bioRxiv preprint doi: https://doi.org/10.1101/212092; this version posted November 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

An analytical framework to understand regulatory novelty accompanying allopolyploidization

Guanjing Hu[1] and Jonathan F. Wendel[1]*

[1] Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA 50011

* Author for Correspondence: Jonathan F. Wendel, Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA 50011, USA, Tel: 515-294-7172, Fax: 515-294-1337, E-mail address: jfw@iastate.edu

bioRxiv preprint doi: https://doi.org/10.1101/212092; this version posted November 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

Allopolyploidy is a prevalent process in plants, having important physiological, ecological, and evolutionary consequences. Massive, genome-wide transcriptomic rewiring in response to genomic merger and doubling has been demonstrated in many allopolyploid systems, encompassing a diversity of phenomena including homoeolog expression bias, genome dominance, expression-level dominance, and revamping of co-expression networks. Here we present an analytical framework to investigate expression change in allopolyploids as governed by distinct sets of intra- and inter-subgenome *cis*-*trans* relationships. The analytical framework devised is a novel extension of classic allele-specific expression analysis to incorporate and distinguish the separate effects of parental regulatory as well as expression differences in *cis* or *trans*, while providing the conceptual basis and tools to reconcile various patterns of regulatory novelty following hybridization and allopolyploidy.

Keywords: allopolyploidy, allele-specific expression (ASE), *cis* and *trans*, homoeolog expression bias, non-additive expression, expression-level dominance

bioRxiv preprint doi: https://doi.org/10.1101/212092; this version posted November 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Polyploidy, or whole-genome duplication (WGD), is exceptionally common in plants, having important physiological, ecological and evolutionary consequences (Stebbins 1940; Levin 1983; Ramsey and Schemske 2002; Leitch and Leitch 2008; Van de Peer et al. 2009; Madlung 2013; Soltis et al. 2014; Van de Peer et al. 2017; Soltis and Soltis 2016). Two types of polyploidy have long been recognized, autopolyploidy, resulting from the multiplication of one progenitor chromosome set, and allopolyploidy, involving hybridization and duplication of divergent parental genomes, classically from different species (Wendel and Doyle 2005). Allopolyploidy in particular is thought to provide avenues for regulatory novelty and hence phenotypic innovation, as evidenced by myriad non-additive and non-Mendelian responses, including gene loss and silencing (Schnable et al. 2011; Freeling et al. 2012; Liu et al. 2014; Mirzaghaderi and Mason 2017; Koh et al. 2010; Szadkowski et al. 2010; Tate et al. 2009; Anssour et al. 2009; Buggs et al. 2009; Eilam et al. 2009), activation of transposable elements (Senerchia et al. 2015; Parisod et al. 2010; Kawakami et al. 2010), epigenetic modifications (Song et al. 2017; Jackson 2017; Madlung et al. 2002; Chen 2007; Salmon et al. 2005; Rapp and Wendel 2005; Bottley 2014; Fulnecek et al. 2009; Kovarik et al. 2008; Yu et al. 2010; Shcherban et al. 2008), and massive, genome-wide transcriptomic rewiring. The latter encompasses a diversity of phenomena, including biased expression of homoeologs on a genic (Yoo and Wendel 2014; Flagel et al. 2008; Akama et al. 2014; Combes et al. 2013; Wang et al. 2016) or even genomic ("genome dominance") scale (Edger et al. 2017; Yang et al. 2016; Zhang et al. 2015; Flagel and Wendel 2010; Schnable et al. 2011; Cheng et al. 2016; Garsmeur et al. 2014), the poorly understood phenomenon of "expression level dominance" (Rapp et al. 2009; Yoo et al. 2013; Grover et al. 2012; Zhang et al. 2016; Akhunova et al. 2010; Liu et al. 2014), and the modification of duplicated gene co-expression networks as reviewed in Gallagher et al. (2016).

bioRxiv preprint doi: https://doi.org/10.1101/212092; this version posted November 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

A hallmark of all of these phenomena is deviation from vertical transmission of preexisting patterns, or the "parental legacy", inherited from the two progenitors (Buggs et al. 2014). These deviations collectively represent regulatory novelty that either accompanied or evolved following genome merger and doubling.

Notwithstanding this progress with respect to the scope and scale of expression alteration accompanying allopolyploidization, there remains a need to further develop the conceptual foundation for understanding gene expression evolution in allopolyploids. Here we propose a new analytical framework based on consideration of the duplicated sets of *cis-trans* relationships that are formed via allopolyploidy, which create both intra- and inter-genomic possibilities. We extend to the polyploid level the classical allele-specific expression analysis (ASE) model in diploids, which aims to distinguish both parental and newly formed *cis*- and *trans*-regulatory divergence. Our model provides a conceptual basis and methods to reconcile various patterns of regulatory novelty following hybridization and allopolyploidy.

As illustrated in Fig. 1 (upper left panel), we use the cotton (*Gossypium* L.) allopolyploid system as an example. Allotetraploid ("AD genome") cottons originated ~1-2 million years ago from a hybridization event between two diploid species ("A" and "D") followed by whole-genome duplication (Wendel and Grover 2015; Wendel and Cronn 2003; Wendel et al. 2010). The descendants of the parental diploid species remain extant ("A2" and "D5"), from which a synthetic F1 hybrid was generated and has previously been used to disentangle expression changes due to hybridization from those arising later from polyploidy and subsequent evolution (Yoo et al. 2014; Flagel et al. 2008; Flagel and Wendel 2010). Within the synthetic F hybrid and

bioRxiv preprint doi: https://doi.org/10.1101/212092; this version posted November 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

natural tetraploid cottons, the expression of each pair of duplicated genes (homoeologs "At" and "Dt"; t denotes subgenome) is governed by four sets of *cis-trans* relationships, including two intra-subgenome interactions derived from each of the parental diploids, and two newly formed inter-subgenome interactions. Noting that sequence evolution and other forms of immediate genetic changes are not expected in a synthetic F1 hybrid, our analytic model allows new mutations to be introduced into both the *cis* and *trans* elements in natural allopolyploids, as a possible source of expression novelty.

