ON THE NATURE OF THINGS: ESSAYS

New Ideas and Directions in Botany

The wondrous cycles of polyploidy in plants[1]

Jonathan F. Wendel[2]

Readers of the *American Journal of Botany* likely will agree that we are living in an exhilarating era to be a plant biologist. The seemingly endless stream of remarkable breakthroughs appearing in the pages of this and other journals, fueled in part by technological advances, has led to increasing integration among formerly disparate disciplines and across scales ranging from the molecular to the biosphere. Nowhere is the pace of progress more apparent than in plant evolutionary genomics, where the application of massively parallel sequencing approaches and advances in computational and bioinformatic capabilities have led to breathtaking insights about the structure, evolution, and function of plant genomes and their various genomic residents. Here I aim to highlight one key realization of this "genomics era" in which we find ourselves, namely, that *all* modern flowering plant genomes derive from processes set in motion by a history of repeated, episodic whole-genome doubling, or polyploidy.

Botanists have long appreciated that chromosome doubling is important in angiosperm diversification and that it remains an active mode of speciation in many genera. Polyploidy has been apparent for more than a century, based mostly on comparative analyses of chromosome numbers within and between many genera and families. Polyploids are known to originate within individuals, or following hybridization between closely related populations (autopolyploidy), or from interspecific hybridization events (allopolyploidy). Yet despite this historical recognition that polyploidy is important in many groups, the genomics era and its attendant explosion in expressed sequence tag (EST) and genome sequencing projects ushered in a surprising new realization, namely, that seed plant and angiosperm evolution are replete with whole-genome-doubling events (Jiao et al., 2011; Paterson et al., 2012), with each subsequent polyploidy superimposed on the genomic remnants surviving from earlier rounds of polyploidy. Thus, the modern view of the canonical angiosperm genome is that it has experienced multiple episodic polyploidy events, evidenced by a tell-tale pattern of nested intragenomic syntenies, often shared among close relatives but varying widely and in a lineage-specific fashion among different angiosperm groups. Writ large, over the grand sweep of angiosperm history (Fig. 1), we now understand that modern angiosperm genomes range in genomic complexity from those that have experienced few genomic multiplication events (e.g., *Amborella*, *Allium*, *Olea*, *Theobroma*) to others that reflect as many as 128 (*Saccharum*), 144 (*Gossypium*), and even 288 (*Brassica*) genomic multiples. This genome multiplicity is an extraordinary realization, with myriad implications for our understanding of genomic architecture, plant gene family structure, and plant function.

Given this new and improved understanding of angiosperm genomic architecture, why was it not classically recognized? The answer to this question lies in the equally surprising spectrum of genomic processes set in motion by polyploidy (Wendel, 2000; Jackson and Chen, 2010; Soltis and Soltis, 2012). It is useful to partition these many processes according to their temporal dynamics, which range in scale from immediate to long term. Several of the immediate and short-term consequences of the merger of two diverged genomes (in the case of allopolyploidy) are modeled in Fig. 1 (upper right), using work from my own laboratory (Wendel et al., 2012) on *Gossypium* (cotton) as a convenient example, noting that many of these same phenomena now have been observed in multiple polyploid systems including *Tragopogon*, *Glycine*, *Arabidopsis*, *Triticum*, *Brassica*, *Nicotiana*, and *Oryza* (Soltis and Soltis, 2012). Responses at the genomic (DNA) level include mutagenic gene silencing or deletional loss, intergenomic transfer of repetitive elements such as transposable elements, differential rates of accumulation of synonymous or nonsynonymous nucleotide substitutions, and various forms of homoeologous (duplicated copies generated by polyploidy) interaction or gene conversion that generate sequence chimeras or duplicated genes (shown as "recombination"). The scale and scope of these phenomena vary among systems, genes, and genomic regions, and in most cases there is little understanding of phenotypic consequence or ecological or evolutionary significance. Similarly, new and young polyploids typically display novel transcriptomic (Grover et al., 2012; Coate and Doyle, 2014) and/or proteomic (Hu et al., 2011; Koh et al., 2012) phenomena, including biased gene expression with respect to homoeolog (duplicate gene copies resulting from polyploidy), expression level dominance with respect to the two different progenitor diploid genomes, and expression subfunctionalization (partitioning of aggregate ancestral expression among homoeologs) or neofunctionalization

[1] Manuscript received 7 July 2015; revision accepted 18 August 2015.

Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, Iowa 50011

[2] E-mail for correspondence (jfw@iastate.edu)

doi:10.3732/ajb.1500320

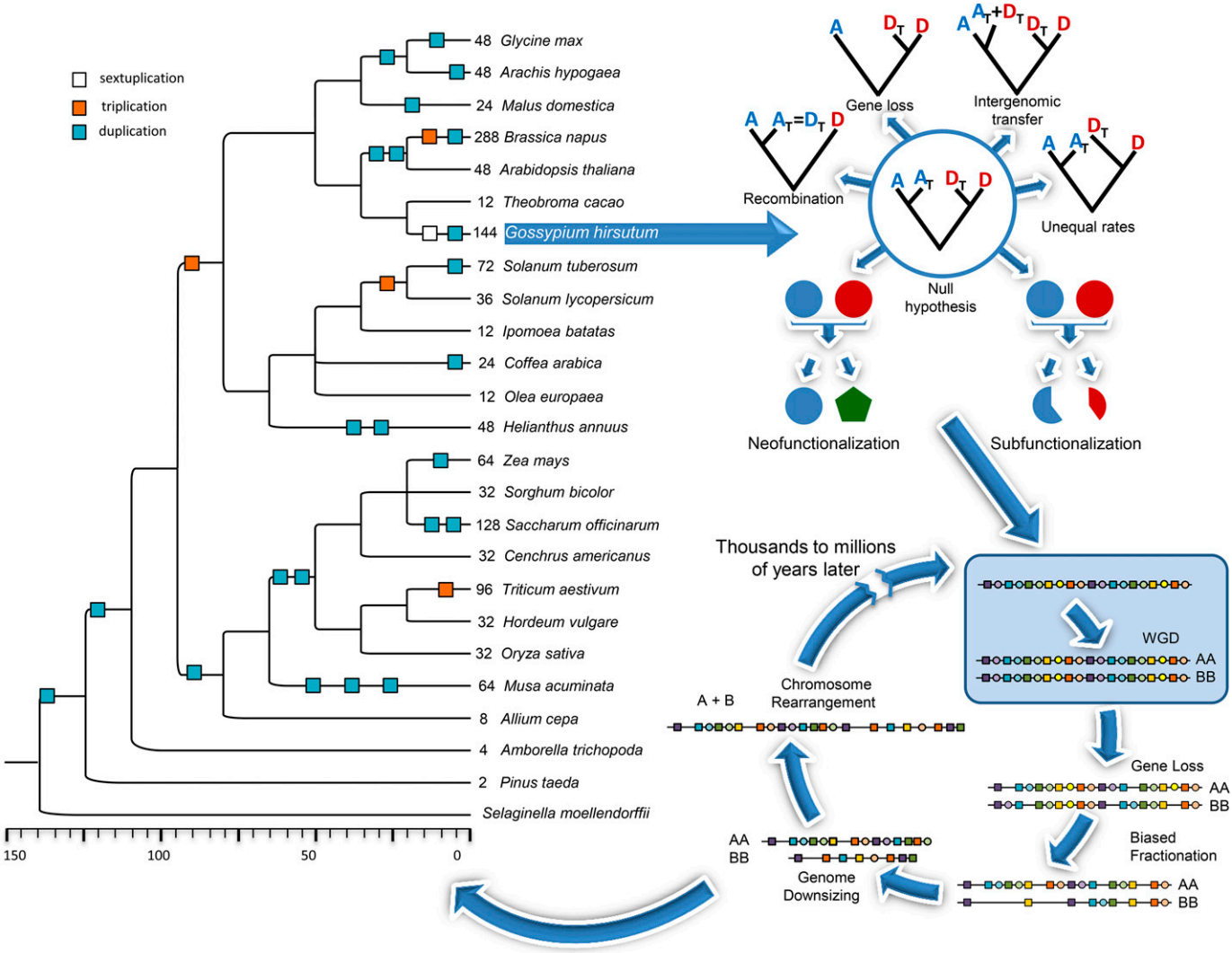

FIGURE 1 Processes and patterns of polyploidy in plants. Left: the evolutionary history of angiosperms includes multiple, repeated polyploidy events, only a fraction of which are illustrated here. Numbers at branch tips indicate the number of genome equivalents derived by multiplication of all previous whole-genome doubling events. Some of the phenomena commonly observed in young allopolyploids are modeled in the top right panel, using *Gossypium* (cotton) as an example. Allopolyploid cottons contain two homoeologous genomes (here A_t and D_t) inherited from their original diploid genome ancestors (A, D); thus, in the absence of intergenomic interactions, gene copies (for example) from A and A_t should be phylogenetically sister to each other, as should those from D and D_t. Responses to polyploidy include those at the genic and genomic level (e.g., homoeologous recombination, gene loss, or silencing, intergenomic colonization or transfer of genic or repetitive elements, differential rates of accumulation of nucleotide substitutions) and those at the transcriptomic or higher levels (e.g., biased homoeolog expression, subfunctionalization, neofunctionalization). These short-term responses lead to the longer-term phenomena diagrammed in the lower right panel, which illustrates gene loss and differential retention of duplicated genes (here, more circles than squares lost), biased genome fractionation (loss of the B genome), the generalized phenomenon of genome downsizing, and chromosome reduction through massive rearrangement. As shown by the arrow at the bottom and in the phylogeny to the left, these processes typically are cyclical, occurring repeatedly on timescales of thousands to millions of years.

(novel expression domain or protein function) (the latter two illustrated in Fig. 1). Even more varied yet poorly understood is the gamut of short-term evolutionary responses to polyploidy at other "omics" levels, including a diverse suite of epigenomic (Salmon et al., 2005; Chen and Ni, 2006; Chen, 2007) and small RNA (Ng et al., 2012) alterations with cascading effects that propagate through proteomic, physiological, and metabolomic networks to ultimately affect plant phenotype and function. Developing an understanding of these myriad interactions and their evolutionary consequences is an exciting area for investigation in the coming years, as diverse technologies and systems approaches are brought to bear on multiple and divergent polyploid systems.

