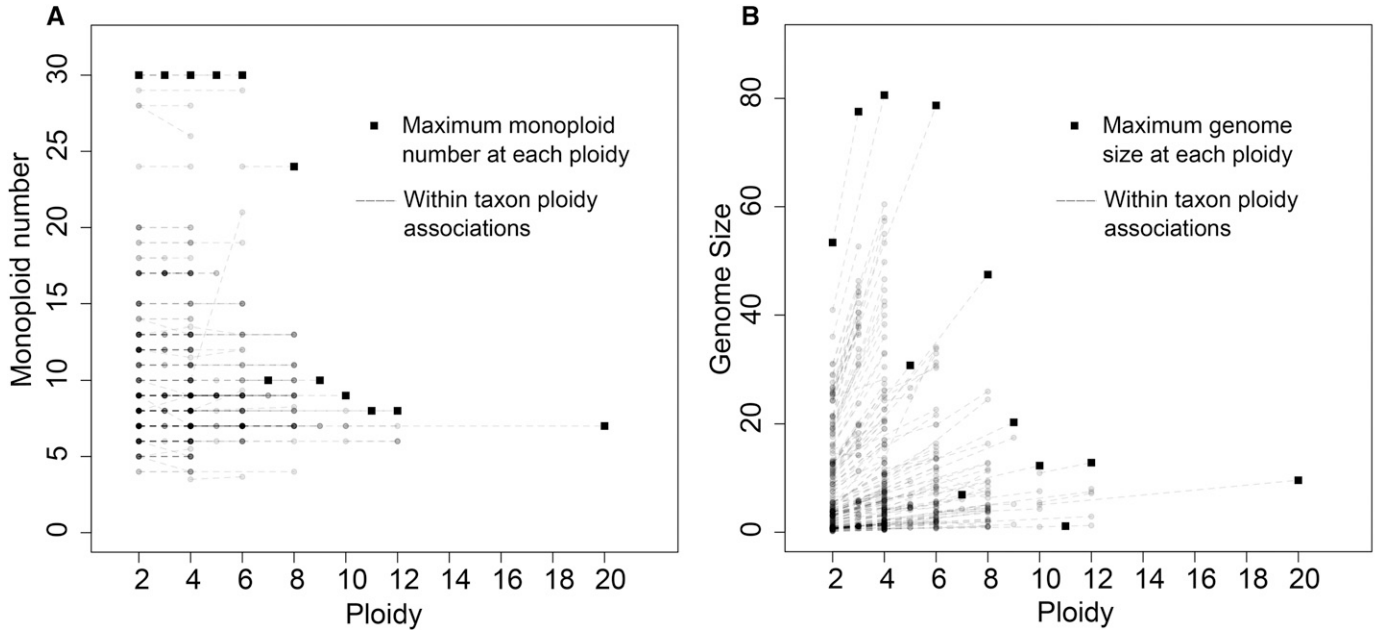

FIGURE 1 (A) Monoploid number for polyploid series taxa. Each dotted line represents the changes in ploidy for each of the 304 taxa. More than 90% of ploidy changes do not change the monoploid number (horizontal dotted lines). (B) Genome size for polyploid series taxa. Each dotted line represents the changes in ploidy for each of the 304 taxa. More than 92% of the ploidy changes do not change genome size relative to the change in ploidy.

genome size after a ploidy change as $y^{j}_{\text{expected}} = \frac{j}{i} y^{i}_{\text{observed}}$, where $i < j$ are two different ploidy levels for a single taxon. In the example of *Artemisia dracunculus*, the diploid genome size is 2.97 pg, so the expected tetraploid genome size from autopolyploidy would be 5.94 pg; however, the observed tetraploid genome size is 5.91 pg. Later, we matched the data set of observed and expected genome sizes to a recent angiosperm phylogeny at the species level (Zanne et al., 2014), which contained 121 of the polyploidy series taxa (Appendix S1). We performed phylogenetic generalized least squares (PGLS) using the linear model

$$y^{j}_{\text{observed}} = (1-r)\left(y^{j}_{\text{expected}}\right) + \varepsilon \qquad \text{(Eq. 1)}$$

with three correlation structures: Brownian motion (BM) with parameter σ representing drift; Ornstein–Uhlenbeck (OU) with parameter α representing strength of selection and/or phylogenetic signal; and Pagel transformation with parameter λ representing phylogenetic signal (Pagel, 1997; Martins and Hansen, 1997; Freckleton et al., 2002). Parameter $(1 - r)$ represents the retention of genome size after the genome size increase; thus $r \times 100\%$ is the genome downsizing percentage. We obtained maximum likelihood estimates for r, σ^2, α, and λ (Fig. 2) and the Akaike information criterion (AIC) for the linear model in Eq. 1 under each of the three phylogenetic correlation structures to determine the best model.

Simulation experiment—Proposing a linear model to estimate the genome downsizing percentage requires the comparison of the expected and observed genome sizes within the same taxon. Because ancestral genome sizes and history of whole-genome duplications are not available for all angiosperm taxa in the angiosperm C-value data set, we designed 16 nonparametric bootstrap simulation experiments that approximated mean genome size at all ploidies under the presence of genome downsizing and thresholds. Therefore, simulations are an approximation to the linear model shown in Eq. 1, in the absence of genome duplication information. None of our simulations consider phylogenetic relatedness since we found strong evidence for phylogenetic independence from the linear model that uses polyploidy series taxa (see results and Fig. 2). Nine simulations were performed at the species level (by sampling from species values) and the remaining seven at the genus level (by sampling from a single species per genus). We performed simulations at the genus level to remove possible biases toward highly sampled or studied genera. For all the simulation experiments, we only considered even ploidy levels, which contained far more data in the C-value database than odd ploidy levels, like other studies have considered in the past (Leitch and Bennett, 2004; Meyers and Levin, 2006). For simplicity, in our simulations, we assumed that ploidy only increases between two subsequent even ploidy levels; that is, ploidy changes only happen from an even ploidy to the next even ploidy (e.g., $2x$ to $4x$, or $4x$ to $6x$). Preliminary results from a modeling study in angiosperms indicate that increasing to the nearest even ploidy is the most common type of ploidy change (R. Zenil-Ferguson, unpublished data), which is also consistent with the high number of even to even ploidy associations found in polyploidy series taxa (Table 1).

Simulation for the genome downsizing hypothesis—The first simulation experiments examined the genome downsizing hypothesis (GD). If genome downsizing after a WGD is responsible for the constancy of angiosperm genome size mean at different ploidy levels, reducing (downsizing) the genome sizes by a given percentage after the ploidy increases would approximate the observed mean of genome sizes at the increased ploidy. Under this scenario, for the species level simulation, we took a random sample with replacement of the diploid genome sizes. The sample size is denoted as $n^{4x} = 917$, which corresponds to the number of observed tetraploids in the C-values data set (Bennett and Leitch, 2012). We denote the random sample of diploid genome sizes as $(y^{2x}_1, y^{2x}_2, \ldots, y^{2x}_{917})$. This diploid genome size sample was subsequently multiplied by two, representing a WGD. Then, the sample was "downsized" by multiplying each sampled genome size by $(1-r)$. We tried this approach for all genome downsizing percentages r, a parameter with values ranging between 0 and 1 with a precision of 0.001. It then follows that our simulated sample of tetraploid genomes sizes is $y^{4x}_i = 2(1-r)y^{2x}_i$, with $i = 1,2,\ldots,917$.

For subsequent even ploidies (j = 6x, 8x, 10x, 12x), we drew a random sample with replacement of genome sizes $(y^{j-2}_1, y^{j-2}_2, \ldots, y^{j-2}_{n^j})$, of size n^j, the number of observed polyploids from the C-value data set at ploidy j. Thus, each new simulated sample $(y^{j}_1, y^{j}_2, \ldots, y^{j}_{n^j})$, at ploidy level j was created by increasing ploidy by a factor $\frac{j}{j-2}$ and subsequently applying the genome downsizing percentage loss $1-r$. Hence,

$$y^{j}_i = \frac{j}{j-2}(1-r)\, y^{j-2}_i \text{ for } i = 1, 2, \ldots, n^j. \qquad \text{(Eq. 2).}$$

Note that in Eq. 2, we assumed that ploidy increases additively by two. We repeated this sampling protocol 10,000 times to obtain

TABLE 1. Ploidy associations in polyploidy series taxa.