Although *cis* and *trans* regulatory divergence between species has long been the focus of classic ASE analysis (Wittkopp et al. 2008), the effects of inter-subgenome *cis-trans* interactions in an inter-specific F1 are inherently neglected based on the assumption that *trans* factors act equally on the *cis* elements of both alleles (Fig. 1, right upper panel; see "if inter = intra"). That is, for diploids, orthologous genes that became two alleles of the same gene in a F1 hybrid experience the same *trans* environment as aggregated *trans* regulatory factors of different parental origins. Thus, allele-specific differences in F1 (in the form of allelic expression ratios At/Dt) that mirror interspecific difference between parents, must reflect differences that are linked in *cis* to the gene itself (i.e., mutations in the gene or its flanking regulatory sequences). Additional differences that are observed between orthologs in the parental species (A2/D5 $\neq$ 1), but not within the hybrid (At/Dt = 1), thereby correspond to the effects of *trans* divergence. The combinatorial effect of *cis* x *trans*, estimated as allele-specific differences in the hybrid that do not mirror the patterns observed between parents (i.e. At/Dt $\neq$ A2/D5) has been consistently observed in different hybrid and allopolyploid systems (Shi et al. 2012; Lemmon et al. 2014; Xu et al. 2014; Combes et al. 2015).

bioRxiv preprint doi: https://doi.org/10.1101/212092; this version posted November 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Here we present a conceptual model that considers allelic divergence in these systems to be determined by all four possible sets of *cis-trans* relationships, regardless of whether inter- and intra- interactions control gene expression differently. If the combination of *trans* factors has the same effects for the two co-resident homoeologs, as assumed in ASE analysis, regulatory divergence in the F1 hybrid reflects parental divergence in *cis* only, *or* expression ratios will be equal to 1 due to the identical *trans* effects. If, however, a *trans* factor acts on the *cis* element of its own origin differently from the *cis* element of its homoeolog, unexpected regulatory patterns might be observed that are attributed to divergence between parental regulatory factors. So with respect to allopolyploids (and hybrids), the category of *cis* x *trans* is interpreted to reflect inter-subgenome interactions. Our analytical framework applies the same algebraic inference as classical ASE analysis, but the conceptual difference is nonetheless meaningful. Insight into actual parental divergence (by only *cis* and only *trans* effects) will be obscured if the relative contribution of *cis* and *trans* effects are mistakenly derived from the combinatorial category of *cis* x *trans*. For example, Lemmon et al. (2014) examined the contribution of *cis* regulatory divergence between maize and teosinte by measuring the proportion of F1 allelic divergence in parental divergence (equivalent to $\frac{\mathrm{At/Dt}}{\mathrm{A2/D5}}$ in the cotton example), which masked the fact that over 30% of regulatory divergence in F1 (At/Dt) is in fact attributed to novel interactions rather than only parental divergence in *cis*; instead, regulatory changes attributed to novel interactions (represented by *cis* x *trans*) should be excluded when examining interspecific divergence between parental species. Using instead our analytical framework, changes accompanying genome merger and doubling are conceptually distinguished from intra-genomic genic and regulatory evolution of *cis* or *trans* elements themselves.

bioRxiv preprint doi: https://doi.org/10.1101/212092; this version posted November 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

We illustrate this approach in Fig. 1 for synthetic F1 hybrid and natural allopolyploid cotton (upper left panel). The resulting regulatory categories and their relative magnitudes will provide insights into the temporal dimension of evolutionary change between duplicated genes and genomes, i.e., to determine whether significant changes (or even lack of changes) in regulatory divergence are primarily associated with the onset of genomic merger, or with subsequent genome doubling followed by other long-term evolutionary changes. In cotton leaf transcriptomes, 28 and 611 genes exhibited novel expression relative to parental diploids in synthetic F1 and natural allotetraploid cottons, respectively (see Fig. 1 upper left; Yoo et al, 2013); this and previous results (Flagel and Wendel 2010; Flagel et al. 2008) consistently suggest a larger overall effect of expression modulation associated with long-term allopolyploid evolution. One might, for example, test the hypothesis that hybridization introduces regulatory divergence between homoeologs of the 28 genes due to the differences in parental regulatory machineries, whereas subsequent (if evolution proceeded by this route (Ramsey and Schemske 1998; Mason and Pires 2015; Kreiner et al. 2017; Kolar et al. 2017) to allopolyploidy) change, following genome doubling and natural selection, may resolve or amplify some regulatory incompatibility (Sun et al. 2017) while permitting the evolutionary exploration of additional expression space, as for the 611 genes having novel expression. Because homoeolog expression patterns in natural allopolyploid species are further complicated by sequence evolution following polyploidy, it has been difficult to distinguish post-polyploidy effects from immediate changes. Our analytical framework allows the evolutionary genetic changes of parental *cis* and *trans* factors to be modelled and integrated into the analysis of expression evolution accompanying allopolyploidy, bringing us a step closer to understanding longer-term post-polyploidy effects.