Some of the longer-term consequences of repeated, cyclical polyploidy become evident when one considers that lineages having experienced multiple historical rounds of polyploidy (Fig. 1, phylogeny) do not exhibit especially high chromosome numbers nor genome sizes. If one assumes that the ancestral angiosperm chromosome number is 5–7, as often suggested (Stebbins, 1971; Raven, 1975), then modern *Gossypium hirsutum*, for example, should have 5–7 times 144, or 720–1008 chromosomes in its haploid

complement, in the absence of massive, repeated chromosome number reduction, instead of the 26 it actually contains. Similarly, and notwithstanding the uncertainty regarding actual ancestral angiosperm genome sizes (Soltis et al., 2003), if one postulates a small ancestral genome size of, say, 300 megabases, in the absence of countervailing forces the modern *G. hirsutum* genome should be enormous, 43.2 gigabases (Gb), instead of the 2.4 Gb that it actually contains. Thus, it is immediately apparent that polyploidy is, at least in part, reversible, and that over time it is followed by *massive* chromosomal rearrangements, reductions in chromosome number, and the large-scale loss of both repetitive sequences and duplicated genes (Leitch and Leitch, 2008; Freeling et al., 2012) all leading to genome downsizing (Leitch and Leitch, 2008), as shown in Fig. 1, lower right. This diploidization phenomenon reflects mechanistically diverse processes, operating collectively and over the long haul, ultimately leading to contemporary descendants that behave cytogenetically as normal diploids while harboring in their genomes the vestigial evidence of past polyploidy events.

One intriguing and only partially understood dimension of the genomic diploidization process is that it may be nonrandom with respect to the types of genes retained in duplicate and those returned to single-copy status by mutation. In general, single-copy genes are enriched for essential housekeeping functions and include genes targeted to the chloroplast and those involved in DNA repair and replication; they also are more highly and broadly expressed than are those retained in duplicate (De Smet et al., 2013). This pattern is illustrated in Fig. 1, lower right, where more "circle" than "square" genes are lost. A number of nonmutually exclusive explanations have been forwarded to account for these patterns, including differences in the propensity of genes to experience neutral or selectively imposed subfunctionalization or neofunctionalization, constraints imposed by dosage considerations, and the need for stoichiometric balance in multiprotein complexes (Birchler and Veitia, 2012; Conant, Birchler, and Pires, 2014).