To From	3*x*	4*x*	5*x*	6*x*	7x	8x	9*x*	10*x*	11*x*	12x	Total associations
2*x*	50 (21.3%)	145 (61.9%)	7 (2.9%)	24 (10.2%)	1 (0.4%)	6 (2.5%)	0 (0.0%)	1 (0.4%)	0 (0.0%)	0 (0.0%)	234
3*x*		20 (76.9%)	5 (19.2%)	1 (3.8%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	26
4*x*			17 (17.7%)	46 (47.9%)	2 (2.0%)	20 (20.8%)	0 (0.0%)	10 (10.4%)	0 (0.0%)	1 (1.0%)	96
5*x*				8 (61.5%)	3 (23.0%)	2 (15.3%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	13
6*x*					4 (11.1%)	21 (58.3%)	4 (11.1%)	4 (11.1%)	1 (2.7%)	2 (5.5%)	6
7*x*						2 (66.6%)	1 (33.3%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	3
8*x*							1 (14.2%)	3 (42.8%)	1 (14.2%)	2 (28.5%)	7
9*x*								0(100.0%)	0 (0.0%)	0 (0.0%)	0
10*x*									1 (25.0%)	3 (75.0%)	4
11*x*										1 (100.0%)	1

Notes: Associations between two ploidy values within each polyploidy series taxon are shown. Associations were counted from lower to higher ploidies observed within a single taxon. Total association percentages are shown in parentheses.

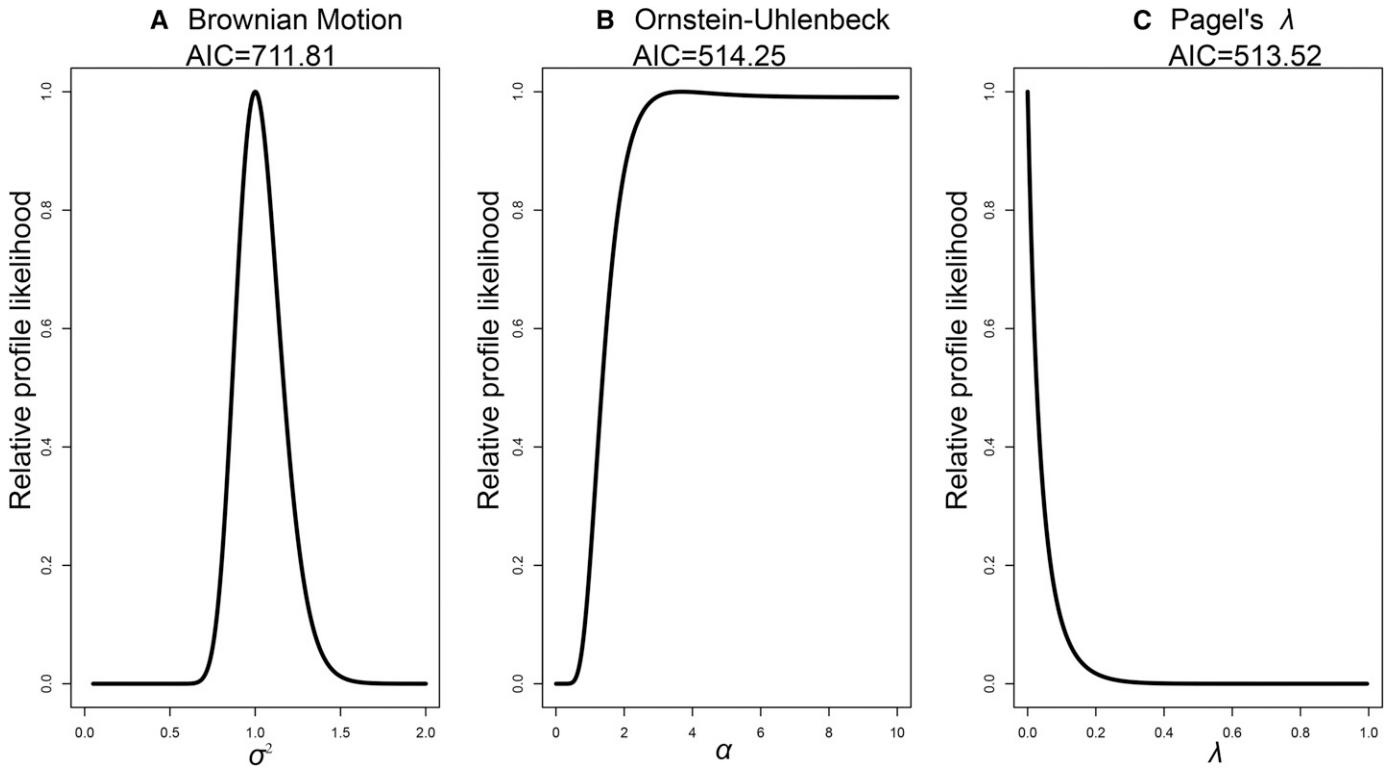

FIGURE 2 Profile likelihood for covariance transformations based on phylogeny for a linear model of genome size change based on polyploidy series taxa, assuming (A) Brownian motion with drift σ^2; (B) Ornstein-Uhlenbeck with pull α; and (C) Pagel's λ transformations. Estimates in (B) and (C) show lack of phylogenetic signal among taxa.

replicates for each genome downsizing percentage $(1-r)$ that we examined and for each even ploidy up to $12x$. We calculated this number of replicates because the simulated genome size mean at each ploidy j defined as

$$\bar{Y}_{GD}^{j} = \frac{1}{n^j}\sum_{i=1}^{n^j} y_i^j \qquad \text{(Eq. 3)}$$

was bootstrap consistent at each ploidy, meaning that, each time the simulation is repeated, the resulting mean value changes only by 10^{-3}.

For the genus level simulations, we randomly sampled one individual for each genus without replacement at ploidy level $j-2$ and then applied Eq. 2 to obtain a new sample for the subsequent even ploidy j. For example, for $j-2=2x$ we took only one diploid taxon from each of the 893 genera and then applied Eq. 2 to this sample to create new tetraploids under the GD hypothesis. For the new sample of tetraploids, we calculated a new genome size mean

$$\bar{Y}_{\mathrm{GD}}^{4x} = \frac{1}{893}\sum_{k=1}^{893} y_k^{2x} \qquad \text{(Eq. 4).}$$

Notice that the average in Eq. 4 is different from the average in Eq. 3. At the species level, the average in Eq. 3 weighs each genus based on their number of taxa, so genera with many species with genome size data contribute more to Eq. 2. In contrast, the genus level mean calculated in Eq. 3 weighs each genus equally (For comparison of averages at the species and genus levels, see Appendix S1 in Supplemental Data with online version of this article).

Simulations for the threshold hypothesis—To test the threshold hypothesis at the species level, we considered thresholds on the monoploid number (MNT) and on genome size (GST). To determine whether constraints on the maximum monoploid number can explain the observed genome size mean at each ploidy level, we first determined the maximum monoploid numbers of the 304 polyploidy series data (Fig. 1A) and set these as the threshold limits. Based on these thresholds, we performed a simulation study to evaluate monoploid number threshold, in which WGD could happen if and only if the monoploid number of a specific taxon is smaller than the proposed monoploid number threshold at each ploidy level. For instance, at ploidy level $2x$, a genome size y_i^{2x} of taxon i was considered for sampling only if its monoploid number x_i^{2x} was less than or equal to 30 ($x_i^{2x} \le x_{\mathrm{polyploid}}^{2x} = 30$), which is the maximum monoploid number observed from polyploidy series taxa at $2x$ ploidy (Fig. 1A). After sampling from taxa that were bounded by the established thresholds, we created the new simulated sample $(y_1^j, y_2^j, \ldots, y_{n^j}^j)$ by increasing the genome size using

$$y_i^j = \frac{j}{j-2} y_i^{j-2}, \text{ with } j = 4x, 6x, \ldots, 12x \qquad \text{(Eq. 5).}$$

The simulated sample sizes at each ploidy were equal to sample sizes observed at each ploidy level n^j. This nonparametric bootstrap

simulation was performed 10,000 times, until calculations of genome size mean $\bar{Y}^{j}_{\mathrm{MNT}}$ by ploidy were consistent. We also repeated this experiment using different monoploid number thresholds, which were based on different percentiles of the distribution of monoploid numbers from all taxa at each ploidy (Fig. 3A). Percentile thresholds used ranged from 70 to 100%, in increments of 5%. Using different percentiles as thresholds allowed us to determine whether preventing more taxa from undergoing a WGD can better explain the genome size mean at the following even ploidy.