bioRxiv preprint doi: https://doi.org/10.1101/212092; this version posted November 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Although several key phenomena that characterize novel patterns of duplicated gene expression have been extensively studied in various plants hybrids and allopolyploids, such as additive and non-additive expression, expression-level dominance, genome dominance, transgressive expression and homoeolog expression bias, as reviewed (Yoo et al. 2014), interpreting these patterns across systems remains difficult due to terminological inconsistency (Grover et al. 2012), among other factors. Conceptual relationships among these different phenomena are not well understood, thereby impeding the synthesis required to uncover the underpinnings of duplicate gene expression evolution. The approach outlined here may facilitate an understanding of the interplay between the foregoing phenomena and how ancestral and newly formed *cis-trans* relationships govern expression evolution accompanying genome merger and doubling. As examples, we highlight four specific questions for which our conceptual framework may find utility:

(1) Is homoeolog expression bias in hybrids or allopolyploids primarily determined by ancestral regulatory divergence between parental species, or does it arise from novel inter-subgenome interactions? Homoeolog expression bias is defined as one of two duplicated genes (homoeologs) being expressed more than the other. As outlined in Fig. 1 (upper right and middle panels), genes displaying homoeolog expression bias are classified as being regulated by *cis* and/or inter-subgenome (*cis* x *trans*) effects. Associating the A-biased and D-biased homoeolog expression patterns with categorized regulatory effects, with respect to the ancestral and newly introduced *cis-trans* interactions, will provide insights into the primary mechanism (by *cis* only divergence or

bioRxiv preprint doi: https://doi.org/10.1101/212092; this version posted November 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

by inter-subgenome interactions), both for single pairs of homoeologs and at a genome-wide scale to determine the global balance of A- and D-biases.

(2) What are the *cis* and *trans* regulatory origins of expression level dominance and transgressive expression? Considering the *total* expression aggregated for a pair of homoeologs, expression level additivity is expected in two cases, i.e., if regulatory divergence is determined by only *cis* effects, or when common *trans* effects equilibrate, and hence average, expression of the two ancestral states (Fig. 1, bottom panel). Therefore, non-additive expression patterns, including expression-level dominance and transgressive expression levels, arise from novel inter-subgenomic interactions; thus all four sets of regulatory interactions are involved. That is, non-additive *total* expression patterns are determined by the joint effects of inter- and intra-subgenome interactions, which can lead to large variations in the patterns of non-additive expression, depending on the divergence of parental genomes, the age of allopolyploid events, and condition-specific regulatory interactions. Consistent with this expectation, the fraction of duplicated genes that exhibit non-additive expression varies largely among allopolyploid systems and populations, various tissue types, developmental stages and environmental conditions, ranging from less than 1% to 7% in different allohexaploid wheat species (Chelaifa et al. 2013; Chague et al. 2010), from 23 to 61% among variable cotton tissues (Rambani et al. 2014; Flagel and Wendel 2010; Yoo et al. 2013), and from 42% to 60% under two temperature conditions in coffee (Bardil et al. 2011).

bioRxiv preprint doi: https://doi.org/10.1101/212092; this version posted November 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(3) How is the direction of expression level dominance determined by *cis* and *trans* regulation, and is it related to the direction of homoeolog expression bias? It has been suggested that expression-level dominance toward one parent is mainly caused by up- or down-regulation of the allele of the other parent (Shi et al. 2012; Yoo et al. 2013; Cox et al. 2014; Combes et al. 2015). Taking the A-dominant expression pattern as an example (Fig. 1, bottom panel; see "A-dominant" row), the joint effect of both inter-subgenome interactions and the intra D-genome regulation approximates the effect of intra A-genome regulation; if up- or down- regulation is mainly observed for the Dt homoeolog (as governed by intra D→D and inter A→D), inter-subgenome interactions need to show a stronger effect on Dt than At (inter A→D ≫ inter D→A), indicating that expression level dominance towards one genome (A-dominant) is determined by its predominant *trans* actions (stronger A than D *trans* factor) in inter-subgenome interactions. In other words, previous studies consistently suggest that expression level dominance is associated with the asymmetric effects between two types of inter-subgenome *cis*-*trans* interactions, and preferentially toward the genome acting dominantly in *trans*. Because inter-subgenome interactions can up- or down- regulate the target alleles, the direction of expression level dominance appears *not* to be associated with the direction of homoeolog bias.

(4) Finally, how are gene-to-gene networks rewired by genomic merger and doubling, in terms of the genome-wide collection of inter- and intra- subgenome interactions? As recently reviewed by Gallagher et al. (2016), co-expression network analysis in polyploids not only has the potential to facilitate a better understanding of the complex

bioRxiv preprint doi: https://doi.org/10.1101/212092; this version posted November 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

'omics' underpinnings of phenotypic and ecological traits, but also may provide novel insight into the interaction among duplicated genes and genomes. As demonstrated for bread wheat (Pfeifer et al. 2014) and cotton (Hu et al. 2016) development networks, biased expression of homoeologs and asymmetric recruitment of the parental sub-network structures (modules) may be associated with specific functions and phenotypic traits. Because these wheat and cotton studies are based on aggregated co-expression relationships of homoeologs, one future direction is to generate networks considering homoeolog expression separately, thereby allowing the direct evaluation of inter- and intra-subgenome interactions in the network context. Accordingly, examining the derived network topology will reveal the degree of network conservation of ancestral interactions and how the union of two diverged regulatory systems reshapes network architecture through gain and loss of intra- and inter-subgenome interactions, respectively (Conant 2010; Conant and Wolfe 2008; Conant and Wolfe 2006). Although co-expression relationships do not necessarily represent physical interactions between *cis* and *trans* regulatory elements, the gene-to-gene interconnections that are inferred based on the "guilt-by-association" principle will provide an unprecedented opportunity to understand genome-scale transcriptional responses to genome evolution, when co-analyzed with parental *cis*-*trans* divergence, inter-subgenome incompatibility, homoeolog bias and expression-level dominance, as proposed.