A second intriguing aspect of the diploidization process is that it may be nonrandom with respect to retention of the ancestral progenitor genomes. Phrased alternatively, it has been suggested that one of the two genomes is preferentially retained, often with higher gene expression levels than in the more fractionated genome. This phenomenon, termed biased fractionation, has been widely observed across angiosperm lineages (Schnable et al., 2011; Cheng et al., 2012; Freeling et al., 2012; Renny-Byfield et al., 2015), and leaves evolutionary footprints that may remain visible for at least 60 million years (Renny-Byfield et al., 2015). Biased fractionation is illustrated for the model allopolyploid in Fig. 1, where more genes are shown as being lost from hypothetical progenitor B than from progenitor A.

A major challenge is to connect the genomic processes schematically illustrated in the top right panel of Fig. 1, which operate on a microevolutionary scale, to the longer term processes modeled in the lower right, which ultimately generated the macroevolutionary pattern exemplified by the angiosperm phylogeny shown. While much has been learned about genomic processes accompanying polyploid formation, stabilization and short-term evolution, we have as yet generated little understanding of how these mechanisms generate the ecological, physiological, or morphological variation that is visible to selection and that presumably drives the patterns observed over longer time-scales (Edger et al., 2015; Tank et al., 2015). How do mutation, drift, and selection operate and interact in ecological settings to ultimately generate the longer-term observations of biased fractionation and genome downsizing, for example? These are profoundly challenging and multidimensional research questions, given that selection presumably operates simultaneously at multiple levels, variably across different gene duplicates and gene classes, protein complexes, and genomic regions, and on entire networks of interacting metabolic pathways that generate physiology and phenotypes, all with unique ecological settings and constrained by historical genomic contexts and present population dynamics. Developing a deeper appreciation of modern plant genome architecture from these many perspectives is a grand challenge for plant evolutionary systems biology, one that will likely remain vibrant and revealing for decades to come.

How does our new view of the genesis of modern flowering plant genome organization impact our understanding of plant evolution, form, and function? I suggest that this perspective is central to much of plant biology, a proposal bolstered by the realization that the size and functional diversification of contemporary plant multigene family structures, gene expression patterns, and genic contexts all represent the culmination of this repeated "wash–rinse–repeat" cycle of polyploidization followed by nonrandom and incomplete diploidization. These endpoints, having been shaped by natural selection as well as neutral forces, have created the genic and genomic architecture that underlie diversification and all plant phenotypes, be they physiological, ecological, or morphological (Soltis et al., 2009; De Smet et al., 2013; Rensing, 2014; Edger et al., 2015; Tank et al., 2015).

ACKNOWLEDGEMENTS

I thank my many colleagues and collaborators for their contributions to my thinking and the National Science Foundation and Cotton Incorporated for funding.

LITERATURE CITED

Birchler, J. A., and R. A. Veitia. 2012. Gene balance hypothesis: Connecting issues of dosage sensitivity across biological disciplines. *Proceedings of the National Academy of Sciences, USA* 109: 14746–14753.

Chen, Z. J. 2007. Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annual Review of Plant Biology* 58: 377–406.

Chen, Z. J., and Z. Ni. 2006. Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. *BioEssays* 28: 240–252.

Cheng, F., J. Wu, L. Fang, S. Sun, B. Liu, K. Lin, G. Bonnema, and X. Wang. 2012. Biased gene fractionation and dominant gene expression among the subgenomes of *Brassica rapa*. *PLoS One* 7: e36442.

Coate, J. E., and J. J. Doyle. 2014. Variation in transcriptome size: Are we getting the message? *Chromosoma* 124: 27–43.

Conant, G. C., J. A. Birchler, and J. C. Pires. 2014. Dosage, duplication, and diploidization: Clarifying the interplay of multiple models for duplicate gene evolution over time. *Current Opinion in Plant Biology* 19: 91–98.

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.