Next, we tested whether constraints on the maximum genome size (i.e., C-value) can explain the observed distribution of genome sizes. In each genome size threshold simulation, a genome size y_i^{2x} could potentially be part of the sample if and only if genome size of an individual i is less than or equal to 53.3 pg $\left(y_i^{2x} \leq y^{2x}_{\mathrm{polyploid}} = 53.3\right)$, the maximum genome size among polyploidy series taxa at the $2x$ ploidy (Fig. 1B). Taxa that were bounded by the established thresholds were sampled with replacement, and the new simulated sample $(y_1^j, y_2^j, \ldots, y_{n^j}^j)$ was generated using Eq. 5. This experiment was also repeated using genome size thresholds representing different percentiles of the distribution of genome sizes among all taxa at each ploidy (Fig. 3B). We varied the percentile thresholds from 70 to 100% in increments of 5%. The 70% percentile was the most restrictive threshold hypothesis, meaning that only 70% of the lowest observed genome sizes at a given ploidy could be considered for sampling, whereas the 100% percentile implied that all taxa can be sampled (i.e., there is no threshold). Each nonparametric bootstrap simulation experiment was performed 10,000 times for each polyploidy series and percentile threshold at each ploidy, because calculations of mean genome size $\bar{Y}^{j}_{\mathrm{GST}}$ by ploidy were consistent.

We also performed simulations at the species level considering simultaneously the two thresholds (monoploid number and genome size, MNT + GST), where a taxon could belong to the sample if and only if both its monoploid number was smaller than the polyploidy series monoploid number threshold $(x_i^{j-2} \leq x^{j-2}_{\mathrm{polyploid}})$ and its genome size was smaller than the polyploidy series genome size threshold $(y_i^{j-2} \leq y^{j-2}_{\mathrm{polyploid}})$. We increased the ploidy in the sample through Eq. 5. We ran 10,000 replicates and calculated $\bar{Y}^{j}_{\mathrm{MNT+GST}}$.

For simulations at the genus level, we sampled one individual at each genus per ploidy. Like in the species level simulations, we sampled individuals according to the monoploid number, genome size, and both thresholds, which were determined by polyploidy series taxa (e.g., $x_i^{j-2} \leq x^{j-2}_{\mathrm{polyploid}}, y_i^{j-2} \leq y^{j-2}_{\mathrm{polyploid}}$). We proceeded by increasing ploidy using Eq. 4 and later calculating either $\bar{Y}^{j}_{\mathrm{MNT}}$, $\bar{Y}^{j}_{\mathrm{GST}}$, or $\bar{Y}^{j}_{\mathrm{MNT+GST}}$ based on the threshold selected for each simulation, which all contained 10,000 replicates per ploidy (Table 2).

Simulations for genome downsizing in the presence of a threshold (GD +T)—A third simulation experiment tested the strength of both genome downsizing and the presence of a threshold, either the monoploid number threshold, genome size threshold, or both thresholds (Table 2). For species level simulations, we drew a random sample of genome sizes with replacement $(y_1^{j-2}, y_2^{j-2}, \ldots, y_{n^j}^{j-2})$, of the size n^j, but each genome size had to meet the threshold requirement determined by polyploidy series maxima based on genome sizes $y_i^{j-2} \leq y^{j-2}_{\mathrm{polyploid}}$, monoploid numbers $x_i^{j-2} \leq x^{j-2}_{\mathrm{polyploid}}$, or both (Fig. 1). To obtain the new sample of increased ploidy, we used Eq. 2. We used multiple genome downsizing percentages r with value between 0 and 1 with precision 0.001. After each simulation, the mean $\bar{Y}^{j}_{\mathrm{GD}+T}$ for genome size was calculated from the simulated data. This process was repeated 10,000 times at each even ploidy level $j = 4x, \ldots, 12x$.

At the genus level, we randomly sampled again one taxon per genus and per ploidy, but each taxon was selected only if its genome size was less or equal than the genome size of polyploidy series maximum $(y_i^{j-2} \leq y^{j-2}_{\mathrm{polyploid}})$, its monoploid number was less or equal than the polyploidy series maximum $(x_i^{j-2} \leq x^{j-2}_{\mathrm{polyploid}})$, or both values of genome size and monoploid number met the inequalities above. For the selected random sample, we used Eq. 2 to

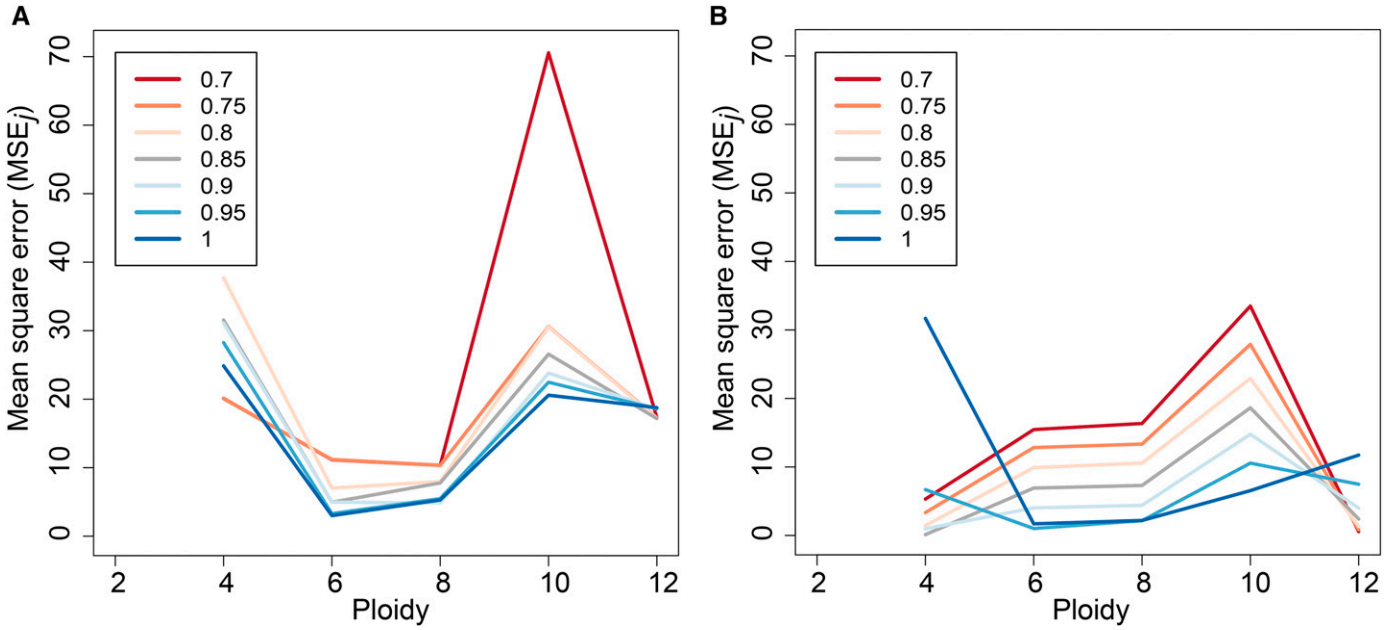

FIGURE 3 (A) Monploid number threshold (MNT) by percentile. The percentile that minimizes MSE by ploidy is 100%, meaning that no monoploid number threshold is necessary. (B) Genome size threshold (GST) by percentile. The smaller percentiles from 75% to 90% minimize MSE at 4*x* and 6*x* ploidies, whereas percentiles 90% and 95% minimize 8*x*, 10*x*, and 12*x* ploidies.

TABLE 2. Summary of genome downsizing simulations at species and genus levels.