bioRxiv preprint doi: https://doi.org/10.1101/212092; this version posted November 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

Akama S, Shimizu-Inatsugi R, Shimizu KK, Sese J (2014) Genome-wide quantification of homeolog expression ratio revealed nonstochastic gene regulation in synthetic allopolyploid *Arabidopsis*. Nucleic Acids Res 42 (6):e46. doi:10.1093/nar/gkt1376

Akhunova AR, Matniyazov RT, Liang H, Akhunov ED (2010) Homoeolog-specific transcriptional bias in allopolyploid wheat. BMC Genomics 11:505. doi:10.1186/1471-2164-11-505

Anssour S, Krugel T, Sharbel TF, Saluz HP, Bonaventure G, Baldwin IT (2009) Phenotypic, genetic and genomic consequences of natural and synthetic polyploidization of *Nicotiana attenuata* and *Nicotiana obtusifolia*. Ann Bot 103 (8):1207-1217. doi:10.1093/aob/mcp058

Bardil A, de Almeida JD, Combes MC, Lashermes P, Bertrand B (2011) Genomic expression dominance in the natural allopolyploid *Coffea arabica* is massively affected by growth temperature. New Phytol 192 (3):760-774. doi:10.1111/j.1469-8137.2011.03833.x

Bottley A (2014) Epigenetic variation amongst polyploidy crop species. In: Epigenetics in Plants of Agronomic Importance: Fundamentals and Applications. Springer International Publishing, pp 33-46. doi:10.1007/978-3-319-07971-4_3

Buggs RJ, Doust AN, Tate JA, Koh J, Soltis K, Feltus FA, Paterson AH, Soltis PS, Soltis DE (2009) Gene loss and silencing in *Tragopogon miscellus* (Asteraceae): comparison of natural and synthetic allotetraploids. Heredity 103 (1):73-81. doi:10.1038/hdy.2009.24

Buggs RJ, Wendel JF, Doyle JJ, Soltis DE, Soltis PS, Coate JE (2014) The legacy of diploid progenitors in allopolyploid gene expression patterns. Philosophical transactions of the Royal Society of London Series B, Biological sciences 369 (1648):20130354. doi:10.1098/rstb.2013.0354

Chague V, Just J, Mestiri I, Balzergue S, Tanguy AM, Huneau C, Huteau V, Belcram H, Coriton O, Jahier J, Chalhoub B (2010) Genome-wide gene expression changes in genetically stable synthetic and natural wheat allohexaploids. New Phytol 187 (4):1181-1194. doi:10.1111/j.1469-8137.2010.03339.x

Chelaifa H, Chague V, Chalabi S, Mestiri I, Arnaud D, Deffains D, Lu Y, Belcram H, Huteau V, Chiquet J, Coriton O, Just J, Jahier J, Chalhoub B (2013) Prevalence of gene expression additivity in genetically stable wheat allohexaploids. New Phytol 197 (3):730-736. doi:10.1111/nph.12108

Chen ZJ (2007) Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. Annu Rev Plant Biol 58:377-406. doi:10.1146/annurev.arplant.58.032806.103835

Cheng F, Sun C, Wu J, Schnable J, Woodhouse MR, Liang J, Cai C, Freeling M, Wang X (2016) Epigenetic regulation of subgenome dominance following whole genome triplication in *Brassica rapa*. New Phytol 211 (1):288-299. doi:10.1111/nph.13884

Combes M-C, Dereeper A, Severac D, Bertrand Bi, Lashermes P (2013) Contribution of subgenomes to the transcriptome and their intertwined regulation in the allopolyploid *Coffea arabica* grown at contrasted temperatures. New Phytologist 200 (1):251-260

Combes M-C, Hueber Y, Dereeper A, Rialle S, Herrera J-C, Lashermes P (2015) Regulatory divergence between parental alleles determines gene expression patterns in hybrids. Genome biology and evolution 7 (4):1110-1121

bioRxiv preprint doi: https://doi.org/10.1101/212092; this version posted November 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Conant GC (2010) Rapid reorganization of the transcriptional regulatory network after genome duplication in yeast. Proceedings of the Royal Society of London B: Biological Sciences 277 (1683):869-876

Conant GC, Wolfe KH (2006) Functional partitioning of yeast co-expression networks after genome duplication. PLoS biology 4 (4):e109-e109

Conant GC, Wolfe KH (2008) Turning a hobby into a job: how duplicated genes find new functions. Nat Rev Genet 9 (12):938-950. doi:10.1038/nrg2482

Cox MP, Dong T, Shen G, Dalvi Y, Scott DB, Ganley ARD (2014) An interspecific fungal hybrid reveals cross-kingdom rules for allopolyploid gene expression patterns. PLoS Genet 10 (3):e1004180-e1004180

Edger PP, Smith RD, McKain MR, Cooley AM, Vallejo-Marin M, Yuan Y-W, Bewick AJ, Ji L, Platts AE, Bowman MJ, others (2017) Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-old naturally established neo-allopolyploid monkeyflower. The Plant Cell Online:tpc-00010