Edger, P. P., H. M. Heidel-Fischer, M. Bekaert, J. Rota, G. Glöckner, A. E. Platts, D. G. Heckel, et al. 2015. The butterfly plant arms-race escalated by gene and genome duplications. *Proceedings of the National Academy of Sciences, USA* 112: 8362–8366.

Freeling, M., M. R. Woodhouse, S. Subramaniam, G. Turco, D. Lisch, and J. C. Schnable. 2012. Fractionation mutagenesis and similar consequences of mechanisms removing dispensable or less-expressed DNA in plants. *Current Opinion in Plant Biology* 15: 131–139.

Grover, C., J. Gallagher, E. Szadkowski, M. Yoo, L. Flagel, and J. Wendel. 2012. Homoeolog expression bias and expression level dominance in allopolyploids. *New Phytologist* 196: 966–971.

Hu, G., N. L. Houston, D. Pathak, L. Schmidt, J. J. Thelen, and J. F. Wendel. 2011. Genomically biased accumulation of seed storage proteins in allopolyploid cotton. *Genetics* 189: 1103–1115.

Jackson, S., and Z. J. Chen. 2010. Genomic and expression plasticity of polyploidy. *Current Opinion in Plant Biology* 13: 153–159.

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.

Koh, J., S. Chen, N. Zhu, F. Yu, P. S. Soltis, and D. E. Soltis. 2012. Comparative proteomics of the recently and recurrently formed natural allopolyploid *Tragopogon mirus* (Asteraceae) and its parents. *New Phytologist* 196: 292–305.

Leitch, A., and I. Leitch. 2008. Genomic plasticity and the diversity of polyploid plants. *Science* 320: 481–483.

Ng, D. W., J. Lu, and Z. J. Chen. 2012. Big roles for small RNAs in polyploidy, hybrid vigor, and hybrid incompatibility. *Current Opinion in Plant Biology* 15: 154–161.

Paterson, A. H., X. Wang, J. Li, and H. Tang. 2012. Ancient and recent polyploidy in monocots. *In* P. S. Soltis and D. E. Soltis [eds.], Polyploidy and genome evolution, 93–108. Springer, Berlin, Germany.

Raven, P. H. 1975. The bases of angiosperm phylogeny: Cytology. *Annals of the Missouri Botanical Garden* 62: 724–764.

Renny-Byfield, S., L. Gong, J. P. Gallagher, and J. F. Wendel. 2015. Persistence of sub-genomes in paleopolyploid cotton after 60 million years of evolution. *Molecular Biology and Evolution* 32: msv001.

Rensing, S. A. 2014. Gene duplication as a driver of plant morphogenetic evolution. *Current Opinion in Plant Biology* 17: 43–48.

Salmon, A., M. L. Ainouche, and J. F. Wendel. 2005. Genetic and epigenetic consequences of recent hybridization and polyploidy in *Spartina* (Poaceae). *Molecular Ecology* 14: 1163–1175.

Schnable, J. C., N. M. Springer, and M. Freeling. 2011. Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proceedings of the National Academy of Sciences, USA* 108: 4069–4074.

Soltis, D. E., V. A. Albert, J. Leebens-Mack, C. D. Bell, A. H. Paterson, C. Zheng, D. Sankoff, et al. 2009. Polyploidy and angiosperm diversification. *American Journal of Botany* 96: 336–348.

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.

Soltis, P. S., and D. E. Soltis [eds.]. 2012. Polyploidy and genome evolution. Springer, Berlin, Germany.

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

Tank, D. C., J. M. Eastman, M. W. Pennell, P. S. Soltis, D. E. Soltis, C. E. Hinchliff, J. W. Brown, et al. 2015. Nested radiations and the pulse of angiosperm diversification: increased diversification rates often follow whole genome duplications. *New Phytologist* 207: 454–467.

Wendel, J. F. 2000. Genome evolution in polyploids. *Plant Molecular Biology* 42: 225–249.

Wendel, J. F., L. E. Flagel, and K. L. Adams. 2012. Jeans, genes, and genomes: Cotton as a model for studying polyploidy. *In* P. S. Soltis and D. E. Soltis [eds.], Polyploidy and genome evolution, 181–207. Springer, Berlin, Germany.