Simulations	Downsizing percentage $(1-r)\times 100\%$	Percentile tested	Simulations in Figure	Optimal downsizing percentage	Optimal percentile
Species simulations					
GD	0.0, 0.1,…,99.9	—	3A, 3B	30.1	—
MNT (2 different simulations)	—	70%, 75%,…,100%, and polyploidy series maxima MNT (threshold in Fig. 1A)	2A, 3A	—	100
GD+MNT	0.0, 0.1,…,99.9	Polyploidy series monoploid number maxima MNT (threshold in Fig. 1A)	3A, 3B	30.1	—
GST (2 different simulations)	—	70%,75%,…,100%, and polyploidy series maximum GST	2B	—	**90, 95**
GD+GST	0.0, 0.1,…,99.9	Polyploidy series maxima GST (threshold in Fig. 1B)	3A, 3B	**16.8**	—
MNT+GST	—	Polyploidy series maxima MNT+GST (thresholds in Fig. 1)	3A	—	—
GD+MNT+GST	0.0, 0.1,…,99.9	Polyploidy series maxima MNT+GST (thresholds in Fig. 1)	3A, 3B	**15.4**	—
Genus simulations					
GD	0.0, 0.1,…,99.9	—	3C, 3D	0.3	
MNT	—	Polyploidy series maxima MNT (threshold in Fig. 1A)	3C	—	
GD+MNT	0.0, 0.1,…,99.9	Polyploidy series maxima MNT (threshold in Fig. 1A)	3C, 3D	0.0	
GST	—	Polyploidy series maxima GST (threshold in Fig. 1B)	3C	—	
GD+GST	0.0, 0.1,…,99.9	Polyploidy series maxima GST (threshold in Fig. 1B)	3C, 3D	**0.1**	
MNT+GST	—	Polyploidy series maxima MNT+GST (thresholds in Fig. 1)	3C		
GD+MNT+GST	0.0, 0.1,…,99.9	Polyploidy series maxima MNT+GST (thresholds in Fig. 1)	3C, 3D	**0.1**	

Notes: Summary of bootstrap simulations of genome downsizing (GD), monoploid number threshold (MNT), genome size threshold (GST), and combinations of them. Simulations that include GD assume genome downsizing percentages from 0.0% (no downsizing) to 99.9% (almost total genome downsizing). Optimal downsizing percentage minimizes the value of Total MSE. Simulations that consider MNT and/or GST assume that the thresholds are the determined by observed maximum values of polyploidy series taxa (in bold). Only in the case of species simulations, percentiles of the distribution of genome sizes and monoploid numbers were tested as alternative thresholds. Optimal percentiles are the percentile values that minimize Total MSE (in bold).

find the new subsequent simulated even ploidy values. The mean $\overline{Y}^{j}_{\mathrm{GD+T}}$ was calculated as in Eq. 4.

Estimating the optimal genome downsizing percentage—Determining the genome downsizing percentage loss that best explains the observed genome size means (denoted by $\overline{Y}^{j}_{\mathrm{observed}}$) in the genome downsizing and genome downsizing threshold simulations was done by finding the value of r that minimizes the mean squared error for the mean genome size, $\mathrm{MSE}^{j}(r) = E\left[\left(\overline{Y}^{j}_{\mathrm{observed}} - \overline{Y}^{j}_{\mathrm{GD\ or\ GD+T}}(r)\right)^{2}\right]$, at each even ploidy level $j = 4x,6x,\ldots 12x$. By definition, $\mathrm{MSE}^{j}(r)$ values are close to zero if the observed genome size mean at each ploidy j is equal to the simulated genome size mean under a hypothesis. This is similar to the approach taken in decision theory (Berger, 1985) where a parameter (e.g., genome downsizing percentage) is estimated via minimization of an expected loss function (e.g., mean squared error). Furthermore, minimizing $\mathrm{MSE}^{j}(r)$ for genome downsizing percentage r is equivalent to minimizing least squares, or finding the maximum likelihood estimate in a linear model if errors are distributed Gaussian-like (Casella and Berger, 2002). Hence, the optimal genome downsizing percentage r in the genome downsizing or genome downsizing with thresholds simulations is the one that minimizes the $\mathrm{MSE}^{j}(r)$ at all ploidies (minimizes the negative log-likelihood; Fig. 4B, 4D). Since $\mathrm{MSE}^{j}(r)$ is positive for all ploidies, the optimal downsizing percentage rate is the one that minimizes the sum of all $\mathrm{MSE}^{j}(r)$, defined as Total $\mathrm{MSE}(r)$ (Table 3). Furthermore, the optimal threshold for monoploid number threshold or genome size threshold in the absence of genome downsizing was calculated using the mean squared error at each ploidy, $\mathrm{MSE}^{j} = E\left[\left(\overline{Y}^{j}_{\mathrm{observed}} - \overline{Y}^{j}_{T}\right)^{2}\right]$ (Table 4, notice the absence of dependency of r), and the total mean squared error, $\mathrm{Total\ MSE} = \sum_{j=4x}^{12x} \mathrm{MSE}^{j}$, for each of the polyploidy series thresholds and the percentile thresholds 70%,75%,…100% (Table 3). The threshold that best explains the distribution of genome sizes is the one that minimizes the Total MSE (Fig. 3).

For genus level simulations, we use the same equations for MSE^{j} and Total MSE (Tables 3 and 4), but the mean values $\overline{Y}^{j}_{\mathrm{GD\ or\ GD+T}}(r)$ were calculated as described in Eq. 4, where they are weighed uniformly across all genera. In the case of genera, the values $\overline{Y}^{j}_{\mathrm{observed}}$ represent an average of the averages. In other words, we calculated the mean genome size averages per genus and per ploidy, and those resulting mean values were later averaged. This calculation allowed us to propose a $\overline{Y}^{j}_{\mathrm{observed}}$ that is not biased toward highly sampled genera (Appendix S1).

Comparison of hypotheses—The genome downsizing hypothesis was evaluated based on the genome downsizing simulation with the percentage loss that resulted in the minimum Total MSE (r) (Table 3). The threshold hypothesis was evaluated based on the monoploid