Eilam T, Anikster Y, Millet E, Manisterski J, Feldman M (2009) Genome size in natural and synthetic autopolyploids and in a natural segmental allopolyploid of several *Triticeae* species. Genome 52 (3):275-285. doi:10.1139/G09-004

Flagel L, Udall J, Nettleton D, Wendel J (2008) Duplicate gene expression in allopolyploid *Gossypium* reveals two temporally distinct phases of expression evolution. BMC Biol 6:16. doi:10.1186/1741-7007-6-16

Flagel LE, Wendel JF (2010) Evolutionary rate variation, genomic dominance and duplicate gene expression evolution during allotetraploid cotton speciation. New Phytol 186 (1):184-193. doi:10.1111/j.1469-8137.2009.03107.x

Freeling M, Woodhouse MR, Subramaniam S, Turco G, Lisch D, Schnable JC (2012) Fractionation mutagenesis and similar consequences of mechanisms removing dispensable or less-expressed DNA in plants. Curr Opin Plant Biol 15 (2):131-139. doi:10.1016/j.pbi.2012.01.015

Fulnecek J, Matyasek R, Kovarik A (2009) Faithful inheritance of cytosine methylation patterns in repeated sequences of the allotetraploid tobacco correlates with the expression of DNA methyltransferase gene families from both parental genomes. Mol Genet Genomics 281 (4):407-420. doi:10.1007/s00438-008-0420-8

Gallagher JP, Grover CE, Hu G, Wendel JF (2016) Insights into the ecology and evolution of polyploid plants through network analysis. Mol Ecol 25 (11):2644-2660. doi:10.1111/mec.13626

Garsmeur O, Schnable JC, Almeida A, Jourda C, D'Hont A, Freeling M (2014) Two evolutionarily distinct classes of paleopolyploidy. Mol Biol Evol 31 (2):448-454. doi:10.1093/molbev/mst230

Grover CE, Gallagher JP, Szadkowski EP, Yoo MJ, Flagel LE, Wendel JF (2012) Homoeolog expression bias and expression level dominance in allopolyploids. New Phytol 196 (4):966-971. doi:10.1111/j.1469-8137.2012.04365.x

Hu G, Hovav R, Grover CE, Faigenboim-Doron A, Kadmon N, Page JT, Udall JA, Wendel JF (2016) Evolutionary conservation and divergence of gene coexpression networks in *Gossypium* (cotton) seeds. Genome biology and evolution 8 (12):3765-3783. doi:10.1093/gbe/evw280

Jackson SA (2017) Epigenomics: dissecting hybridization and polyploidization. Genome Biol 18 (1):117. doi:10.1186/s13059-017-1254-7

bioRxiv preprint doi: https://doi.org/10.1101/212092; this version posted November 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Kawakami T, Strakosh SC, Zhen Y, Ungerer MC (2010) Different scales of *Ty1*/*copia*-like retrotransposon proliferation in the genomes of three diploid hybrid sunflower species. Heredity (Edinb) 104 (4):341-350. doi:10.1038/hdy.2009.182

Koh J, Soltis PS, Soltis DE (2010) Homeolog loss and expression changes in natural populations of the recently and repeatedly formed allotetraploid Tragopogon mirus (Asteraceae). BMC Genomics 11:97. doi:10.1186/1471-2164-11-97

Kolar F, Certner M, Suda J, Schonswetter P, Husband BC (2017) Mixed-ploidy species: progress and opportunities in polyploid research. Trends Plant Sci. doi:10.1016/j.tplants.2017.09.011

Kovarik A, Dadejova M, Lim YK, Chase MW, Clarkson JJ, Knapp S, Leitch AR (2008) Evolution of rDNA in *Nicotiana* allopolyploids: a potential link between rDNA homogenization and epigenetics. Ann Bot 101 (6):815-823. doi:10.1093/aob/mcn019

Kreiner JM, Kron P, Husband BC (2017) Evolutionary dynamics of unreduced gametes. Trends Genet 33 (9):583-593. doi:10.1016/j.tig.2017.06.009

Leitch AR, Leitch IJ (2008) Genomic plasticity and the diversity of polyploid plants. Science 320 (5875):481-483. doi:10.1126/science.1153585

Lemmon ZH, Bukowski R, Sun Q, Doebley JF (2014) The role of *cis* regulatory evolution in maize domestication. PLoS Genet 10 (11):e1004745. doi:10.1371/journal.pgen.1004745

Levin DA (1983) Polyploidy and novelty in flowering plants. The American Naturalist 122 (1):1-25

Liu S, Liu Y, Yang X, Tong C, Edwards D, Parkin IA, Zhao M, Ma J, Yu J, Huang S, Wang X, Wang J, Lu K, Fang Z, Bancroft I, Yang TJ, Hu Q, Wang X, Yue Z, Li H, Yang L, Wu J, Zhou Q, Wang W, King GJ, Pires JC, Lu C, Wu Z, Sampath P, Wang Z, Guo H, Pan S, Yang L, Min J, Zhang D, Jin D, Li W, Belcram H, Tu J, Guan M, Qi C, Du D, Li J, Jiang L, Batley J, Sharpe AG, Park BS, Ruperao P, Cheng F, Waminal NE, Huang Y, Dong C, Wang L, Li J, Hu Z, Zhuang M, Huang Y, Huang J, Shi J, Mei D, Liu J, Lee TH, Wang J, Jin H, Li Z, Li X, Zhang J, Xiao L, Zhou Y, Liu Z, Liu X, Qin R, Tang X, Liu W, Wang Y, Zhang Y, Lee J, Kim HH, Denoeud F, Xu X, Liang X, Hua W, Wang X, Wang J, Chalhoub B, Paterson AH (2014) The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. Nat Commun 5:3930. doi:10.1038/ncomms4930

Madlung A (2013) Polyploidy and its effect on evolutionary success: old questions revisited with new tools. Heredity 110 (2):99-104. doi:10.1038/hdy.2012.79

Madlung A, Masuelli RW, Watson B, Reynolds SH, Davison J, Comai L (2002) Remodeling of DNA methylation and phenotypic and transcriptional changes in synthetic *Arabidopsis* allotetraploids. Plant Physiol 129 (2):733-746. doi:10.1104/pp.003095

Mason AS, Pires JC (2015) Unreduced gametes: meiotic mishap or evolutionary mechanism? Trends Genet 31 (1):5-10. doi:10.1016/j.tig.2014.09.011

Mirzaghaderi G, Mason AS (2017) Revisiting pivotal-differential genome evolution in wheat. Trends Plant Sci 22 (8):674-684. doi:10.1016/j.tplants.2017.06.003

Parisod C, Alix K, Just J, Petit M, Sarilar V, Mhiri C, Ainouche M, Chalhoub B, Grandbastien MA (2010) Impact of transposable elements on the organization and function of allopolyploid genomes. New Phytol 186 (1):37-45. doi:10.1111/j.1469-8137.2009.03096.x

bioRxiv preprint doi: https://doi.org/10.1101/212092; this version posted November 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Pfeifer M, Kugler KG, Sandve SR, Zhan B, Rudi H, Hvidsten TR, International Wheat Genome Sequencing C, Mayer KF, Olsen OA (2014) Genome interplay in the grain transcriptome of hexaploid bread wheat. Science 345 (6194):1250091. doi:10.1126/science.1250091

Rambani A, Page JT, Udall JA (2014) Polyploidy and the petal transcriptome of *Gossypium*. BMC Plant Biol 14:3. doi:10.1186/1471-2229-14-3

Ramsey J, Schemske DW (1998) Pathways, mechanisms and rates of polyploidy formation in flowering plants. Annual Review of Ecology and Systematics 29 (1):467-501. doi:10.1146/annurev.ecolsys.29.1.467

Ramsey J, Schemske DW (2002) Neopolyploidy in flowering plants. Annual Review of Ecology and Systematics 33 (1):589-639. doi:10.1146/annurev.ecolsys.33.010802.150437

Rapp RA, Udall JA, Wendel JF (2009) Genomic expression dominance in allopolyploids. BMC Biol 7:18. doi:10.1186/1741-7007-7-18

Rapp RA, Wendel JF (2005) Epigenetics and plant evolution. New Phytol 168 (1):81-91. doi:10.1111/j.1469-8137.2005.01491.x

Salmon A, Ainouche ML, Wendel JF (2005) Genetic and epigenetic consequences of recent hybridization and polyploidy in *Spartina* (Poaceae). Mol Ecol 14 (4):1163-1175. doi:10.1111/j.1365-294X.2005.02488.x

Schnable JC, Springer NM, Freeling M (2011) Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proc Natl Acad Sci USA 108 (10):4069-4074. doi:Doi 10.1073/Pnas.1101368108

Senerchia N, Felber F, Parisod C (2015) Genome reorganization in F1 hybrids uncovers the role of retrotransposons in reproductive isolation. Proc Biol Sci 282 (1804):20142874. doi:10.1098/rspb.2014.2874

Shcherban AB, Badaeva ED, Amosova AV, Adonina IG, Salina EA (2008) Genetic and epigenetic changes of rDNA in a synthetic allotetraploid, *Aegilops sharonensis* x *Ae. umbellulata*. Genome 51 (4):261-271. doi:10.1139/G08-006

Shi X, Ng DW, Zhang C, Comai L, Ye W, Chen ZJ (2012) *Cis*- and *trans*-regulatory divergence between progenitor species determines gene-expression novelty in *Arabidopsis* allopolyploids. Nat Commun 3:950. doi:10.1038/ncomms1954

Soltis PS, Liu X, Marchant DB, Visger CJ, Soltis DE (2014) Polyploidy and novelty: Gottlieb's legacy. Phil Trans R Soc B 369 (1648):20130351-20130351

Soltis PS, Soltis DE (2016) Ancient WGD events as drivers of key innovations in angiosperms. Curr Opin Plant Biol 30:159-165. doi:10.1016/j.pbi.2016.03.015

Song Q, Zhang T, Stelly DM, Chen ZJ (2017) Epigenomic and functional analyses reveal roles of epialleles in the loss of photoperiod sensitivity during domestication of allotetraploid cottons. Genome Biol 18 (1):99-99

Stebbins GL (1940) The significance of polyploidy in plant evolution. The American Naturalist 74 (750):54-66

Sun Y, Wu Y, Yang C, Sun S, Lin X, Liu L, Xu C, Wendel JF, Gong L, Liu B (2017) Segmental allotetraploidy generates extensive homoeologous expression rewiring and phenotypic diversity at the population level in rice. Mol Ecol 26 (20):5451-5466. doi:10.1111/mec.14297