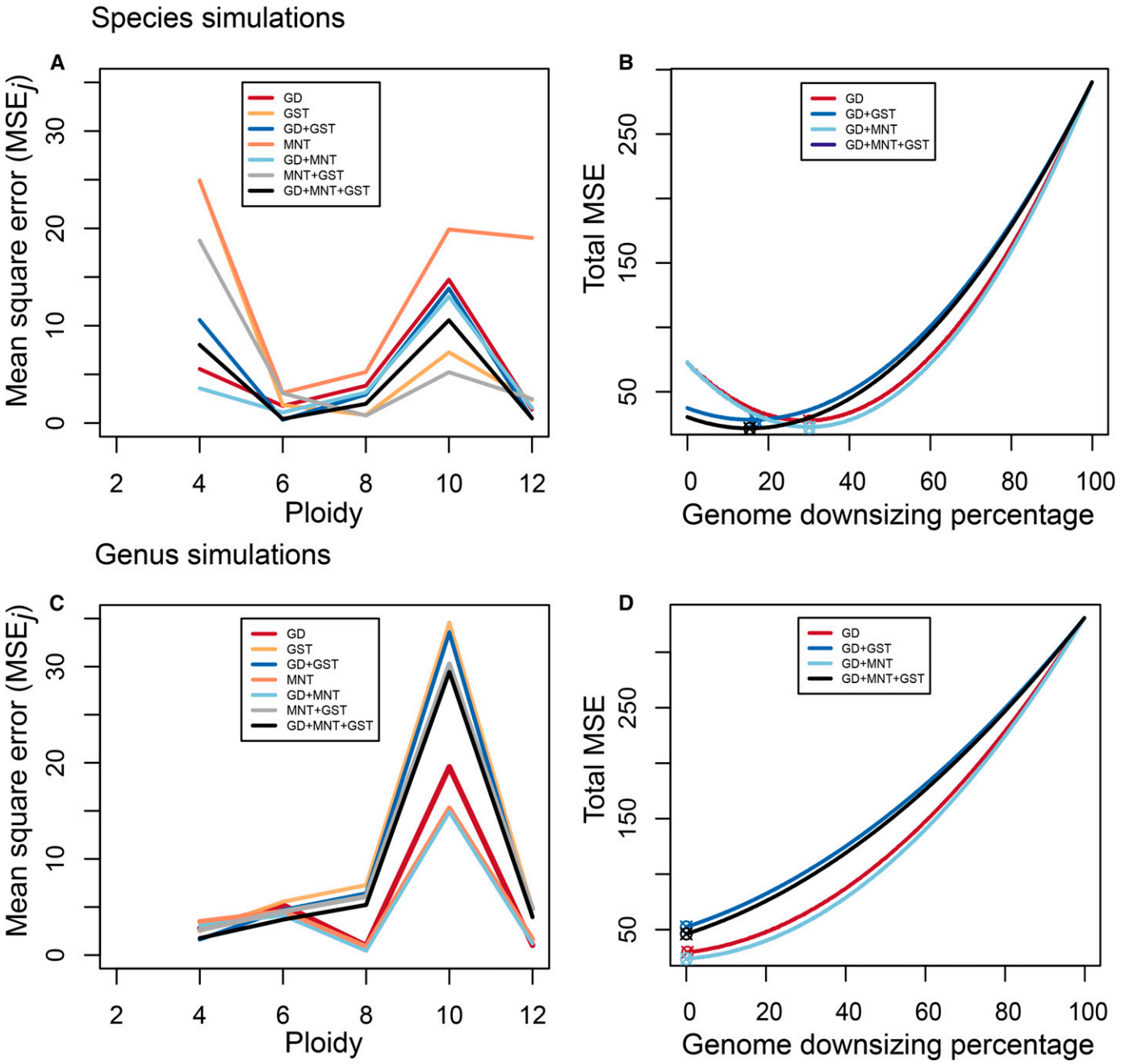

FIGURE 4 (A) Comparison of the genome downsizing (GD), threshold (T), and genome downsizing with threshold hypotheses (GD+T) in the species level simulations. Overall genome downsizing when thresholds are added (GD+T) minimizes MSE at all ploidies, performing better than GD, or any of T alone. (B) Detection of best GD percentage loss in the species level simulations. The genome downsizing percentage loss that performs best under the different hypotheses is the one that minimizes the Total MSE (the sum of the MSE at all ploidies). When a threshold is added to GD, the genome downsizing percentage loss is 50% smaller. (C) Comparison of hypotheses GD, T, and GD+T in the genus level simulations. MNT is the hypothesis that best explains the mean of genome size per ploidy and per genus at all ploidies except for 4*x*. (D) Detection of best GD at species level. The best genome downsizing percentage is 0% (i.e., no genome downsizing) under all the hypotheses.

number threshold, genome size threshold, and monoploid number and genome size threshold simulations (Table 3; Fig. 4A, 4C). Finally, the genome downsizing and threshold hypotheses were evaluated based on the genome downsizing with genome size threshold, genome downsizing with monoploid number threshold, and genome downsizing with both monoploid and genome size thresholds simulations using the genome downsizing percentage loss that minimized the Total MSE (Table 3; Fig. 4B, 4D). The best hypotheses for the species level and genus level at a given ploidy were determined based on individual minimum MSE^j at each ploidy (Table 4) and the Total MSE (Table 3; Fig. 4) at all ploidies simultaneously. Furthermore, comparisons between hypotheses were made using

TABLE 3. Optimal genome downsizing percentage and comparison of hypotheses.

Simulations	Rate at best Total MSE (%)	MSE at best Total MSE					Minimum Total MSE	~AIC
		4*x*	6*x*	8*x*	10*x*	12*x*		
Species								
GD	30.1	5.558	1.757	3.835	14.736	1.374	27.260	56.520
GST	—	24.947	1.847	0.832	7.262	2.366	37.254	84.508
GST+GD	16.8	10.579	0.329	2.931	13.813	0.471	28.123	68.246
MNT	—	24.877	3.098	5.237	19.895	19.015	72.123	154.247
MNT+GD	30.1	3.572	1.091	3.142	12.980	1.565	22.350	56.699
GST+MNT	—	18.739	3.062	0.765	5.221	2.471	30.258	70.516
GST+MNT+GD	15.4	8.035	0.425	1.970	10.564	0.494	**21.488**	**54.976**
Genus								
GD	0.3	2.758	5.202	0.974	19.561	1.023	29.518	61.035
GST		2.050	5.124	6.854	4.146	4.400	52.574	115.147
GST+GD	0.0	2.049	5.125	6.835	33.985	4.369	52.362	116.724
MNT		3.058	4.108	0.449	14.862	1.224	**23.701**	**57.402**
MNT+GD	0.0	3.059	4.129	0.437	14.880	1.196	23.701	59.401
GST+MNT		2.187	4.116	5.665	29.832	4.374	46.175	102.340
GST+MNT+GD	0.0	2.185	4.117	5.644	29.861	4.398	46.204	104.408

Notes: Values of MSE^{J} by ploidy at minimum Total MSE as shown in Fig. 3A and 3C. For species level simulations, GD with both MNT and GST minimize the Total MSE (in bold). For genus-level simulations, Total MSE is minimized when the MNT hypothesis is considered (in bold).

an approximation to Akaike information criterion (~AIC) assuming that Total MSE is equivalent to approximating the negative log-likelihood, that allowed us to compare all hypotheses fairly since biases by number of parameters were corrected.

RESULTS

Polyploidy series data—The most frequent associations between ploidies in the polyploidy series data are between subsequent even ploidies, followed by associations between an odd ploidy and the next even ploidy (e.g., 3*x* to 4*x*, or 5*x* to 6*x*; Table 1). For all even ploidies except 12*x*, the thresholds determined by maximum genome sizes and monoploid number in polyploidy series taxa are larger than the 85% percentiles observed genome sizes at each ploidy (see Appendix S1 for exact percentile values of genome sizes).

For the PGLS analysis under Brownian motion evolution, we found significant evidence for drift ($\hat{\sigma}^2 = 1.06$, Fig. 2A) and a maximum likelihood estimate for $(\widehat{1-r}) = 0.79$, $p < 10^{-13}$) that resulted

TABLE 4. Summary of genome downsizing percentage estimates per ploidy.

Simulations	Quantity	4*x*	6*x*	8*x*	10*x*	12*x*
Species						
GD	Percentage loss	51.3	15.4	16.1	18.1	33.9
	Minimum MSE^{J} at each ploidy	0.080	0.314	1.983	12.934	1.076
GST+GD	Percentage loss	48.5	15.3	0.4	0.1	15.2
	Minimum MSE^{J} at each ploidy	**0.065**	0.315	0.817	7.168	**0.454**
MNT+GD	Percentage loss	48.0	19.8	18.1	24.9	35.5
	Minimum MSE^{J} at each ploidy	0.078	0.294	1.936	12.359	1.120
GST+MNT+GD	Percentage loss	44.50	19.0	1.5	0.10	14.7
	Minimum MSE^{J} at each ploidy	**0.060**	**0.289**	**0.755**	**5.262**	0.486
Genus						
GD	Percentage loss	29.5	0.0	0.2	0.3	6.0
	Minimum MSE^{J} at each ploidy	0.007	5.125	0.937	19.561	**0.713**
GST+GD	Percentage loss	26.6	0.0	0.0	0.1	0.0
	Minimum MSE^{J} at each ploidy	0.007	5.125	0.937	19.561	**0.713**
MNT+GD	Percentage loss	30.5	0.0	0.0	0.0	6.3
	Minimum MSE^{J} at each ploidy	0.007	4.129	**0.437**	**14.880**	0.821
GST+MNT+GD	Percentage loss	27.20	0.0	0.0	0.10	0.0
	Minimum MSE^{J} at each ploidy	**0.006**	**4.117**	5.644	29.808	4.398

Notes: Minimum values of MSE^{J} by ploidy and corresponding genome downsizing percentage loss for each hypothesis (minima showed in bold). Evidence for genome downsizing mostly comes from simulation of tetraploids for which more than 25% percentage downsizing can explain the mean genome size of tetraploid level with or without the presence of thresholds. For ploidies larger than 8*x*, thresholds become more critical since MSE^{J} is minimized for hypotheses considering thresholds.

in a genome downsizing percentage of 20.1%, $CI^{95\%}$ = (11.5%, 28.8%). This model has an AIC of 711.81. When assuming an OU transformation for the covariance among taxa maximum likelihood estimate for $\hat{\alpha} = 3.69$, and its profile likelihood is shown in Fig. 2B. Large values of α represent very low phylogenetic correlation between two individuals under this model of evolution of genome size since the covariance between two individuals is defined as $v_{ij} = \sigma^2\left(e^{-\alpha t_{ij}}\right)$(Martins and Hansen, 1997). The AIC of the OU model is 514.25, better than Brownian motion. The maximum likelihood estimate for genome downsizing percentage in this case is 2.2%, $\left(\widehat{1-r}\right) = 0.97$, $p < 10^{-13}$, $CI^{95\%}$ = (0%, 4.7%).

To further assess whether the phylogenetic signal is strong for the same linear model in Eq. 1, we used Pagel's λ transformation. This resulted in a maximum likelihood estimate for $\hat{\lambda} = 0$ (Fig. 2C), and the genome downsizing percentage estimate was 2.2%, $\left(\widehat{1-r}\right) = 0.97$, $p < 10^{-13}$, $CI^{95\%}$ = (0%, 4.7%). The AIC for the linear model with Pagel's λ was 513.52, approximately equal to AIC of the model with OU transformation.