Szadkowski E, Eber F, Huteau V, Lode M, Huneau C, Belcram H, Coriton O, Manzanares-Dauleux MJ, Delourme R, King GJ, Chalhoub B, Jenczewski E, Chevre AM (2010) The first meiosis of resynthesized *Brassica napus*, a genome blender. New Phytol 186 (1):102-112. doi:10.1111/j.1469-8137.2010.03182.x

bioRxiv preprint doi: https://doi.org/10.1101/212092; this version posted November 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Tate JA, Joshi P, Soltis KA, Soltis PS, Soltis DE (2009) On the road to diploidization? Homoeolog loss in independently formed populations of the allopolyploid *Tragopogon miscellus* (Asteraceae). BMC Plant Biol 9:80. doi:10.1186/1471-2229-9-80

Van de Peer Y, Maere S, Meyer A (2009) The evolutionary significance of ancient genome duplications. Nat Rev Genet 10 (10):725-732. doi:10.1038/nrg2600

Van de Peer Y, Mizrachi E, Marchal K (2017) The evolutionary significance of polyploidy. Nat Rev Genet 18 (7):411-424. doi:10.1038/nrg.2017.26

Wang X, Zhang H, Li Y, Zhang Z, Li L, Liu B (2016) Transcriptome asymmetry in synthetic and natural allotetraploid wheats, revealed by RNA-sequencing. New Phytol 209 (3):1264-1277. doi:10.1111/nph.13678

Wendel FJ, Doyle JJ (2005) Polyploidy and evolution in plants. In: Henry RJ (ed) Plant diversity and evolution: genotypic and phenotypic variation in higher plants. CABI Publishing, Wallingford, UK, pp 97-117. doi:10.1079/9780851999043.0097

Wendel JF, Brubaker CL, Seelanan T (2010) The origin and evolution of *Gossypium*. In: Stewart JM, Oosterhuis DM, Heitholt JJ, Mauney JR (eds) Physiology of Cotton. Springer Netherlands, pp 1-18. doi:10.1007/978-90-481-3195-2_1

Wendel JF, Cronn RC (2003) Polyploidy and the evolutionary history of cotton. Advances in Agronomy 78:139-186

Wendel JF, Grover CE (2015) Taxonomy and evolution of the cotton genus, *Gossypium*. In: Fang DD, Percy RG (eds) Cotton. Agronomy Monograph, vol 57, 2nd edn. American Society of Agronomy, Madison, WI, pp 25-42. doi:10.2134/agronmonogr57.2013.0020

Wittkopp PJ, Haerum BK, Clark AG (2008) Regulatory changes underlying expression differences within and between *Drosophila* species. Nat Genet 40 (3):346-350. doi:10.1038/ng.77

Xu C, Bai Y, Lin X, Zhao N, Hu L, Gong Z, Wendel JF, Liu B (2014) Genome-wide disruption of gene expression in allopolyploids but not hybrids of rice subspecies. Mol Biol Evol 31 (5):1066-1076. doi:10.1093/molbev/msu085

Yang J, Liu D, Wang X, Ji C, Cheng F, Liu B, Hu Z, Chen S, Pental D, Ju Y, Yao P, Li X, Xie K, Zhang J, Wang J, Liu F, Ma W, Shopan J, Zheng H, Mackenzie SA, Zhang M (2016) The genome sequence of allopolyploid *Brassica juncea* and analysis of differential homoeolog gene expression influencing selection. Nat Genet 48 (10):1225-1232. doi:10.1038/ng.3657

Yoo MJ, Liu X, Pires JC, Soltis PS, Soltis DE (2014) Nonadditive gene expression in polyploids. Annu Rev Genet 48:485-517. doi:10.1146/annurev-genet-120213-092159

Yoo MJ, Szadkowski E, Wendel JF (2013) Homoeolog expression bias and expression level dominance in allopolyploid cotton. Heredity 110 (2):171-180. doi:10.1038/hdy.2012.94

Yoo MJ, Wendel JF (2014) Comparative evolutionary and developmental dynamics of the cotton (*Gossypium hirsutum*) fiber transcriptome. PLoS Genet 10 (1):e1004073. doi:10.1371/journal.pgen.1004073

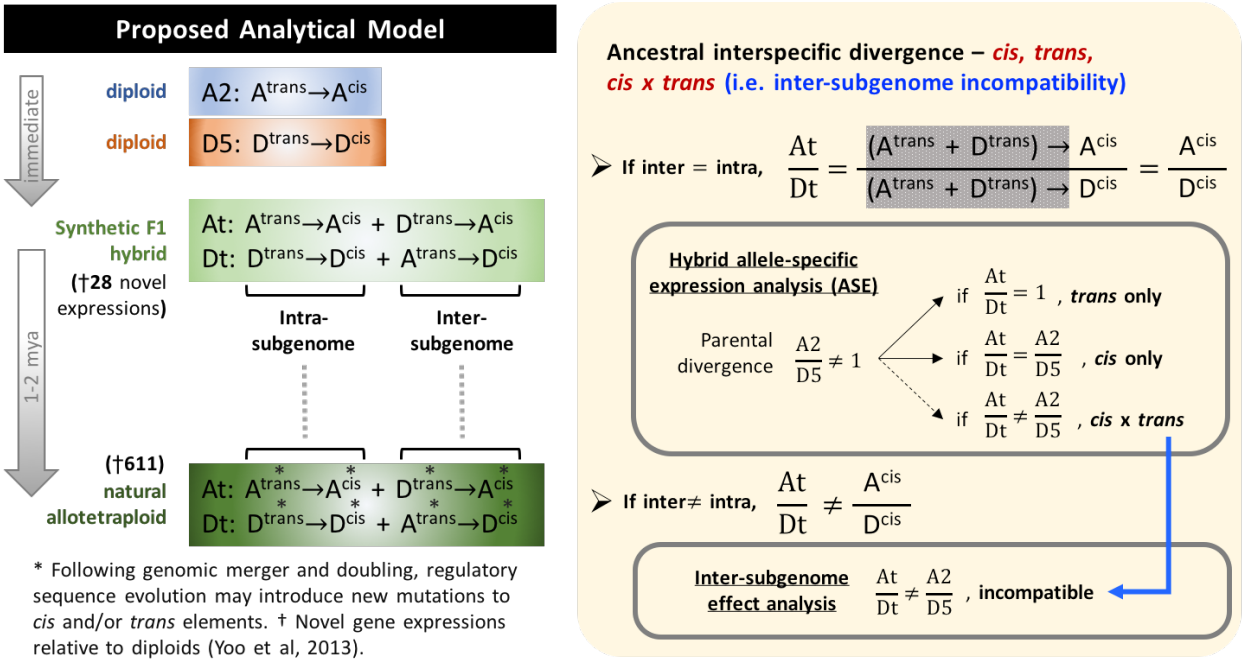
Yu Z, Haberer G, Matthes M, Rattei T, Mayer KF, Gierl A, Torres-Ruiz RA (2010) Impact of natural genetic variation on the transcriptome of autotetraploid *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences of the United States of America 107 (41):17809-17814. doi:10.1073/pnas.1000852107