Simulation for genome downsizing—For the genome downsizing at species level simulation experiments, the optimal downsizing percentage r is 30%, $CI^{95\%}$ = (23.6%, 35.8%) (Table 3; Fig. 4B). This value was determined by minimizing the Total MSE(r) at all ploidies. However, during simulations for $4x$, the optimal percentage loss was near 50% based on the minimum MSE^j obtained for tetraploids, whereas smaller percentage losses (<35%) were optimal in simulations of larger ploidies (Table 4). For the genome downsizing bootstrap simulations at the genus level, the optimal downsizing percentage across all ploidies was 0, i.e., no downsizing; $CI^{95\%}$ = (0%, 4.3%) (Table 3; Fig. 4). In the $4x$ simulations, the optimal percentage loss was near 30%, but it was less than 7% in the simulations of ploidies larger than $4x$ (Table 4; see Appendix S1 for calculation of confidence intervals).

Simulation for thresholds—In the species level simulations, we found that across all ploidies, the 100% percentile of monoploid number threshold minimizes Total MSE, indicating that there is no maximum monoploid number threshold and adding a monoploid number threshold to simulations does not approximate average genome size across all ploidies (Table 2). However, at $4x$ ploidy, the 70% and 75% percentile thresholds, which represent the most restrictive threshold for monoploid number at species level, minimize the MSE^j (Fig. 3A). In general, monoploid number threshold is not restrictive for ploidies larger than $4x$, since larger percentiles 90%, 95%, and 100% have a smaller value of MSE^j (Fig. 3A). The monoploid number threshold determined by polyploidy series taxa performs like a 99% percentile (Appendix S1), and it is the hypothesis that performs worst at recovering the mean genome size of species at all ploidies (Table 3; Fig. 3A). In contrast, in the genus level simulations, monoploid number threshold determined by polyploidy series minimizes Total MSE, and thus represents the hypothesis that best recovers the observed average genome sizes at different ploidies (Table 3, Fig. 4C).

For the species level genome size threshold simulations, unrestrictive thresholds of 90% and 95% percentiles minimize Total MSE across all ploidies; however, threshold percentiles from 70 to 90% minimize the MSE^j at ploidies $4x$ and $12x$, while 90%, 95%, and 100% minimize the MSE^jat ploidies j = 6x, 8x,10x (Fig. 3B). Using the genome size threshold determined by the maxima in the polyploidy series taxa performs similarly to genome size threshold at the 95% percentile in simulations at the species level (Appendix S1), approximating the mean genome size better than monoploid number (Table 3, Fig. 4A). In contrast, in the genus level simulations, genome size threshold defined by polyploidy series is the worst hypothesis at approximating mean genome size (Table 3).

Assuming that the genome size and monoploid number thresholds were determined by polyploidy series taxa allowed us to detect whether there is an interaction effect at the species level that might improve the prediction of mean genome size. Our species level simulation using both thresholds improves the MSE^jonly at ploidies $j = 8x,10x$ (Fig. 4A). In the genus level simulations, the interaction of genome size and monoploid number thresholds is only better than monoploid number threshold for $4x$ (Table 3; Fig. 4C). Thus, genome size and genome size with monoploid number thresholds are better for the species level simulations, but the strength and type of the threshold depends on ploidy (Table 3). For genus level simulations, monoploid number threshold is the only critical threshold for ploidies larger than $6x$.

Simulation for genome downsizing in the presence of thresholds—In the species level simulations, when genome size threshold of polyploidy series is added to the genome downsizing hypothesis, the optimal genome downsizing percentage loss drops to 16.8% [from 30% for genome downsizing alone, $CI^{95\%}$ = (8.6%, 22.7%)], but if monoploid number threshold alone is added to genome downsizing, the optimal genome downsizing percentage remains 30%, $CI^{95\%}$ = (24.7%, 35.8%) (Table 3; Fig. 4B). If genome size and monoploid number thresholds are added to the GD simulations, the best genome downsizing percentage is 15.4%, $CI^{95\%}$ = (8.2%, 22.4%) (Fig. 4B). Adding the two thresholds (genome size and monoploid number) reduces the $MSE^j(r)$ at higher ploidies (Table 3; Fig. 4A). In general, when a threshold is used, downsizing percentage loss values between 8 and 23% minimize the Total $MSE^j(r)$ more than using only genome downsizing in the simulations (Fig. 4B).

In the genus level simulations, genome downsizing does not seem to have an important role when combined with the genome size or monoploid number thresholds (Fig. 4C). In all cases, the best genome downsizing percentage loss is 0, $CI^{95\%}$ = (0%, 4.8%) (Fig. 4D). However, in the presence of any genome downsizing, the monoploid number threshold hypothesis best explains the mean genome size per genus and per ploidy (Fig. 4D). Using either monoploid number threshold and/or genome size threshold, the largest genome downsizing percentages that minimize MSE^j are at $4x$ ploidy (Table 4).

Comparison of hypotheses—In the species level simulations, the genome downsizing hypothesis with both monoploid number and genome downsizing threshold (~AIC = 54.97) with optimal genome downsizing rate of approximately 15.4% (Fig. 4B) is the hypothesis that best approximates the observed distribution of genome sizes at all ploidies. Under the genome downsizing hypothesis alone, the optimal downsizing percentages minimizing Total MSE(r) is 30.1% (~AIC = 56.52). Genome downsizing hypothesis is indistinguishable from genome downsizing with monoploid number threshold (~AIC = 56.69). The reduction of genome downsizing percentage is mostly due to the addition of the genome size threshold (Table 3). The genome downsizing hypothesis with genome size threshold alone (~AIC = 68.24) did not perform well compared with hypotheses that added monoploid number threshold

(Table 3). For species level simulations, hypotheses that only considered thresholds perform worse than hypotheses that include genome downsizing (Table 3).

In contrast, for the genus level simulations, the monoploid number threshold hypothesis best explains the mean genome size per genus and per ploidy by reducing the MSE^j at all ploidies (Fig. 4C, ~AIC = 57.40). Adding genome downsizing to monoploid number threshold slightly worsens the fit to observed genome size distributions (~AIC = 59.40, Table 3). Interestingly, if genome size threshold is added to any hypothesis, that produces the worst approximations (Table 3).

DISCUSSION

In this study we used simulations to assess the effects of genome downsizing after WGDs and maximum genome size or chromosome number thresholds on the relationship between average genome size and ploidy number in angiosperms. Genome downsizing appears to affect the average genome sizes, especially at tetraploid level, but our results emphasize the importance of constraints on genome size or monoploid number. While the role of genome downsizing following WGD is widely acknowledged (Petrov et al., 2000; Leitch and Bennett, 2004; Wang et al., 2005; Leitch et al., 2008; Renny-Byfield et al., 2011; Ibarra-Laclette, 2013), the role of genomic constraints on WGDs has received relatively little attention (but see De Smet et al., 2013). Yet thresholds on monoploid number alone, without any genome downsizing, best explain the average genome size at different ploidies in the genus level simulations, and failing to consider maximum genome size thresholds can result in overestimating the effects of genome downsizing in the species level simulations.

There have been relatively few direct estimates of genome downsizing percentages or rates from specific plant species (e.g., Leitch et al., 2008; Hawkins et al., 2009; Renny-Byfield et al., 2011). Leitch et al. (2008) found examples of genome downsizing ranging from 1.9 to 14.3% after WGDs in *Nicotiana*. In the recent allopolyploid *Nicotiana tabacum*, genome downsizing appears to be largely due to reductions in repetitive DNAs, such as retroelements, mostly from the *N. tomentosiformis* parental genome (Renny-Byfield et al., 2011). Hawkins et al. (2009) studied *gypsy*-like retrotransposon *Gorge3* to determine the directionality of genome size change in three species of *Gossypium*. Using an exponential model in a phylogenetic context, the authors found a decreasing rate of genome size change for *G. ramondii* (−4.96 Mb), whereas *G. herbaceoum* and *G. exigum* showed an increasing rate (3.91Mb and 11.12Mb, respectively). They concluded that some lineages might be more successful than others at reducing or preventing proliferation of transposable elements that ultimately contribute to genome size determination in cotton. Our genome downsizing simulations indicate optimal genome downsizing rates between 0 and 30% across polyploidy series taxa, species, and genus analyses. Thus, genome downsizing does not necessarily have to be large to explain the lack of a strong relationship between average genome size and ploidy level, and at the genus level, it may not play a role at all.