Zhang D, Pan Q, Tan C, Zhu B, Ge X, Shao Y, Li Z (2016) Genome-wide gene expressions respond differently to A-subgenome origins in *Brassica napus* synthetic hybrids and natural allotetraploid. Front Plant Sci 7:1508. doi:10.3389/fpls.2016.01508

bioRxiv preprint doi: https://doi.org/10.1101/212092; this version posted November 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Zhang T, Hu Y, Jiang W, Fang L, Guan X, Chen J, Zhang J, Saski CA, Scheffler BE, Stelly DM, Hulse-Kemp AM, Wan Q, Liu B, Liu C, Wang S, Pan M, Wang Y, Wang D, Ye W, Chang L, Zhang W, Song Q, Kirkbride RC, Chen X, Dennis E, Llewellyn DJ, Peterson DG, Thaxton P, Jones DC, Wang Q, Xu X, Zhang H, Wu H, Zhou L, Mei G, Chen S, Tian Y, Xiang D, Li X, Ding J, Zuo Q, Tao L, Liu Y, Li J, Lin Y, Hui Y, Cao Z, Cai C, Zhu X, Jiang Z, Zhou B, Guo W, Li R, Chen ZJ (2015) Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. Nat Biotechnol 33 (5):531-537. doi:10.1038/nbt.3207

bioRxiv preprint doi: https://doi.org/10.1101/212092; this version posted November 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



Partitioned homoeolog expression patterns – Homoeolog expression bias

Class	Hypothesis	Analytic model	Interpretation
At bias	At/Dt > 1	$A^{trans}\rightarrow A^{cis} + D^{trans}\rightarrow A^{cis} > D^{trans}\rightarrow D^{cis} + A^{trans}\rightarrow D^{cis}$	Determined by Intra- (ancestral divergence) or inter- effects, or both.
Dt bias	At/Dt < 1	$A^{trans}\rightarrow A^{cis} + D^{trans}\rightarrow A^{cis} < D^{trans}\rightarrow D^{cis} + A^{trans}\rightarrow D^{cis}$	

Aggregated homoeolog expression patterns – Additivity, dominance, and transgressive expression

Expression description	Measure	Model
Total expression (T)	T= At+Dt	$A^{trans}\rightarrow A^{cis} + D^{trans}\rightarrow A^{cis} + D^{trans}\rightarrow D^{cis} + A^{trans}\rightarrow D^{cis}$
Parental A-genome expression	A2	$2 \times (A^{trans}\rightarrow A^{cis})$
Parental D-genome expression	D5	$2 \times (D^{trans}\rightarrow D^{cis})$
Mid parental expression	(A2+D5)/2	$A^{trans}\rightarrow A^{cis} + D^{trans}\rightarrow D^{cis}$

Factor of 2 is used to account for normalization across ploidy levels

Class	Hypothesis	Analytic model	Interpretation
Additivity	T = (A2+D5)/2	$D^{trans}\rightarrow A^{cis} + A^{trans}\rightarrow D^{cis} = 0$	inter- effects equal to 0
Transgressive expression	T > max(A2,D5) *or* T < min(A2,D5)	$D^{trans}\rightarrow A^{cis} + A^{trans}\rightarrow D^{cis} > 0$ or $D^{trans}\rightarrow A^{cis} + A^{trans}\rightarrow D^{cis} < 0$	inter- effects NOT equal to 0
A-dominant	T = A2	$A^{trans}\rightarrow A^{cis} = D^{trans}\rightarrow A^{cis} + D^{trans}\rightarrow D^{cis} + A^{trans}\rightarrow D^{cis}$	Asymmetric inter-effects have been consistently reported.
D-dominant	T = D5	$D^{trans}\rightarrow D^{cis} = A^{trans}\rightarrow D^{cis} + D^{trans}\rightarrow A^{cis} + A^{trans}\rightarrow A^{cis}$	

Figure 1. A proposed analytical model for understanding regulatory novelty accompanying hybridization and allopolyploidy, using *Gossypium* diploids A2 and D5, and their derived homoeologs At and Dt. Shown is an extension of the classic allelic-specific expression (ASE) analysis (upper right panel). Applications for understanding homoeolog expression bias (middle panel) and non-additive expression patterns (bottom panel) are illustrated.