It is also possible to underestimate genome downsizing percentage if ploidy is misclassified. Ploidy has been traditionally defined based on chromosome number multiples (Stebbins, 1938). With more information about genome and gene duplications, redefining ploidy for many taxa will be critical for accurately estimating genome downsizing and thresholds. In the near future, models estimating genome downsizing based on ploidy change will also have to consider how diploidization (Wolfe, 2001) further expands the variance of diploid genome sizes and how diploidization can bias estimations of genome downsizing.

Maximum genome size and monoploid number thresholds also affect the relationship between average genome size and ploidy level. The evidence for the importance of genome size threshold mainly comes from the simulations of tetraploids, for which genome sizes thresholds best explained the mean genome size in both the genus and species simulations. Plants with large genomes may have low speciation rates, are not well adapted to extreme environments, and have low maximum photosynthetic rates (Knight et al., 2005). These factors may affect new tetraploids as they are trying to become established. Furthermore, a species with a large genome size might require a large supply of nutrients (e.g., phosphorus), preventing its establishment in nutrient-limited environments (Šmarda et al., 2013). Thus, there may be selection against duplicating genome sizes in species with especially large content (Gaut and Ross-Ibarra, 2008). Importantly, if the genome size threshold is considered, it greatly reduces the amount of genome reduction needed to explain the distribution of average genome sizes across different ploidy levels in the species level simulations. However, adding monoploid number to the genome downsizing hypothesis allowed us to better approximate mean genome size at all ploidies. These results demonstrate potentially complex interactions between genome downsizing, thresholds, and ploidy.

Interestingly, the results for species and genus level simulations differ with respect to the monoploid number threshold hypothesis (Table 3; Fig. 3A, 3B). While monoploid number threshold alone does not explain mean genome size at any ploidy species level simulations, this threshold best explains the mean genome size in the genus level simulations at large ploidies. This result indicates the amount of genome size variability per monoploid number found in specific plant lineages (Greilhuber et al., 2005). Still, monoploid number threshold added to genome downsizing is the second best hypothesis for the species level simulations, even when it does not affect estimates of the genome downsizing percentage. The importance of the monoploid number threshold suggests that the number and repetition of chromosomes can eventually inhibit or prevent multiple WGD through failed meiosis (Cifuentes et al., 2010; Zielinski and Schied, 2012; Moore, 2013) or that particular karyological characteristics of a genus directly relate to the possibility of WGD, as recently found in *Carex* (Lipnerová et al., 2013). In the context of the C-value paradox (Lynch and Walsh, 2007), the monoploid number threshold modifies genome size evolution in two directions. On the one hand, large genome sizes associated with small monoploid numbers will remain in the population, increasing mean genome size over time, similarly to Lynch and Conery (2003) arguments about initial nonadaptive processes acting on genome size evolution. On the other hand, despite genome size, many species might not be undergoing WGDs due to the large monoploid number, ultimately decreasing the diversity of genome sizes in higher ploidies.

The strength of genome size and monoploid number thresholds heavily depends on the way we represent the distribution of genome sizes. Statistics like the mean genome size of species may be biased toward highly sampled genera. Whereas using the mean genome size weighted by genus may ameliorate some taxonomic biases, it also greatly decreases sample size. Although we may get a

more direct perspective on the processes affecting genome sizes after WGDs by comparing genomes of polyploidy taxa with those of their parental diploid genomes or closer relatives (e.g., Vitte and Bennetzen, 2006; Grover et al., 2007; Leitch et al., 2008; Eilam et al., 2008) or by phylogenetic approaches like the ones presented here for polyploidy series taxa, the diploid ancestors for many current polyploids are extinct or unknown (as shown in *Persicaria* allopolyploids; Kim et al., 2008). Modeling the process of genome downsizing and threshold restrictions throughout a phylogeny and integrating sampling error or variation on genome sizes of the tips will further help evaluate the importance of the monoploid number and genome size threshold hypotheses to explain the current distribution of genome sizes. Currently, comparative method models for a semidiscrete trait like genome size are being developed (Landis et al., 2013); however, these methods do not currently incorporate changes after WGDs or allow for thresholds.

This study presents a new step toward resolving the paradoxical relationship between average genome size and ploidy level in angiosperms, suggesting that constraints on maximum genome size and maximum monoploid number for WGDs, and not only genome downsizing, limit angiosperm genome sizes in the face of frequent WGDs. We demonstrate that a large-scale simulation approach can provide some insights into the effects of WGDs on genome size. Still, our simulation experiment was based on a number of assumptions that greatly simplify the process of WGD. For example, we include only one type of ploidy change (even ploidy increasing to the next highest even ploidy), although other ploidy changes can occur (Stebbins, 1966, 1971; Ramsey and Schemske, 1998; Husband et al., 2013). We also assume a single percentage of genome downsizing occurs immediately after all WGDs, no matter the initial genome size, ploidy level, or type of WGD. It is possible that the genome downsizing percentage is time dependent (Soltis et al., 2003), ploidy dependent (Leitch et al., 2008), or genome size dependent (Ibarra-Laclette et al., 2013), and it may occur differently in allopolyploidy vs. autopolyploidy (Meyers and Levin, 2006; Husband et al., 2013). Our simulations implicitly assume autopolyploidy, but considering simulations for allopolyploidy implies the selection of two diploid genomes that are "compatible". Unfortunately, what should be considered allopolyploid compatibility among genome sizes is not easy to define. We hope that the methodology presented here can be extended to explore more factors potentially affecting the relationship between polyploidy and genome size evolution.

ACKNOWLEDGEMENTS

The authors thank P. S. Soltis and M. M. Miyamoto for insightful comments about genome size dynamics and B. C. O'Meara and J. M. Ferguson for advice about the methods. They also thank M. S. Barker and two anonymous reviewers for thoughtful comments. Funding for R.Z.-F. provided by NSF IGERT program (DGE-0801544) and NSF DDIG (DEB-1501547).

LITERATURE CITED

Adams, K. L., and J. F. Wendel. 2005. Polyploidy and genome evolution in plants. *Current Opinion in Plant Biology* 8: 135–141.

Arrigo, N., and M. S. Barker. 2012. Rarely successful polyploids and their legacy in plant genomes. *Current Opinion in Plant Biology* 15: 140–146.

Bennett, M. D., and I. J. Leitch. 1995. Nuclear DNA amounts in angiosperms. *Annals of Botany* 76: 113–176.

Bennett, M. D., and I. J. Leitch. 2012. Plant DNA C-values database, release 6.0, December 2012. Kew Botanic Gardens, Kew, UK. Available at data.kew.org/cvalues/.

Bennetzen, J. L. 2002. Mechanisms and rates of genome expansion and contraction in flowering plants. *Genetica* 115: 29–36.

Bennetzen, J. L., and E. A. Kellogg. 1997. Do plants have a one-way ticket to genomic obesity? *Plant Cell* 9: 1509–1514.

Berger, J. O. 1985. Statistical decision theory and Bayesian analysis. Springer Science & Business Media, New York, New York, USA.

Casella, G., and R. L. Berger. 2002. Statistical inference, vol. 2. Duxbury, Pacific Grove, California, USA.

Cifuentes, M., L. Grandont, G. Moore, A. M. Chèvre, and E. Jenczewski. 2010. Genetic regulation of meiosis in polyploid species: New insights into an old question. *New Phytologist* 186: 29–36.

De Smet, R., K. L. Adams, K. Vandepoele, M. C. Van Montagu, S. Maere, and Y. Van de Peer. 2013. Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. *Proceedings of the National Academy of Sciences, USA* 110: 2898–2903.

Eilam, T., Y. Anikster, E. Millet, J. Manisterski, and M. Feldman. 2008. Nuclear DNA amount and genome downsizing in natural and synthetic allopolyploids of the genera *Aegilops* and *Triticum*. *Genome* 51: 616–627.

Freckleton, R. P., P. H. Harvey, and M. Pagel. 2002. Phylogenetic analysis and comparative data: A test and review of evidence. *American Naturalist* 160: 712–726.

Gaut, B. S., and J. Ross-Ibarra. 2008. Selection on major components of angiosperm genomes. *Science* 320: 484–486.

Greilhuber, J., J. Doležel, M. A. Lysak, and M. D. Bennett. 2005. The origin, evolution and proposed stabilization of the terms 'genome size'and 'C-value' to describe nuclear DNA contents. *Annals of Botany* 95: 255–260.

Grover, C. E., H. Kim, R. A. Wing, A. H. Paterson, and J. F. Wendel. 2007. Microcolinearity and genome evolution in the *AdhA* region of diploid and polyploid cotton (*Gossypium*). *Plant Journal* 50: 995–1006.

Hawkins, J. S., C. E. Grover, and J. F. Wendel. 2008. Repeated big bangs and the expanding universe: Directionality in plant genome size evolution. *Plant Science* 174: 557–562.

Hawkins, J. S., H. Kim, J. D. Nason, R. A. Wing, and J. F. Wendel. 2006. Differential lineage-specific amplification of transposable elements is responsible for genome size variation in Gossypium. *Genome Research* 16: 1252–1261.

Hawkins, J. S., S. R. Proulx, R. A. Rapp, and J. F. Wendel. 2009. Rapid DNA loss as a counterbalance to genome expansion through retrotransposon proliferation in plants. *Proceedings of the National Academy of Sciences, USA* 106: 17811–17816.

Husband, B. C., S. J. Baldwin, and J. Suda. 2013. The incidence of polyploidy in natural plant populations: Major patterns and evolutionary processes. *In* I. J. Leitch, J. Greihuber, J. Dolezel, and J. Wendel [eds.], Plant genome diversity, vol. 2, 255–276. Springer, Vienna, Austria.

Ibarra-Laclette, E., E. Lyons, E. Hernández-Guzmán, C. A. Pérez-Torres, L. Carretero-Paulet, T. Chang, T. Lan, et al. 2013. Architecture and evolution of a minute plant genome. *Nature* 498: 94–98.

Jiao, Y., N. J. Wickett, S. Ayyampalayam, A. S. Chanderbali, L. Landherr, P. E. Ralph, L. P. Tomsho, et al. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* 473: 97–100.

Kim, S., S. E. Sultan, and M. J. Donoghue. 2008. Allopolyploid speciation in *Persicaria* (Polygonaceae): Insights from a low-copy nuclear region. *Proceedings of the National Academy of Sciences, USA* 105: 12370–12375.

Knight, C. A., and J. M. Beaulieu. 2008. Genome size scaling through phenotype space. *Annals of Botany* 101: 759–766.

Knight, C. A., N. A. Molinari, and D. A. Petrov. 2005. The large genome constraint hypothesis: Evolution, ecology and phenotype. *Annals of Botany* 95: 177–190.

Landis, M. J., J. G. Schraiber, and M. Liang. 2013. Phylogenetic analysis using Lévy processes: Finding jumps in the evolution of continuous traits. *Systematic Biology* 62: 193–204.

Leitch, I. J., and M. D. Bennett. 1997. Polyploidy in angiosperms. *Trends in Plant Science* 2: 470–476.

Leitch, I. J, and M.D. Bennett. 2004. Genome downsizing in polyploid plants. *Biological Journal of the Linnean Society* 82: 651–663.

Leitch, I. J., L. Hanson, K. Y. Lim, A. Kovarik, M. W. Chase, J. J. Clarkson, and A. R. Leitch. 2008. The ups and downs of genome size evolution in polyploid species of *Nicotiana* (Solanaceae). *Annals of Botany* 101: 805–814.

Lipnerová, I., P. Bureš, L. Horová, and P. Šmarda. 2013. Evolution of genome size in *Carex* (Cyperaceae) in relation to chromosome number and genomic base composition. *Annals of Botany* 111: 79–94.

Lynch, M., and J. S. Conery. 2003. The origins of genome complexity. *Science* 302: 1401–1404.

Lynch, M., and B. Walsh. 2007. The origins of genome architecture, vol. 98. Sinauer, Sunderland, Massachusetts, USA.

Martins, E. P., and T. F. Hansen. 1997. Phylogenies and the comparative method: A general approach to incorporating phylogenetic information into the analysis of interspecific data. *American Naturalist* 149: 646–667.

Mayrose, I., M. S. Barker, and S. P. Otto. 2010. Probabilistic models of chromosome number evolution and the inference of polyploidy. *Systematic Biology* 59: 132–144.

Meyers, L. A., and D. A. Levin. 2006. On the abundance of polyploids in flowering plants. *Evolution* 60: 1198–1206.

Michael, T. P. 2014. Plant genome size variation: Bloating and purging DNA. *Briefings in Functional Genomics* 13:308–317.

Moore, G. 2013. Meiosis in polyploids. *In* Z. J. Chen and J. A. Birchler [eds.], Polyploid and hybrid genomics, 241–255.

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

Pagel, M. 1997. Inferring evolutionary processes from phylogenies. *Zoologica Scripta* 26: 331–348.

Petrov, D. A., T. A. Sangster, J. S. Johnston, D. L. Hartl, and K. L. Shaw. 2000. Evidence for DNA loss as a determinant of genome size. *Science* 287: 1060–1062.

Ramsey, J., and D. W. Schemske. 1998. Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual Review of Ecology, Evolution, and Systematics* 29: 467–501.

Renny-Byfield, S., M. Chester, A. Kovarík, S. C. Le Comber, M.-A. Grandbastien, M. Deloger, R. A. Nichols, et al. 2011. Next generation sequencing reveals genome downsizing in allotettraploid *Nicotiana tabacum*, predominantly through the elimination of paternally derived repetitive DNAs. *Molecular Biology and Evolution* 28: 2843–2854.

Ross-Ibarra, J. 2007. Genome size and recombination in angiosperms: A second look. *Journal of Evolutionary Biology* 20: 800–806.

Sanmiguel, P., and J. L. Bennetzen. 1998. Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. *Annals of Botany* 82: 37–44.

Šmarda, P., M. Hejcman, A. Březinová, L. Horová, H. Steigerová, F. Zedek, P. Bureš, P. Hejcmanová, and J. Schellberg. 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.

Soltis, D. E., R. J. A. Buggs, W. B. Barbazuk, S. Chamala, M. Chester, J. P. Gallagher, P. S. Schnable, and P. S. Soltis. 2012. The early stages of polyploidy: Rapid and repeated evolution in *Tragopogon*. *In* P. S. Soltis and D. E. Soltis [eds.], Polyploidy and genome evolution, 271–292. Springer, New York, New York, USA.

Soltis, D. E., and P. S. Soltis. 1999. Polyploidy: Recurrent formation and genome evolution. *Trends in Ecology & Evolution* 14: 348–352.

Soltis, D. E., P. S. Soltis, M. D. Bennett, and I. J. Leitch. 2003. Evolution of genome size in the angiosperms. *American Journal of Botany* 90: 1596–1603.

Stebbins, G. L. 1938. Cytological characteristics associated with the different growth habits in the dicotyledons. *American Journal of Botany* 25: 189–198.

Stebbins, G. L. 1966. Chromosomal variation and evolution. *Science* 152: 1463–1469.

Stebbins, G. L. 1971. Chromosomal evolution in higher plants. Edward Arnold, London, UK.

Tiley, G. P., and J. G. Burleigh. 2015. The relationship of recombination rate, genome structure, and patterns of molecular evolution across angiosperms. *BMC Evolutionary Biology* 15: 194–208.

Vitte, C., and J. L. Bennetzen. 2006. Analysis of retrotransposon structural diversity uncovers properties and propensities in angiosperm genome evolution. *Proceedings of the National Academy of Sciences, USA* 103: 17638–17643.

Wang, X. Y., X. L. Shi, B. L. Hao, S. Ge, and J. C. Luo. 2005. Duplication and DNA segmental loss in the rice genome: Implications for diploidization. *The New Phytologist* 165: 937–946.

Wolfe, K. H. 2001. Yesterday's polyploids and the mystery of diploidization. *Nature Reviews. Genetics* 2: 333–341.

Zanne, A. E., D. C. Tank, W. K. Cornwell, J. M. Eastman, S. A. Smith, R. G. FitzJohn, D. J. McGlinn, et al. 2014. Three keys to the radiation of angiosperms into freezing environments. *Nature* 506: 89–92.

Zielinski, M. L., and O. M. Scheid. 2012. Meiosis in polyploid plants. *In* P. S. Soltis and D. E. Soltis [eds.], Polyploidy and genome evolution, 33–55. Springer, Berlin, Germany.